



Numeric Data Products and Services

A SPEC Kit compiled by

Michael N. Cook
Social Science Bibliographer
Mann Library
Cornell University

John J. Hernandez
Coordinator, U.S. & International Documents
Librarian for Politics & Public Administration
Elmer Holmes Bobst Library
New York University

Shawn Nicholson
Michigan Documents Librarian
Sociology and Social Work Bibliographer
Michigan State University Library

August 2001

Series Editor: Lee Anne George

SPEC Kits are published by the

Association of Research Libraries
OFFICE OF LEADERSHIP AND MANAGEMENT SERVICES
21 Dupont Circle, NW, Suite 800
Washington, D.C. 20036-1118
(202) 296-2296 Fax (202) 872-0884
<<http://www.arl.org/spec/index.html>>
<pubs@arl.org>

ISSN 0160 3582

Copyright © 2001

The papers in this compilation are copyrighted by the Association of Research Libraries. ARL grants blanket permission to reproduce and distribute copies of these works for nonprofit, educational, or library purposes, provided that copies are distributed at or below cost, and that ARL, the source, and copyright notice are included on each copy. This permission is in addition to rights of reproduction granted under Sections 107, 108, and other provisions of the U.S. Copyright Act.



The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

SPEC

SUPPORTING EFFECTIVE LIBRARY MANAGEMENT FOR OVER TWENTY YEARS

Committed to assisting research and academic libraries in the continuous improvement of management systems, OLMS has worked since 1970 to gather and disseminate the best practices for library needs. As part of its commitment, OLMS maintains an active publications program best known for its SPEC Kits. Through the OLMS Collaborative Research/Writing Program, librarians work with ARL staff to design SPEC surveys and write publications. Originally established as an information source for ARL member libraries, the SPEC series has grown to serve the needs of the library community worldwide.

WHAT ARE SPEC KITS?

Published six times per year, SPEC Kits contain the most valuable, up-to-date information on the latest issues of concern to libraries and librarians today. They are the result of a systematic survey of ARL member libraries on a particular topic related to current practice in the field. Each SPEC Kit contains an executive summary of the survey results (previously printed as the SPEC Flyer); survey questions with tallies and selected comments; the best representative documents from survey participants, such as policies, procedures, handbooks, guidelines, websites, records, brochures, and statements; and a selected reading list—both in print and online sources—containing the most current literature available on the topic for further study.

SUBSCRIBE TO SPEC

Subscribers tell us that the information contained in SPEC Kits is valuable to a variety of users, both inside and outside the library. SPEC purchasers use the documentation found in SPEC Kits as a point of departure for research and problem solving because they lend immediate authority to proposals and set standards for designing programs or writing procedure statements. SPEC Kits also function as an important reference tool for library administrators, staff, students, and professionals in allied disciplines who may not have access to this kind of information.

SPEC Kits can be ordered directly from the ARL Publications Distribution Center. To order, call (301) 362-8196, fax (301) 206-9789, email <pubs@arl.org>, or go to <<http://www.arl.org/pubscat/index.html>>.

Information on SPEC and other OLMS products and services can be found on the ARL website at <<http://www.arl.org/olms/infosvcs.html>>. The website for SPEC is <<http://www.arl.org/spec/index.html>>. The executive summary or flyer for each kit after December 1993 can be accessed free of charge at the SPEC website.



Kit 263

Numeric Data Products
and Services
August 2001

Survey

EXECUTIVE SUMMARY	9
SURVEY RESULTS	14
RESPONDING INSTITUTIONS.....	23

Representative Documents

COLLECTION POLICIES

University of British Columbia <i>Collection Development Policy: Data Services</i>	29
University of California, San Diego <i>Collection Development Statement for Machine Readable Data Files</i>	33
Harvard University <i>Collection Development Policy for Social Science Data Sets</i>	35
University of Pennsylvania <i>Collection Development Policy: Digital Information Resources</i>	38
University at Albany, State University of New York <i>Electronic Media</i>	41

USER GUIDES

University of British Columbia <i>Numeric Data Services at UBC: An Introductory Guide</i>	47
Colorado State University <i>Statistical Sources</i>	48

Duke University	
<i>Data Services at Perkins</i>	52
University of Massachusetts	
<i>Government CD-ROM User Guides</i>	53
University of Michigan	
<i>Numeric Data Services</i>	55
University of Virginia	
<i>Geostat User Guides</i>	57
Yale University	
<i>Help Using SSDA Data Files on the STATLAB Network</i>	59
DATA SERVICE DESCRIPTIONS	
Emory University	
<i>Electronic Data Center</i>	63
Harvard University	
<i>Harvard-MIT Data Center Home Page</i>	64
University of Manitoba	
<i>Data Library Services</i>	65
University of Michigan	
<i>Numeric Data Services. University Library</i>	66
University of Pennsylvania	
<i>Quantitative and Geospatial Data Holdings</i>	67
University of Tennessee	
<i>Data Services Policy</i>	69
University of Toronto	
<i>Data Library Service</i>	71
University of Virginia	
<i>Geospatial & Statistical Data Center</i>	75
University of Waterloo	

<i>Overview of TriUniversity Data Resources</i>	76
Yale University	
<i>The Social Science Data Archive at Yale</i>	77
JOB DESCRIPTIONS	
University of California, San Diego	
<i>75% Data Services Librarian</i>	81
University of Georgia	
<i>Data Services Librarian</i>	83
McMaster University	
<i>Digital Resources Specialist</i>	85
University of Virginia	
<i>Director, Geospatial and Statistical Data Center</i>	88
ACCESS POLICIES	
University of British Columbia	
<i>General Licence Agreement Conditions</i>	91
University of Pennsylvania	
<i>Access to Electronic Resources</i>	92
Washington State University	
<i>Holland Library Data Archive</i>	93
University of Waterloo	
<i>Data Access & Use Restrictions. TDR Licensing Agreements</i>	94
York University	
<i>Access Restrictions</i>	95

Selected Resources

BOOKS, JOURNAL ARTICLES, AND WEBSITES.....	99
--	----



SURVEY



Executive Summary

Introduction

As scholarship places increasing emphasis on quantitative analysis, the need for numeric data grows across academic disciplines. Students and faculty require better access to sources of numeric data ranging from simple descriptive statistics to large sets of raw data. A “numeric data product” may be defined as any information resource—print or nonprint—with considerable numeric content. This definition focuses on the content of numeric data products, irrespective of format. As is evident from the survey results, many perceive data as strictly limited to machine-readable files. Such a perception ignores the fact that data comes in a wide variety of formats.

Academic departments, campus computing centers, and libraries have all provided support to varying degrees for researchers using numeric data. While the demand grows, specialized data services are quickly becoming a major operation. On many campuses, the library has become a central access point for numeric data users by coordinating collections of data resources and services to researchers. As more academic libraries join the data game, they look to their peer institutions for models of service, collection development, access policies, budgets, and staffing patterns.

Academic libraries have historically collected and provided access to bibliographic sources that contain some amount of numeric data. Libraries that are also government document depositories have often served as gateways to numeric data, as many of the publications distributed through depository programs are statistical in nature. The emergence of major statistical indexes has helped libraries, both depository and nondepository, provide improved access to sources of numeric data. Such products as the *American Statistics Index*, published by the Congres-

sional Information Service, allow users to identify and access publications that contain substantial numeric content.

As with bibliographic resources, the demand among numeric data users for immediate, full-text access is on the rise. Since the release of 1990 Census data on CD-ROM, academic institutions have faced an increasing demand for numeric data, especially in a machine-readable form. Data consortia, such as the Inter-university Consortium for Political and Social Research (ICPSR), have also provided more widespread access to numeric data by archiving and distributing data files.

The survey results indicate that most academic libraries collect and service numeric data, but a disparity exists in how they define numeric data, what specific services are provided, what levels of staffing and budget are needed, and how they evaluate data collections and services. It is hoped that the survey results will help identify some best practices, assuage some fears, and foster collaboration among institutions and individuals working to provide numeric data services.

Survey

While many academic libraries collect and provide services for numeric data resources, little has been written about how such resources are selected or which specific services are provided. The investigators have attempted to find answers to several specific questions about numeric data in academic libraries. What relationships exist between libraries and other academic units that have responsibility for numeric data? What are the principle sources for numeric data? How much are libraries spending to acquire numeric data resources? How much staff time are libraries devoting to numeric data? How do libraries evaluate their own performance regard-

ing numeric data? The investigators designed their survey to try to capture some of this information.

During the spring of 2001, an eighteen-question survey was distributed to ARL member institutions in the United States and Canada. The survey was divided into six topical areas: Background, Numeric Data Product Collections, Access to Machine-readable Numeric Data Products, Staffing, Budget, and Evaluation of Numeric Data Collections & Services. In addition to basic information gathering about data collections and services, respondents were presented with an opportunity to rank their service levels based on desired and perceived performance. The investigators draw from the *LibQUAL+* methodology recently developed by ARL. Sixty-six of the 121 ARL member libraries (55%) responded to the survey.

Responsibility for Data

Sixty-four out of 66 libraries (97%) indicated they collect or provide access to numeric data. All of these respondents provide some assistance for users trying to locate numeric data. Apart from helping users locate data, the most frequently reported functions among libraries were: cataloging numeric data (98%), providing access to data (98%), collecting numeric data products (97%), storing numeric data (95%), answering reference questions about data (95%), and assisting users to retrieve data (92%).

The survey asked respondents to indicate which of the other campus units also perform various functions associated with data services. When it comes to assisting users with data analysis, about a third (31%) of libraries provide this service, while close to half of the respondents (45%) indicated that academic departments perform this function. This seems to correspond with the typical limitation many academic librarians place on reference service; namely, to help users get to information, but leave the interpretation of that information to the user. Other less frequently reported functions performed by libraries include: archiving data (53%), developing web-based tools to access numeric data (52%), and collecting codebooks (78%).

Among the other campus units, academic departments appear to provide a significant share of data services. Fifty-nine percent of respondents indicated

that departments also collect data products, 47% said they store data, 36% said they provide access to data, and 31% said they provide user instruction. Overall, libraries appear to take the lead in handling numeric data. It should be noted, however, that this trend in responses is very likely a result of the fact that all the organizations surveyed were libraries. If a similar survey had been administered to the other campus units, a similar skew would probably have resulted. Nevertheless, the trend is telling, as libraries often are not aware of what other units on campus are doing with regard to data. Libraries considering providing more data-related services would do well to investigate what the other units are doing, and develop partnerships with them.

While 48% of responding libraries indicated they have no formal relationship with another academic unit responsible for numeric data, over a third (39%) hold both data and codebooks that are used by the data unit, and 34% include the data unit as an administrative part of the library. Among the explanatory responses, the majority of those commenting indicated they had an informal but strong relationship to other units, whether or not they are within the library.

Data Collection Development

Obtaining information about the number of data products libraries are collecting proved to be very difficult. On average, only about a third of the respondents were able to provide any numbers in response to this question. An overwhelming number of respondents indicated that approximating the number of products collected is difficult at best and perhaps impossible, as data products are distributed in various forms and from various sources.

Of those that could respond (22 out of 64), about half (48%) indicated they receive between 100–500 data products through a government depository program, 13% said they got 3,000 or more. Two-thirds (67%) said they receive 50 or fewer products via international governmental organizations (such as the UN, IMF, and OECD). About half (48%) indicated they received 20 or fewer products from commercial sources. Given the difficulty in answering this question, perhaps some other method of estimating the

percentages of the library's total collection received from these various sources, rather than exact numbers, could shed better light.

The majority of respondents (89%) indicated their institutions were members of the ICPSR. Of those, 60% indicated that their official representative to the ICPSR was a member of the library staff. Over a quarter (27%) said their institutions participated with the Roper Center for Public Opinion Research. Other consortia included the European Consortium for Political Research, R-Cade, Statistics Canada's Data Liberation Initiative, Sociometrics' Electronic Social Science Data Library and statewide consortia.

Over a quarter (27%) of respondents have a collection development policy covering numeric data. A preponderance (98%) of libraries collects numeric data in the social sciences. Over three-quarters (77%) collect in the sciences, while a little over half (56%) collect numeric data in the humanities. Business was mentioned frequently as another distinct subject area, while others included environment, public health, agriculture and law.

With regard to formats in which data is collected, online (98%) and CD-ROM (97%) are in the majority, followed by paper (81%), diskette (75%), microform (70%), and even magnetic tape (23%). The only other format mentioned was DVD, which reflects an emerging format of choice. Some respondents also commented that data only comes electronically, while differentiating that from statistics, which are acquired in print, microform etc. This is an interesting distinction to make as it reflects a prevailing notion that data is something relegated only to the world of computers.

Machine-readable Data

Since machine-readable numeric data poses particular issues for access and bibliographic control, two questions were asked specifically concerning this format. Nearly all responding libraries (97%) provide cataloging of these numeric data resources in their online public-access catalogs, and just under a quarter (23%) have some other specialized catalog specifically for data. Many respondents indicated they rely on lists presented on web pages, and several respondents said they had some sort of catalog-

ing scheme in development.

Thirty-four percent of respondents indicated they have no formal policy with regard to patron access to machine-readable data. More than half (58%) said that access is freely available to the institution's community, both on and off campus. Only a third (33%) restricted access to students, faculty, and staff. Over half of the respondents noted that walk-in users have access to their resources, both locally mounted and remotely accessed. Since there is a high degree of variability among license agreements (especially with regard to commercially produced data resources) nearly every data source, in effect, has a different access policy. Government document resources were often singled out as having no access restrictions.

Staffing and Budgeting

Just over a third of the responding libraries (38%) reported they employ someone specifically to acquire numeric data products. Of those, 79% hold the M.L.S. degree or its equivalent. Close to half (48%) said they employ someone specifically to handle service requests for numeric data, 65% of whom hold the M.L.S degree or its equivalent. Among other credentials held by these individuals are Bachelor's and advanced degrees in political science, economics, and statistics.

The most commonly provided job title was Data Librarian. Judging from the responses that provided the same title for both acquisitions and reference service responsibility, 13 of the above individuals (54%) are responsible for both areas. Many respondents expressed concern over use of the term "specifically," reporting that the tasks of acquiring and providing service for numeric data are often distributed among several librarians and staff, without having any one individual specifically charged with this responsibility. On average, libraries reported that they devoted 2 full time equivalent (FTE) librarians, 1 FTE support staff, and 1 FTE student staff to these responsibilities.

The same difficulty in determining the number of data products in the collection reappears when trying to estimate a budget for data products. Many respondents indicated that it is difficult to derive a budgeted amount since numeric data is scattered throughout their collections. Only 27 respondents

out of 64 (42%) could provide some specific budget information. For those giving figures, the budgeted amounts range from zero to \$200,000. The average amount is \$31,825 with a median of \$12,000. The majority (89%) budget less than \$100,000 and 30% budget below \$2,000.

Evaluation of Data Services

Respondents were asked to indicate what techniques or measures they used to evaluate the quality of their data collections and services. Nearly half (45%) admitted they do not perform any formal evaluation. Of those that did perform some evaluation, most (69%) preferred usage statistics to measure the quality of their data services.

Working with the LibQUAL+ methodology, respondents were asked to rank their numeric data collections and services on a nine-point Likert scale, with 1 being the lowest level of service and 9 the highest. Seven aspects of data collections and services were addressed: convenient access to data, providing reference services, acquiring data products, creating partnerships with other data units, collecting user feedback, adding products on request, and maintaining error-free catalog records. For each category, respondents were asked to indicate what they thought should be a minimum service level, a desired service level, and the perceived performance level at their own library.

For the most part, respondents tended to rank the minimum level of service at around 4–6, the desired level at 7–8, and the perceived level of performance generally landed somewhere in between. A look at the difference between average rankings for each of the three service levels reveals some interesting patterns and could indicate how well libraries are performing. In the area of providing convenient access to data products, the average ranking for a minimum level of service is 5.32, the desired level is 8.23, and the perceived level is 5.80. The relatively high desired level of service would indicate that libraries consider convenient access to be important. Since the perceived level of performance is fairly close to the minimum level, one could say that libraries are providing convenient access at about the minimal level of service. In other words, there is some room

for improvement.

Similarly, looking at the difference between rankings across the various categories could offer some insight into how libraries value the different aspects of data collections and services. With regard to providing error-free catalog records, the average rank for minimal service is 6.17, the desired level is 8.55, and the perceived level is 6.60. Since the average minimal level and the average desired level are both higher for cataloging than they are for convenient access, one might infer that the libraries surveyed generally value accurate cataloging somewhat more highly than convenience. In fact, this seems consistent with the findings above that show most libraries indicate they catalog numeric data products and provide cataloging for machine-readable data as well.

With regard to collecting user feedback, the average minimum level is 4.56, the desired level is 7.54, and the perceived level is 4.00. These numbers would indicate that not only do the libraries seem to value collecting feedback less than they value cataloging, but also that they are not collecting feedback as well as they think they should. This seems entirely consistent with the earlier findings that show almost half of the libraries do not employ any specific measure to evaluate their collections and services. Likewise, in the area of creating partnerships, the minimum service level is 4.86, the desired is 7.80, and the perceived is 5.56. This would indicate that libraries may not value the creation of partnerships as highly as other aspects, and that they are a bit more satisfied with their performance in this area. Again this appears to be consistent with the findings above that show about half have no formal relationships with the other data units on campus, but seem to rely somewhat on informal arrangements.

Issues Raised

The results of this survey raise several important issues for libraries to consider while planning to create or expand numeric data collections and services. Among these are: creating a working definition of numeric data, determination of services to be provided, relationships with other campus data units, staffing and budget, access policies, and evaluation mechanisms.

Perhaps the most fundamental place to begin thinking about data collections and services is a decision on how the library will define what it means by "numeric data." Many respondents aired their views regarding this survey's definition of data. Some found it confusing while others found it out of conformity with traditional definitions. Interestingly, whereas some indicated that numeric data has never been thought of as a separate, distinct part of the collection, others pointed out that numeric collections are very different from traditional library collections. The result was that both groups found some questions difficult to answer. While the precise definition is certainly open to debate, the crucial matter is for libraries to have some understanding of what they will consider data to be, and to create or modify their collection development policies and service plans appropriately.

Libraries must decide for themselves what specific services they will provide with respect to data. Will they only provide reference assistance; limiting themselves to guiding users to sources of data and helping them copy or download? Will they take the next step and provide assistance in analyzing and presenting data in meaningful ways? Will they place data files on a server and create an interface for users to access the data? Will they only hold the data products and leave users to handle things on their own? Whatever services libraries decide to provide, they should have a firm sense of what these will be before proceeding.

Libraries should investigate what other units on campus are doing with regard to data. Has the academic computing department set up a data lab for students? Might someone there have expertise in data analysis to whom the library could refer users? Does the library have a subscription to a service that could help the data lab locate additional sources? Collaborative relationships between the library and other data units would go a long way towards providing more cohesive service to the university's researchers. They could also help spread the costs so these departments might not feel the pinch as hard as if they operated alone.

Staffing and budgeting issues must also be settled. Once libraries know what services they plan

to provide, they can determine what qualifications their staff would require to carry them out. Do they need to have a data librarian specifically charged to handle data, or could these functions be distributed among several librarians? Would they need to hire someone with a statistics background? The same is true for determining what kind of equipment the library would need and, consequently, how much they will need to spend. Will they need a new server, cabinets for CD-ROMs, extra terminals?

The library should decide what limitations they might place on access to data resources. Many licensing agreements require that access be limited only to university affiliates, while most government-issued resources are freely accessible. There are various, and sometimes conflicting, obligations imposed upon such resources that make creation of an access policy difficult. Will the library require a sign-in sheet for some resources and leave others open? Will they only provide web access to free resources while limiting use of commercial resources to on-site use only? At a time when the software industry is seeking stricter enforcement of their intellectual property rights, licensing restrictions on data products are becoming a very hot issue.

Finally, one issue that is often put on the back burner is evaluation. Many of the respondents do not have a mechanism for evaluating the quality of their collections or services, but having some method of evaluation is key for any future planning. Libraries might consider providing for an additional column on their reference statistics sheets to cover data-related questions. They could use server statistics to show how much usage their online data products are getting. A focus group could help show what other products or services users need. These techniques are also especially useful to aid libraries in deciding if they want to provide any specific data-related services at all.

Conclusion

Numeric data is quickly becoming an area for academic libraries to venture into, as more emphasis in scholarship is placed on quantitative analysis. As with other more traditional bibliographic resources, numeric data is available in a variety of formats,

each of which bringing its peculiar limitations and requirements. Libraries planning to enter the data game would do well to consider the issues raised by this survey as they proceed. A review of the accompanying documents provided by respondents may help others to identify some examples of good practices, as well as become more aware of the collections and services other institutions offer and some of the challenges they have encountered.