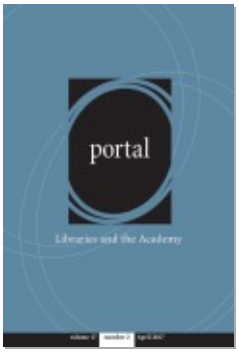
Pioneers in the Wild West: Managing Data Collections

Karen Hogenboom, Michele Hayslett



portal: Libraries and the Academy, Volume 17, Number 2, April 2017, pp. 295-319 (Article)

Published by Johns Hopkins University Press
DOI: *https://doi.org/10.1353/pla.2017.0018*

Pioneers in the Wild West: Managing Data Collections

Karen Hogenboom and Michele Hayslett

abstract: During the last few years, many academic libraries have accepted the challenge of helping their users locate and acquire the numeric data they need. To meet their users' ever-increasing need for data, librarians are purchasing data sets one at a time ("small data"). This service, though important to our users, raises many issues in the areas of collection scope, acquisition procedures, and discovery and access. The authors conducted a survey of data librarians in summer 2015 and followed up by interviewing five data librarians in depth to report on how academic libraries collect and manage small data and to explore the strengths and weaknesses of various approaches.

Introduction

Thirty years ago, relatively few social scientists used numeric or spatial data in their research. Currently, however, there is a strong quantitative strain in the research done in almost every social science field. Academic librarians have evolved with this trend, and many academic libraries have at least one data librarian who works with researchers to obtain and use the data sets they need to answer their research questions. These librarians have many different titles, from *social sciences data librarian* to *government information librarian* to *GIS* (geographic information systems) *librarian*. But all of them consider it part of their mission to collect data as well as more traditional library materials.

Many researchers need data created by a third party, whether they are survey results, genome sequences, weather observations, or something else. Some third-party data are freely available on the Internet, but other valuable data are kept behind a paywall and cannot be seen without a paid subscription, controlled by organizations that may have collected the data for business purposes or for other nonacademic reasons. Several library vendors offer data products that compile statistics from many different sources to be downloaded, visualized—that is, presented in a graphical or pictorial format—or both. Much of the data used by researchers, however, lives "in the wild." Researchers

portal: Libraries and the Academy, Vol. 17, No. 2 (2017), pp. 295–319.

may need a high-resolution satellite image of a landscape feature in South America or a parcel-level description of properties that were foreclosed on during the United States housing crisis, for example. These data sets are only available from small distributors, who may know little about standard library acquisitions practices and may have terms in their sales agreements and licenses that are difficult or impossible for a library to accept. For example, some international licenses require any adjudication to take place outside the United States. Data sets also differ from print-based library materials because they need thorough documentation to be used correctly and effectively, documentation which is often lacking from the vendor. Like other library materials, data sets need a place to be stored and a way to let users know they are available.

Although many social scientists and humanists used data 30 years ago, there were few data archives available where they could store their data or look for data that others had created. Some large studies provided data useful for a variety of research questions in many disciplines, including the United States Census and the General Social Survey, a sociological study that gathered information about Americans' concerns, experiences, attitudes, and practices. Since then, data files have evolved from tapes, to compact discs (CDs) and DVDs, to e-files downloadable online. However, the ease of accessing data does not always correspond to easy discovery or easy use. Data can be found on association websites, in CDs in the back of print books, from commercial compilers, and as supplemental material on academic journal websites, for example. Once researchers discover the data set they need, access may or may not be free, and good documentation may or may not be included. More academic libraries have established data services in the last 10 years in correlation with the increased discoverability of data "in the wild" and the growth of data use, in combination with the issues in licensing, accessing, and storing data. The situation of data services in the libraries at the University of North Carolina (UNC) at Chapel Hill may demonstrate the general trend: the first librarian here who wrangled data was one of the government documents librarians who helped users with census and other government data. When exactly the position became oriented more specifically toward data is difficult to pinpoint, but data had become an important component of the job by the late 1990s. Only in 2008 did the position receive the outright title *data librarian*, and the position no longer fell under the umbrella of government documents but instead formed part of a data services unit with the GIS librarian. The GIS librarian position itself was instituted in 2002 as part of the Reference Department (as is the government documents unit). When created in 2009, the data services unit remained within the Reference Department (now called Research and Instructional Services). Prior to the middle to late 1990s, as personal computers became ubiquitous and data retrieval easier, most campus data services across the United States and Canada belonged to data centers, such as the University of Michigan's Institute for Social Research in Ann Arbor

The ease of accessing data does not always correspond to easy discovery or easy use. Data can be found on association websites, in CDs in the back of print books, from commercial compilers, and as supplemental material on academic journal websites, for example.

or UNC's Odum Institute for Research in Social Science in Chapel Hill, rather than to libraries. The University of Illinois in Urbana-Champaign took the same path from government documents to data services, but eight or nine years later.

In most academic data services, collecting data has become an important component. We surveyed academic data librarians to discover how they identify, obtain, and provide access to small data sets for their institutions. Five survey respondents agreed to be interviewed and discussed data collection in more detail than is possible in a survey. Almost all of the literature on this topic discusses specific projects undertaken by individual institutions, so the results of this study are a chance to look for best practices as well as describe a "state of the issue" for management of data collections.

Literature Review

In 2007, Hilary Davis and John Vickery pointed to several signs that data were becoming the "currency" of scholarly communication. They noted that commercial vendors were beginning to create products with which researchers could access and manipulate data. They also mentioned legislation attempting to make data subject to copyright; the exponential growth of data produced; and some publishers' then-new practice of requiring data to be submitted with manuscripts for review. Vendors have found a niche in repackaging public data so that the information can be more easily discovered and used, both within the product and after downloading into a statistical or GIS package.[1] A recent article on managing collections of data stated this point even more strongly: "Research, development and innovation are fed by high-quality primary data. Reuse of this detailed, unit-record data—microdata—can advance science, lead to innovations that boost economies, and offer solutions to pressing social problems."[2]

While many general principles of collection development apply to data collections, such as buying material that is aligned with user needs and is available to the entire campus, the peculiarities of data sets add extra considerations to the collector's task. The data set might be a spreadsheet too large to attach to an e-mail or open in Microsoft Excel, a JPEG image file, or a .txt file containing unformatted text. The data cannot be used without accompanying documentation about how, when, and where they were collected, and how the data are structured. Many researchers want data about individual units of observation, or microdata, but these data often contain personally identifiable information, and access must be carefully monitored. Furthermore, there are no catalogs or approval plans for data sets, so it is difficult for librarians to identify these materials for purchase unless and until a user requests them.

In the 1960s and 1970s, faculty researchers were mostly on their own in managing what were known then as "machine-readable data files." Few libraries had the expertise to assist, and few data centers existed. Initially, many researchers ran analysis via punch cards and relied on the data processing staff in the precursors to today's information technology departments. Libraries were concerned about data files, though. Sue Dodd's 1982 landmark publication, *Cataloging Machine-Readable Data Files: An Interpretive Manual*,[3] was the culmination of almost 10 years of conversations among catalogers about how to wrestle the hydra of this odd medium so that research data could be discovered, accessed, and reused by researchers.

It is important to remember that, until the work carried out by Dodd and others, there were no widely known and systematically organized catalogs, inventories, or bibliographies of data files in the United States. The International Social Science Council (ISSC), which aims to advance the behavioral and economic sciences, recognized the need to establish data archives, largely due to the efforts of the Norwegian political scientist Stein Rokkan. In the United States, the Council of Social Science Data Archives made similar efforts. A publication calling for bibliographic conventions and standards was written by David Nasatir under contract to UNESCO to study "overcoming the barriers to realizing the fullest utilization of machine-readable social science data." Nasatir called not only for the preparation of bibliographic details but also for archives to enable variable-level searching across studies and across archives.[4]

The cataloging of data marks the beginning of wider academic library involvement in data services. Libraries quickly gained footing in offering data services as secondary data—data sets available to be reused by researchers other than those who originally created them—became more widely available. The Census Bureau and the Inter-University Consortium for Political and Social Research (ICPSR) were early distributors of data via punch cards and magnetic tape, but few individual researchers could share their data using these media.

Library-based data services became more prolific once distribution became easier. As one measure of academic library involvement in data services, ICPSR's director of marketing and membership, Linda Detterman, provided statistics about the location of the consortium's official representatives in member institutions over time. In 1988 (and earlier), the great majority of official representatives were faculty. Over ensuing years, though, library-based representatives gained ground until they consistently outnumbered faculty official representatives in almost every year between 1998 and 2014.[5]

The rise of the CD-ROM in the late 1980s is another milestone in academic libraries supporting data use. Despite skepticism on the part of some (expressed by both John Lowe and Cary Griffith),[5] CDs quickly took off as a convenient new distribution medium and became nearly ubiquitous in libraries. Ease of processing and use may have been the greatest advantage of CDs. If they had licensing terms at all, these products tended to include click-through licenses that displayed automatically when users opened the product so that libraries had no further responsibility for informing users of license terms. (Some publishers such as Woods & Poole Economics, Inc. still employ this method of informing users about license terms for their products.) Terms were nonnegotiable, prices were usually comparable to prices for reference works in print (for example, books), and librarians generally had little to worry about regarding the acquisitions process other than deciding whether to buy particular products. Even that was easy because vendors often had many CD products and would provide libraries with catalogs of their offerings. CDs generally arrived with labels, documentation, or both

CDs generally arrived with labels, documentation, or both that made it relatively easy for catalogers to build online public access catalog (OPAC) records for them, enabling librarians to process CDs with the same speed as print materials.

that made it relatively easy for catalogers to build online public access catalog (OPAC) records for them, enabling librarians to process CDs with the same speed as print materials, and with the subsequent result that users were able to find these materials easily.

After technical processing, reference staff could often install the discs on library computers so patrons did not themselves need to install the disc they wanted to use but merely find the computer on which it was available. Discs tended to include built-in access software or, if the interface was DOS (disk operating system) based, the CD arrived with a manual explaining the specialized commands necessary to manipulate and extract the data (the Census Bureau's DataEXTRACT product was a good example of a resource that required complicated user instructions). Finally, in direct contrast to the delicate nature of magnetic tape and computer diskettes, this medium proved very durable. A disc could not be inadvertently erased, and the machines that ran them could not easily mangle them. If handled with reasonable care to avoid scratches, CDs remained usable for decades. Capacity was a problem at first, but that improved over time: whereas some large collections on CD initially ran to 10 discs or more, the arrival on the market of DVDs in the middle to late 1990s largely resolved that issue. Overall, libraries could manage discs much like print materials, with only a few slight changes in procedure: perhaps a library would decide (as did Davis Library at UNC at Chapel Hill) that the discs would be accessible to patrons only when a staff member could help, assuring that users had both assistance in dealing with rare technical issues and less opportunity to "liberate" a resource. If a product required input of a password or license number, library staff would safeguard such information, again usually to prevent a patron absconding with a disc.

To illustrate the burgeoning of CD products, in 1989, Ron Rietdyk, then president of SilverPlatter Information, Inc., one of the first companies to produce commercial reference databases on CD-ROMs, reported "that more CD-ROM discs were produced in the first five months of 1989 than all of 1988."[6] While the 1980 Census was distributed only on magnetic tape that was readable mainly in data centers, the 1990 Census was made available primarily on CDs that could be read on any personal computer.

In the past few years, however, CDs have become a more problematic format for libraries. At UNC at Chapel Hill, the staff time necessary to install and maintain access to the immense variety of products has become unsustainable. Sometimes one product is not compatible with others, necessitating installation on a different machine. In recent years, fewer computers come with CD drives at all and, even when they do, operating systems have advanced to the point that they often cannot interpret the systems for which the CDs were originally designed, making the products difficult or impossible to install or use. Much content issued on CD was never updated, too, making currency a serious issue. Consequently, frustration among both staff and users has blossomed. As a result, the reference staff in Davis Library has undertaken drastic weeding of its CD collection in the past few years.

Meanwhile, many vendors have abandoned CDs in favor of online availability. ICPSR used magnetic tape to distribute both informational materials and data from the mid-1970s to the mid-1980s, producing its first diskette for a user in 1984, as noted in the consortium's online Timeline.[7] The consortium only distributed limited numbers of data sets on CD. According to Assistant Director Mary Vardigan, its CD offerings consisted

of "generally special collections of studies, e.g., studies on Black Americans."[9] The consortium initiated file transfer protocol (FTP) downloads in 1996 for exchanging files over the Internet. For commercial distributors, this move online sometimes led to exponential price increases. One well-known vendor that in 2000 provided an index to their data on CD for just over $500 now charges $8,000 per year for one-user-at-a-time online access. Unlimited campus access costs $30,000 per year. Admittedly, the new system enables visualization, direct download, and analysis of the data, significant improvements over the predecessor product, which served merely as a pointer to where the data existed in print. But this example demonstrates the skyrocketing cost of new data products and how vendors leverage the added benefits of new access models to raise prices.

The landscape of data services is changing dramatically and rapidly in other ways as well. Data formats are proliferating exponentially. Resources might arrive formatted for particular statistical analysis software, such as SAS (Statistical Analysis Software), SPSS (originally Statistical Package for the Social Sciences), R (supported by the R Foundation for Statistical Computing), or Stata (a contraction of *statistics* and *data*), or in a specialized GIS format. Advances in qualitative analysis software have led to many new file formats for data, including images, text, audio, and video. Resources might arrive on CD, on an external hard drive, or virtually by e-mail or FTP download. They might have a built-in user interface, or they might not. They might include labels, documentation, both, or neither. Often catalogers cannot tell what the resource is at all if it arrives with a cryptic label or no label, and with no documentation. Technical staff often lack the software simply to open the files. To make their collections available, data services staff must now often build relationships and workflows with technical services personnel to manage a multitude of functions they seldom had responsibility for in the past: licensing, acquisitions, and cataloging, in addition to the usual tasks of making data available to users and assisting users in using and understanding the information.

Other trends are making data more discoverable. While libraries for a time relied on unique, often homegrown databases to list data sets separate from their regular catalogs, Terrence Bennett and Shawn Nicholson allude to a growing trend in the sophistication of library OPACs that enables users to limit searches to data sets.[10] Some libraries have pushed boundaries in negotiating licenses to gain rights to virtualize distribution, broadening availability to all university affiliates wherever they are, whether in the library, in their offices or dorm rooms, or in another state or country. However, it is unclear from the literature how many libraries leverage new technologies and publisher negotiations to provide wider access, how many have developed more detailed catalog records or evaluated license terms, and how many actively inform their users about license restrictions. The survey and interviews were designed to get a sense of how libraries handle these issues and to identify any best practices that emerged from the results.

Methodology

The survey asked U.S. and Canadian academic librarians about their "libraries' practices and policies around purchasing and providing access to data sets other than large packages sold by library vendors." (Statista and ProQuest's *Statistical Abstract of the United States* were used as examples of large packages.) Questions covered individuals' roles in

selecting and acquiring data sets and their practices in collecting data sets across many stages: identification of what is available; collection development, weeding, and scope; budget; licensing; discoverability; and access.

The survey was pretested by two academic data librarians in North Carolina. In early June 2015, the authors sent an appeal to the e-mail list of the International Association for Social Science Information Services and Technology (IASSIST), asking U.S. and Canadian academic librarians who purchase data (not including large packages) to follow a link to a Web survey. A follow-up message was sent one month later. The survey was hosted in Qualtrics, and most results were processed within the same program. Survey responses were anonymous unless the respondent chose to provide an e-mail address.

Inquiries in April 2016 show U.S. and Canadian IASSIST membership at just over 300 people, but limiting to those in academic libraries yields about 150 individuals. Twenty-seven people responded to the survey, giving a response rate of roughly 18 percent.[11] Consequently, while these results obviously cannot be termed representative, they are at least descriptive of practices in some libraries and may serve as a benchmark for future studies. No tests of statistical significance were run due to the low response rate. (The authors did not take the survey themselves but do speak from their experience in the course of this article.) Skip patterns took respondents to the next relevant question in the survey, depending on their previous answer. As a result, relatively small numbers of respondents saw certain questions; the survey, with skip logic included, is presented in the Appendix.

Respondents were invited to provide their e-mail addresses at the end of the survey if they were willing to be interviewed. Five individuals provided contact information and were interviewed in late October and early November 2015. All participants were asked what they thought the greatest challenges of collecting data are; about their collection development policies (if they had any); what license terms were most problematic for them; how they inform users about license terms; and what they would like to do differently vis à vis their data collections. Some interviewees were asked follow-up questions about specific responses they gave in the survey. Responses were categorized to understand the diversity of their content and to identify possible trends and best practices.

Results

Demographics

Of 27 respondents, 85 percent (23) indicated that their institutions grant doctorates as the highest degree. Seven percent (2) reported master's degrees as their highest, and another 7 percent (2) indicated baccalaureate degrees. Forty-eight percent of respondents (13 of 27) worked at institutions that serve more than 20,000 students. Thirty percent (8 of 27) held positions at smaller institutions, with enrollments of fewer than 10,000 students, and 22 percent (6 of 27) worked at institutions that serve between 10,000 and 20,000 students. Ninety-two percent of 26 respondents (24) said they purchased numeric or GIS data.[8] Of the two who did not, one indicated that large package sources "have been sufficient to fulfill the needs of our students & faculty. It's not that we wouldn't consider doing it if something came up—we just haven't had to so far." The other noted, "Any

data purchases are done by the respective subject/liaison librarian. Data Services does not have a collections budget."

Of the 24 who bought data, 20 reported when their libraries began this practice. Answers ranged from 1965 to within the last five years. Twenty percent (4 of 20) said their libraries began collecting data before 1985. Another 30 percent (6 of 20) started between the middle 1990s and 2000. The remaining 50 percent (10 of 20) started after 2000. One of these respondents noted, "We've had a librarian paying close attention to numeric data in some capacity since at least 2000, but a formal position started in 2008. If you count gov docs data (Census), it's probably been a long time." This is the case at the University of North Carolina at Chapel Hill and, though a generalization, seems likely to hold true at many institutions.

Collection Development

Roles in the Collections Process

Those respondents who did collect data played a variety of roles in the purchasing process. Eighty-six percent (18 of 21) selected data sets for acquisition, but they played a variety of other roles as well (see Table 1).

In most libraries, people other than the respondent also played major roles in collecting data. Only one of the respondents was the sole person involved. For the others, different staff members formed part of the process at different points. In many libraries, subject librarians selected and funded data purchases. Two respondents mentioned that they bought data only when library users requested it, and five reported that other data librarians besides themselves also selected data. One respondent's interlibrary loan department selected data based on requests from their users to obtain data from other institutions. Data librarians' involvement in purchasing data varies when they are not the primary selector. Some librarians review all purchases by other selectors. Others consult with selectors on request or proactively recommend data sources for acquisition.

Only two respondents stated that they had their own collections budget. The data librarian at the University of Illinois at Urbana-Champaign also has a dedicated collections budget and is responsible for managing the subscriptions to major data sources as well as purchasing small data sets. In other cases, a subject librarian, fund manager, or disciplinary team paid for purchases, or an administrator reviewed and funded them. Only one data librarian responded that he or she negotiated licenses for data. Others used the expertise of their acquisitions staff, electronic resources staff, or administrators. Actual purchases were also done by others, including administrators or people in acquisitions or electronic resources. Cataloging or systems experts provided access to data in most cases. In only one case was the data librarian responsible for providing access to purchased data.

In many cases, the respondents did not consider themselves responsible for licensing, though 59 percent (10 of 17) stated that they were more involved in license review than other librarians without specific responsibility for data. Some had responsibility for knowing the consequences of license terms, download options, and data formats. They might have been the first person to review a data set license, then took their evaluation

Table 1.

Respondents' roles in purchasing data ($n = 21$; multiple answers possible)

Role	Number of respondents	Percentage
Selector	18	86
Advising other selectors	5	24
Managing/making collection accessible	1	4
Negotiating licenses	1	4
Managing/overseeing budget	2	8

to others more involved in licensing. Respondents were often called in when problems came up with a license to articulate the value of a data set so the library could balance the benefits of having it versus the problems of complying with the license terms. Respondents sometimes worked with the vendor's staff responsible for licensing to suggest alternate license terms and to discuss the implications of particular restrictions in proposed licenses. One library assigned each license to a single person in acquisitions, and the respondent ordered the resource using a form specific to data sets. The form elicits information about many of the issues that other libraries address only as they arise.

License Terms

Interviewees were very articulate on the topic of license terms that they preferred and those they could not accept. A few vendors have terms that require researchers to submit any scholarly papers written using the vendor's data to the vendor, either before or after publication. This term was never acceptable, but fortunately seemed rare. Another term that was a deal breaker for everyone who mentioned it is a clause that makes the library responsible if a campus researcher violates the terms of the license. A few vendors require that the library report each user of the data to the vendor, and this was a problem for most survey respondents and interviewees. The broader problem with all of these terms is that they require libraries to monitor or control the behavior of their users, something that is logistically difficult or impossible, and that conflicts with many libraries' policies about user privacy. Some survey respondents stated they would accept terms like those if the data were sufficiently valuable for their campus, but most considered such provisions a deal breaker.

A few vendors have terms that require researchers to submit any scholarly papers written using the vendor's data to the vendor, either before or after publication. This term was never acceptable.

One provision that can be problematic but which occurs in almost every license is the jurisdiction clause, wherein the vendor specifies that any legal claims based on the license must be adjudicated in the state or country where their company has its headquarters. Many respondents work at state universities, which are considered state agencies. Most states require state agencies to sign contracts specifying that claims may only be brought in that state; therefore, libraries at state institutions are prohibited from agreeing to such terms. Our respondents said many data purchases fall through because of this license provision, but many libraries have also succeeded in getting the jurisdiction changed to their home state. For both the University of Illinois and the University of North Carolina at Chapel Hill, this jurisdictional issue has led to failed negotiations with some international vendors but has been less of a problem with U.S distributors.

One interviewee expressed a strong preference for purchasing data over licensing data. She had worked with one vendor who would not sell its data outright, and the library would lose access to all of the data if it did not buy updates each year. She pointed out that many academic researchers want to take a longitudinal approach to data and so need information from past years, but the majority of online access points to data are designed for users who need only current information.

Another interviewee pointed out the variety of license terms that address who can access the data. The majority of resources sold by distributors accustomed to the library market allow access to anyone who is at a computer on campus or anyone who logs in to the campus network from home. However, many data licenses from vendors unfamiliar with libraries restrict access to computers with a campus IP (Internet protocol) address that are logged into by an individual. This means that students and faculty cannot obtain the data from off campus, and community users cannot access the data at all. Several respondents said that they have signed licenses despite this kind of restriction if the data were essential to their campus users. One interviewee always tries to negotiate access for community residents but occasionally must sign a license that is restricted to campus users. Some vendors insist that the data be kept on one computer, with access mediated by the library, or require researchers to create individual accounts using their university credentials. Licenses that restrict access to only one user (not merely one user at a time) were also almost always impossible for librarians to sign.

Many academic researchers want to take a longitudinal approach to data and so need information from past years, but the majority of online access points to data are designed for users who need only current information.

When researchers request data and librarians find that they cannot purchase the material because of problematic license terms, several respondents have worked with the researcher to obtain the data for them in other ways. In some cases, the data were free to an individual researcher. In other cases, the researcher had funds that could be used to acquire the data. The University of North Carolina Libraries has occasionally advised faculty about negotiating their own licenses in such cases.

Collection Development Policies

Five of the 17 librarians who collect data had a formal collection development policy, and three of those policies were available online. An additional librarian submitted a collection development policy that was not online. Two of the online collection development statements covered numeric data only. The simplest statement was that of the University of North Carolina at Greensboro, which appears as a sentence in a service description: "The Data Services Librarian will attempt to purchase data sets for [university] researchers dependent on funding and the Library's ability to comply with license and use restrictions."[12] The most complex statement has sections for the scope of the collection, factors affecting collection development, levels of collection intensity, and collection management issues.[13] The third policy statement includes geographic information and maps as well as numeric data. This policy also states how the covered collection relates to other collections in the library; policies related to time period, date of publication, and geographic scope; and a gift policy.[14] Another policy, which is not publically available, applies to numeric data only and is directed to subject librarians who are considering data sets for their collections, because there is no separate budget for data. It addresses format, subject, and geographic and chronological scope of the library's data collection, as well as access principles and related resources that should be checked before ordering overlapping data sets.

Most of the respondents lacked a written collection development policy. Two-thirds (8 of 12) relied at least partially on user requests to shape their collection, and requests were the only method of collection development mentioned by six of these respondents. The main deciding factors for the remaining two respondents were a limited budget in one case and a general collection development policy for the library in the other. Three respondents aligned their collections with the subject strengths of the library where they work. One had recently changed jobs and was building a collection of data sets at her new institution using her previous position's resources as a model. Two of these respondents were writing formal collection development policies at the time of the survey.

The value of a collection development policy came up several times in our interviews. One interviewee felt that having a collection development policy was appropriate because her other colleagues who collect other types of materials have such policies. Another interviewee felt that the collection development statement she was drafting would document her decision process and the criteria for price, access, and format. She saw her policy as a document that will be helpful to someone holding her position in the future and a tool that she could consult to articulate the scope of her collection to users.

Respondents had many ways to identify data for purchase within the parameters they had established either informally or formally in their collection development statement. User requests were a common way to choose data (88 percent, 15 of 17 respondents). Other methods of finding data were notices from vendors, professional networks (other librarians and e-mail lists), and disciplinary literature. Three respondents mentioned some type of environmental scan, matching collection areas with disciplinary strengths on campus, comparing their holdings to similar libraries' collections, or doing an informal gap analysis.

Only one respondent had a written weeding policy, which states that the collection is regularly reviewed but does not give specific criteria for review. Three respondents did not weed their data collections. For those who did weed, the most common reasons they would discard a data set are requirements from the vendor, problems with the data (for example, lack of accuracy), and data that are superseded. Other factors that influenced the decision to weed data were problematic access, data that require obsolete software to access them or are unreadable for other reasons, duplication of content, large file size, and lack of future value. One respondent weighed problematic access against usage and would weed a data set that was difficult to access only if it was not heavily used. An interviewee compared her policy for weeding data to general weeding principles, reporting that material will be weeded if it is no longer worthwhile. She has weeded data sets that become available in a more stable format or that are "snapshots" gathered at a particular time with no previous history or future additions expected.

Budget

Survey respondents were evenly split on whether they have a collections budget specifically for purchasing data. All the respondents who knew how their data collections budget compared to the library's collections budget as a whole had 2 percent or less of the total budget; most had 1 percent or less. Respondents who mentioned a specific amount had between $5,000 and $50,000 per year, with $50,000 allocated for data subscriptions as well as small data purchases.

If respondents did not have a data collections budget, the most common way to fund purchases was to work with subject specialists or bibliographers who contributed their knowledge of data sets to the bibliographer's selection process (6 of 8). Two respondents had gift or endowment funds available to acquire data. In one case, the interlibrary loan department had a budget and would purchase data if requested by a local library patron. Some data librarians were also subject specialists and acquired data in their discipline as part of their normal collections budget.

Scope of Collections

Respondents were invited to describe the scope of their collections "in terms of subject areas, number of data sets, or whatever other way makes sense to describe it." Four respondents described the scope of their collections in terms of the number of data sets they contained, ranging from approximately 20,000 to "very few." Because the question did not specify a way to count files, some of these numbers might be individual files and others might be larger collections. The rest of the respondents employed subjects or formats: three mentioned specifically that they collect GIS data, and three stated that their collections are not limited by subject. Four respondents limited their collections to social science data (in one case, Canadian social science data), and one specialized in health sciences data. A final respondent was trying to determine the scope of his or her data collection because data had been purchased and managed by subject librarians in the past and no one had an overall picture of what had been acquired. At Illinois, where requests drive all purchases of small data sets, the collection consists of about 10 data sets in a wide variety of formats and file sizes, from Excel spreadsheets to historical newspa-

pers formatted for text analysis and genomic data. Other data have been purchased as subscriptions hosted by the vendor. At UNC at Chapel Hill, the size of the data collection is about the same as that at Illinois and comes in similar formats.

Access

Mediated/Unmediated Access

Eighty-eight percent of respondents (15 of 17) indicated they provided access to data sets virtually, that is, online without requiring the patron to come to the library. This confirms our expectation that such arrangements will proliferate as users expect more online access generally and as secure technical setups become easier to administer. Technical setups vary, though. Most respondents (60 percent, or 9 of 15) indicated they employed secure servers. Thirty-three percent (5 of 15) reported they had access for walk-in users as well (presumably such access will vary depending on the particular vendor). Twenty percent (3 of 15) indicated they could or did provide data upon request to individuals, by direct e-mail, on a CD, or by secure file transfer protocol (sFTP, which ensures that the data are securely transferred using a safe, private data stream).

Notification of License Terms

Eighty-one percent of respondents (13 of 16) indicated their vendors required that the library inform users of any license terms or restrictions. Twelve of those 13 respondents commented on the terms or restrictions required, the most commonly cited terms being that the data were not for commercial use or were for educational use only (50 percent, or 6 of 12). One institution feels this interpretation is too narrow because some vendors have interpreted the "educational use only" language to prohibit research utilization of their data, specifying that "educational use" refers only to classroom and student use. UNC at Chapel Hill includes a standard fair use clause: "Nothing in this agreement shall be construed to limit the right of the Licensee or any Authorized User to use the information in accordance with the Fair Use provision of U.S. copyright law." If the vendor will not accept that language, the licensing team attempts to insert a statement that users have permission to study and cite the data for "research, teaching, and private study purposes." If the license lacks the fair use clause and will not allow expanded use cases to include research and private study, the purchasing subject librarian must decide if he or she still wants to license the data.

Some vendors have interpreted the "educational use only" language to prohibit research utilization of their data, specifying that "educational use" refers only to classroom and student use.

Other terms respondents mentioned as required by vendors were generalizations. For example, one respondent said, "Some vendors require signatures from each user of the data with a separate DUA [data use agreement]," or "Some [vendors] do [require users be informed of terms or restrictions] but our Electronic Resource Management unit deals with that, for digital content of any sort." A few noted specific requirements such as "confidential data; publishing rights; levels one can report; academic use only."

Of the 16 respondents who said they inform users of any license terms or restrictions beyond those required by the vendor, 25 percent (4 of 16) indicated they told users about all license terms, 19 percent (3 of 16) reported they informed library patrons about some license terms, and 56 percent (9 of 16) said they conveyed no license terms aside from those the distributors require. For those who answered that they inform users about some license terms, the survey asked what license terms or restrictions they tell users about, aside from any required by vendors. All three responded (100 percent), giving the following answers:

1. Not for commercial use, data are for current affiliates only, do not redistribute the data.
2. As determined by the researchers' need to know citations, use restrictions pertaining to the research, deletion dates, and the like.
3. We have a generic statement on all electronic resources informing users that the contents may be subject to copyright and they need to see the specific terms.

The seven people who answered they informed their users of all or some license terms in addition to those required by vendors received an additional question about how they conveyed those license terms. This was a multiple-choice question with the instruction to choose all options that applied. Table 2 shows the 19 responses that were given across five categories. Both responses in the "Other" category indicated staff communicated personally with users. Because most of those we spoke with also referred to communicating personally with users as their most common means of conveying license information, perhaps this category would have been the most popular if the survey had included that option.

Technology

Eighty-eight percent of respondents (15 of 17) indicated that they provided some access to data sets virtually. (The survey defined "virtual access" as both available online and not requiring the patron to come to the library; some respondents described multiple modes of access.) When these 17 respondents were subsequently asked to describe their technical setups, they described as many different arrangements as they had (so the total was more than 100 percent). Fifteen respondents of the 17 who were asked described seven different types of virtual setups. Eleven of the 15 (73 percent) described secure server setups; two (13 percent) reported allowing access only to specific individuals. One of those two indicated they could set up a secure file transfer protocol (sFTP) for users. While probably many data collections have this capability, this does not seem like a scalable approach for delivering a wide variety of data sets to many users because each connection would have to be set up separately. Twenty percent (3 of 15) described virtual access being only through Web-based vendor servers. One of these three respondents indicated looking to future virtual access being provided through their digital collections repository or through OpenGeoportal.org, a website that provides open source geospatial data. (Two setups, service to walk-in users and delivering data on CD-ROM, were not actually virtual, and so these responses were not considered in this calculation.)

Table 2.

Methods for communicating license terms to users ($n = 7$; multiple answers possible)

Method	Number of respondents	Percentage
Pop-up or click-through license display	4	57
Notes in catalog records	4	57
Notes in subject or class pages	4	57
Notes on the download page or README files with the data	5	71
Other (please specify)	2	28

Assessment

Once a data set has been purchased and is accessible, many survey respondents can track how often it is used. Twelve percent (2 of 17) of survey respondents knew exactly how often their data were used, either using Google Analytics, which tracks and reports website traffic, or server statistics if the data are hosted on a library server. Sixty-five percent of respondents (11 of 17) had a general idea of how often their data were used, but 23.5 percent (4 of 17) lacked a sense of how often the data were utilized. When asked how they get a general sense of how frequently data sets are used, interviewees mentioned data sets where they need to mediate access, or the impression they formed from reference and instruction interactions. Some had server counts for some data sets and could tell exactly how much those were used, but for other data sets they relied on word of mouth or requests they received for specific data. This situation with data collections is similar to usage metrics for many library databases. Some libraries get reports on usage statistics from the nonprofit organization COUNTER (Counting Online Usage of Networked Electronic Resources). Many librarians rely on anecdotal information about usage to decide what to buy and retain. In addition, for data sets, the infrastructure of individual institutions heavily influences how data may be mounted and, as a result, how usage statistics may be gathered, so evaluating usage presents a unique challenge for each individual library. This is an area in need of further study.

Challenges

Interviewees mentioned several challenges that stood out from the rest when collecting data. Starting at the beginning of the acquisitions process, two librarians mentioned that simply determining that a data set exists and is available can be a huge challenge. If someone must have food supply data for Africa at a smaller geographic level than

by country, for example, the information may or may not be available from the United Nations or from individual country governments. A librarian might think that some organization must compile data about quantities of commodities shipped between the ports on the Mediterranean Sea, but identifying that organization and persuading it to share its data can be a challenge. Another interviewee saw his biggest challenge as predicting what will be useful to his campus and balancing the money available for data with the needs of the campus.

Simply determining that a data set exists and is available can be a huge challenge.

An interviewee at a large public university felt that licensing is the biggest challenge in collecting data and was eloquent about the wide variety of terms she has come across in licenses. As another interviewee pointed out, prices for data vary widely and seem not to correspond to the effort necessary to compile or publish the data in many cases. She also mentioned the problems of data storage: finding a place where all of campus can access the data, backing up the data, and making sure that people not authorized to use the data cannot see the information. Librarians can buy some products through a consortium, but they can usually only do so with big databases such as Statista or Social Explorer, which provides demographic information about the United States. The kind of data sets we considered for this study are generally produced as small products by individual researchers or vendors. Libraries represent a small percentage of customers for such dealers, and so there is less opportunity to engage with them and influence their terms. No solutions for this challenge present themselves currently, but data librarians talk to one another about specific data sources. Perhaps organizations for data professionals might influence vendors, especially for products that most data librarians purchase.

Prices for data vary widely and seem not to correspond to the effort necessary to compile or publish the data in many cases.

Interviewees' "Wish Lists"

At the end of each interview, interviewees discussed what they wish they could do differently about the way they collect small data sets. Many of these hopes were about making data collections more discoverable. Two interviewees would like to work with information technology (IT) departments, catalogers, or Web designers to provide a robust website or catalog entries that would lead seamlessly to data. As a prerequisite, licenses should consistently allow data sets to be hosted on a server, rather than on removable media in the library. Another related hope is that vendors would provide better metadata for their data sets.

Two interviewees wanted to develop a more standardized workflow for acquiring small data sets, instead of responding mainly to patron requests. Other desires touched on by interviewees included marketing data holdings to users; ways to stream large data sets online to make them more accessible and usable; clear workflows for moving data sets efficiently through acquisitions, cataloging, preservation, and access services to make them available to patrons more quickly; and having checks on data quality from vendors.

Discussion

How an academic library purchases data depends largely on local factors such as its size, the curriculum it supports, its information technology infrastructure, and its budget. But several threads run throughout. The first is the value of writing a collection development statement, both to guide selection and acquisition and to document decision-making principles and procedures. While the librarians without collection development statements could nevertheless articulate how they make decisions and how their library goes about acquiring data, those with statements could also document how they work with other librarians and acquisitions staff, how they prioritize requests, and other practices. As with any collection development statement, articulating these issues is important not only so that everyone involved in data purchases knows the parameters but also to explain to library patrons the scope of the services offered.

Another common trend is using patron requests to shape small data collections. Whether funding is controlled by the data librarian, a subject specialist, or an administrator, if someone requests a data set, our respondents will try to obtain it. Most research libraries have traditionally taken a "just in case" approach to collecting, buying as many books and journals as possible to have them on hand in case a need arises. In these days of shrinking budgets, and with evidence that the majority of library holdings never get used, more libraries are adopting a "just in time" philosophy, buying many resources only when patrons ask for them. Data requests can take a long time to complete, however, because license terms, storage, and access issues extend the process exponentially compared to a request for a book or journal article.

Whether funding is controlled by the data librarian, a subject specialist, or an administrator, if someone requests a data set, our respondents will try to obtain it.

Most respondents indicated that they rely on other staff to purchase data sets, whether subject librarians or staff members in acquisitions, licensing, cataloging, IT, or all four. Data librarianship is one of the most collaborative positions in the library, reaching across departmental boundaries. Libraries are organized to handle books or other physical media and electronic resources with well-defined parameters, and so cataloging or circulation librarians know how to get those materials on the shelf or in the catalog. But digital materials such as data require much more in-depth collaborations almost every step of the way from acquisitions to access. UNC at Chapel Hill has worked to make the acquisition process for data more routine, but given the difficulties involved, it may never achieve the level of ease with data that exists for other types of materials.

Data librarians will need to continue to work hard to communicate vendors' expectations and rights to their users in a multitude of ways.

Although technology is advancing rapidly, making access to small data sets technically possible, data librarians still need a wide variety of avenues to communicate

with their patrons about how data sets may be used. Partly because library patrons are not accustomed to narrow restrictions on how they use materials, restrictions on applications of data resources are unfamiliar and unexpected. Data librarians will need to continue to work hard to communicate vendors' expectations and rights to their users in a multitude of ways.

Notices to users about the following license terms might become future best practices: publishing rights and levels one can report (noted by just one respondent in this survey); citation or copyright attribution (mentioned by two respondents); and prohibitions on depositing data (noted by no respondent). Although no respondent brought up this last issue, UNC at Chapel Hill does include this notice in catalog records for data sets because more journals are beginning to require authors to deposit data in support of articles submitted for publication, and the great majority of vendors do not allow such deposit. Publishing rights and reporting levels are somewhat similar to deposit prohibitions but are more specific about how much of the data may be reported in a publication. Data librarians at UNC at Chapel Hill have negotiated with vendors concerned about users publishing data to identify examples of acceptable reporting levels. In the long run, complete citation or copyright attribution may prove problematic if data sets are continually revised and added onto—for example, how could one realistically cite every single contributor? Nevertheless, most researchers seem to prefer that scholars who perform secondary data analyses acknowledge their original work. Consequently, whether consistent citation of all contributors over time develops will determine whether contributor citation actually becomes a best practice.

While these problems are common to many library materials, few categories of materials have all of these issues, as data sets do. Data librarians face special challenges in becoming knowledgeable in all of these areas. It is imperative that they develop good working relationships with intra-library colleagues who have expertise in these areas. Organizational structure, funding, technological structure, and other factors make resolving these issues highly situational to each library. Administrators should take these issues into account when they consider instituting new data librarian positions, both to involve these areas in planning data services and to prepare training and networking opportunities for a new data librarian.

Future studies of data acquisition could broaden the population to include qualitative data and digital humanities librarians. From a larger study, we could determine which issues with data acquisition are specific to a particular type of data and which are universal. This would give us a way to identify which issues would have the highest impact if they were addressed by the entire community of librarians who collect data and would give us more resources with which to start conversations with vendors.

Conclusion

Academic library patrons do not typically think of their libraries as a place that can help them with data. A few universities, including the University of California, Los Angeles (UCLA) and the University of Michigan, have high-profile data archives and related services that are not part of the library. Most campuses, however, have no central place where researchers can receive help with data. They must rely on the resources of their

individual departments for funding and expertise. This situation leads to data "deserts" in some departments. Libraries are an equalizing force in data collections and other types of digital services, just as they make tangible materials available campus-wide. In addition, libraries can serve to centralize purchasing so that multiple units need not each buy the same content. Nonetheless, one interviewee mentioned the difficulty of getting patrons to think about the library for more than books and journals, the marketing effort that libraries must perpetually make to let people know how many more things we offer. If we have data sets in the library that are useful for campus teaching and research, and if we can make them discoverable and accessible, data collections can be a gateway to other digital services in the library. Highlighting specific data sets that are available and training researchers and instructors in tools to manipulate data sets, to convert results to graphic or pictorial form, and to manage data in the long term will help demonstrate that library services go well beyond providing books and journals. Collecting data sets is not simple and raises different issues from other types of electronic resources, but in a data-centric research environment, these collections are an important way to serve our institutions.

Most campuses . . . have no central place where researchers can receive help with data. They must rely on the resources of their individual departments for funding and expertise. This situation leads to data "deserts" in some departments.

Karen Hogenboom is an associate professor of library administration and head of the Scholarly Commons in the University Library at the University of Illinois at Urbana-Champaign; she may be reached by e-mail at: hogenboo@illinois.edu.

Michele Hayslett is the librarian for numeric data services and data management in Digital Research Services of the University Libraries at the University of North Carolina at Chapel Hill; she may be reached by e-mail at: michele_hayslett@unc.edu.

Appendix

Appendix A – Survey (including skip logic)

Data Collection Practices in Academic Libraries

This research is being conducted by Michele Hayslett from the University of North Carolina at Chapel Hill and Karen Hogenboom from the University of Illinois at Urbana-Champaign. Thank you for taking our survey on library data acquisitions practices and policies. The survey asks about your library's collection development policies and acquisitions procedures for data sets and how you provide access to the data you acquire. Even if your library does not currently collect data, we would appreciate your input into our project. We hope to get a rich picture of the current state of data collection in U.S. and Canadian academic libraries. Completing the questionnaire involves no more than minimal risk. This means that you are unlikely to experience any harm or discomfort that are greater than those ordinarily encountered in daily life, or during the performance of routine physical or psychological examinations or tests. Completing the questionnaire is voluntary. You may stop at any time, and skip the questions you do not want to answer. The decision to participate, decline, or withdraw from participation will have no effect on you professionally or personally. It takes about 45 minutes to complete the questionnaire. Benefits of participation may include informing/improving your own professional practice by helping us identify coming issues and best practices for various aspects of data collection management. At the end of the survey you will have the opportunity to volunteer for an interview which will explore the answers to your survey questions in more detail. We may not interview everyone who volunteers. If you volunteer for an interview, your contact information will be associated with your survey responses until after our interviews but will be removed before we analyze the survey results. Interviews will be conducted by phone, and will take approximately 30 minutes. The interviews will be recorded and transcribed. Transcripts will be coded and analyzed in the aggregate. If any quotations are used from your interview, whether we attribute them to you or not, we will submit them to you for approval before submitting the manuscript to a publisher. By providing your email at the end of the survey, you certify that you are 18 years old or older and that you voluntarily agree to participate in the interview portion of the survey as described above. The results of the survey and interviews will be stored in an online folder only accessible to Michele Hayslett, Karen Hogenboom, and a student employee. All three individuals will be trained in human subjects research, and will keep the survey results confidential. If you have any questions about your rights as a participant in this study or any concerns or complaints, please contact the University of Illinois Institutional Review Board at 217-333-2670 or via email at irb@illinois.edu. You may also contact the University of North Carolina IRB at 919-966-3113 or via email at irb_questions@unc.edu. If you have any questions about the questionnaire, please contact Karen Hogenboom (hogenboo@illinois.edu).

By clicking "I Agree" below you acknowledge that you are at least 18 years old, that you have read and understand the above consent statement, and that you consent to participate in this study. Please print a copy of this statement for your records, if you so desire.

General Information

What is the highest degree your institution awards?
- Baccalaureate (1)
- Master's (2)
- Doctorate (3)

How many students does your library serve?
- Fewer than 10,000 students (1)
- 10,000 to 20,000 students (2)
- More than 20,000 students (3)

Data Acquisitions

Do you purchase numeric or GIS data? (Do not count large database sources such as Statista or ProQuest's Statistical Abstract of the United States.)
- Yes (1)
- No (2)

Answer If Do you purchase numeric or GIS data other than large database sources? (Large database sources m... No Is Selected

Why not?

If If no, why not? Is Displayed, Then Skip To End of Survey

Note: As with most library database acquisitions, most vendors license access to data rather than selling resources outright. In the following questions, we use the term "purchase" to mean both purchasing resources outright (if you do that) and licensing access to data.

About when did your library start purchasing data? Your best guess is fine.

What is your role (if any) in purchasing data?

In your library currently, are people other than you involved in purchasing data?
- Yes (1)
- No (2)

Answer If In your library currently, are people other than you involved in purchasing data? Yes Is Selected

Please explain your answer. We're particularly interested in the roles you all play in purchasing data, from selection to ordering to license review to any special work needed to make the data available to users.

Collection Development Policies

Do you have a formal collection development policy for data?
- Yes (1)
- No (2)

Answer If Do you have a formal collection development policy for data? Yes Is Selected
If you're willing to share it, please provide a URL if it is available online.
- Enter a URL (1) _____
- My policy is not available online. (2)

Answer If If yes and you're willing to share it, please provide a URL if it is available online. My policy is not available online. Is Selected
If you're willing to share it, please upload it here.

Answer If Do you have a formal collection development policy for data? No Is Selected
How do you decide what to collect?

Is a portion of your collections budget designated for purchasing data?
- Yes (1)
- No (2)

Answer If Is a portion of your collections budget designated for purchasing data? Yes Is Selected
What percentage of the total collection budget is set aside?

Answer If Is a portion of your collections budget designated for purchasing data? No Is Selected
What funds are used to purchase data?

How do you identify data for purchase?

What is the scope of your collection in terms of subject areas, number of data sets, or whatever other way makes sense to describe it?

Are there data types you do NOT collect? Examples of exclusions might include very large data sets, materials on certain media, etc.
- Yes, we have limitations on acquiring some data. (1)
- No, we have no limitations and will collect anything. (2)

Answer If Are there data types you do NOT collect? Examples of exclusions might include very large data sets, materials on certain media, etc. Yes, we have limitations on acquiring some data. Is Selected
What are your limitations?

Do you have a formal weeding policy?
- Yes (1)
- No (2)

Answer If Do you have a formal weeding policy? Yes Is Selected
If your weeding policy is not part of your formal collection development policy for data and you are willing to share it, please provide a URL if it is available online.
- Yes (1) _____
- My weeding policy for data is not available online. (2)

Answer If If it is not part of your formal collection development policy for data and you are willing to share it, please provide a URL if it is available online. My weeding policy for data is not available online. Is Selected
If you are willing to share your weeding policy for data, please upload it here.

Answer If Do you have a formal weeding policy? No Is Selected
In what circumstances do/might you de-accession data sets?

License Review

Do you feel you are more involved in license review for new data acquisitions than most subject librarians without specific responsibility for data are for resources in their respective subject areas?
- Yes (1)
- No (2)

Answer If Do you feel you are more involved in license review for new data acquisitions than most subject librarians? Yes Is Selected
In what way?

Do any of your vendors require that you inform users of any license terms or restrictions?
- Yes (1)
- No (2)

Answer If Do any of your vendors require that you inform users of any license terms or restrictions? Yes Is Selected
Please describe the terms or restrictions about which you must inform users.

Do you inform your users of any license restrictions aside from any vendor requirements?
- Yes, we inform users about ALL license terms. (1)
- Yes, we inform users about SOME license terms. (2)
- No, we inform users about no license terms aside from those vendors require. (3)

Answer If Do you inform your users of any license restrictions aside from any vendor requirements? Yes, we inform users about SOME license terms. Is Selected
What license terms/restrictions do you inform users about, aside from any vendor requirements?

Answer If How do you measure such usage? Click to write Choice 1 Is Selected Or How do you measure such usage? Click to write Choice 2 Is Selected
How do you make users aware of license terms? (Please select all that apply.)
- ❑ Pop-up or click-through license display (1)
- ❑ Notes in catalog records (2)
- ❑ Notes in subject or class pages (3)
- ❑ Notes on the download page or README files with the data (4)
- ❑ Other (please specify) (5) _____

Access to & Usage of Data Collections

Do you provide access to any data sets virtually, that is, both online and not requiring the patron to be physically in a library?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you provide access to any data sets virtually, that is, both online and not requiring the patron to be physically in a library? Yes Is Selected
Please describe your technical setup for providing this access. Broad terms are fine.

Do you have a sense of how often the data you collect are used?
- ❍ Yes, I know exactly how often the data I collect are used. (1)
- ❍ Yes, I have a general sense of how often the data I collect are used. (2)
- ❍ No, I have no sense of how often the data I collect are used. (3)

Answer If Do you have a sense of how often the data you collect are used? Yes, I know exactly how often the data I collect are used. Is Selected
If you have a way or ways to measure how often data collections are used, please describe it/them.

If you would be willing to be interviewed about your library's data collection practices, please provide your email address below. Interviews will last no more than 45 minutes.

Notes

1. Hilary M. Davis and John N. Vickery, "Datasets, a Shift in the Currency of Scholarly Communication: Implications for Library Collections and Acquisitions," *Serials Review* 33, 1 (2007): 26–32.
2. Lynn Woolfrey and Jane Fry, "Managing University Research Microdata Collections," *New Review of Academic Librarianship* 21, 2 (2015): 146–59.
3. Sue A. Dodd, *Cataloging Machine-Readable Data Files: An Interpretive Manual* (Chicago: American Library Association, 1982).
4. Margaret O'Neill Adams, "The Origins and Early Years of IASSIST [International Association for Social Science Information Services & Technology]," *IASSIST Quarterly* 30, 3 (2006): 5–14; David Nasatir, *Data Archives for the Social Sciences: Purposes, Operations and Problems*, Reports and Papers in the Social Sciences 26 (Paris: UNESCO, 1973), quoted in Libbie Stephenson, "Social Science Data Files and Bibliographic Control: Contributions Sue A. Dodd," *IASSIST Quarterly* 37, 1–4 (2013): 8.
5. Interuniversity Consortium for Political and Social Research (ICPSR), "ICPSR: The Founding and Early Years," accessed May 28, 2016, http://www.icpsr.umich.edu/icpsrweb/content/about/history/early-years.html.
6. John B. Lowe, "Gambling on CD-ROM," *Library Journal* 113, 12 (1988): 37–39; Cary Griffith, "What's Wrong with CD-ROM?" *Information Today* 9, 7 (1992): 16–18.
7. F. K. Rottman, "SilverPlatter: The History of a Company of a 'Higher Order,'" *Information Today* 8, 2 (1991): 22.
8. Erik W. Austin, "ICPSR: The 1980s," accessed May 29, 2016, http://www.icpsr.umich.edu/icpsrweb/content/membership/history/eighties.html.
9. Mary Vardigan, e-mail message to the authors, June 18, 2015.
10. Terrence B. Bennett and Shawn W. Nicholson, "Research Libraries: Connecting Users to Numeric and Spatial Resources," *Social Science Computer Review* 25, 3 (2007): 302–18.
11. One respondent did not finish the survey.
12. University Libraries, University of North Carolina at Greensboro (UNCG), "Lynda Kellam, Data Services & Government Information Librarian," accessed May 29, 2016, http://uncg.libguides.com/ld.php?content_id=4214146.
12. Michigan State University Libraries, "Collection Development Policy Statement: Data Services (Numeric Data)," accessed May 29, 2016, http://libguides.lib.msu.edu/dataservicescollectiondevpolicy.
13. University of Western Ontario, London, Ontario, "Western Libraries Collections Management Policy for Maps and Data," accessed May 29, 2016, http://www.lib.uwo.ca/files/policies/collections/mapanddata.pdf.