



**Disappearing disciplinary borders in the social sciences:  
Data acquisition – Access – Management**

**Jon Stiles**, Data Archivist for UC DATA, principal archive of social science data at University of California, Berkeley

**Libby Stephenson**, Data Archivist and Distinguished Librarian, University of California, Los Angeles

**International Federation of Library Associations, Social Science Libraries  
Section, Satellite Conference**

**Disappearing disciplinary borders in the social science library - global studies or  
sea change?**

**University of Toronto, Toronto, Canada  
6-7 August 2008**

**Sponsored by:**

**CQ Press  
Worldwide Universities Network (WUN)  
Center for Global Studies (CGS) at the University of Illinois  
University of Illinois Library  
University of Toronto Libraries  
Faculty Information Studies, University of Toronto**

**This work is licensed under the Creative Commons  
Attribution-Noncommercial-Share Alike 3.0 United States License. To  
view a copy of this license, visit  
<http://creativecommons.org/licenses/by-nc-sa/3.0/us/> or send a letter  
to Creative Commons, 171 Second Street, Suite 300, San Francisco,  
California, 94105, USA.**

**Abstract:**

Quantitative approaches in social science research often focus on a cross-national perspective requiring the use of data from many different nations. Such research also tends to bring together researchers from different disciplines; sociology, geography, political science, economics and so forth. Disciplines tend to emerge around or become defined by shared practices/methods/content specialization. Support of research through the provision of discovery, access, collection, preservation, analytic tools and collaborative environments by data archives and libraries mirror many aspects of the divisions of geography and academic discipline toward which that support is directed. However, the boundaries between disciplines and institutional support settings differ in permeability - the extent to which people, goods, methods, and concepts flow across unimpeded - and can strengthen, weaken or disappear over time as these flows transform the settings in which they operate.

In this paper we review the development of data services at UCLA and Berkeley over the last four decades and show how the provision of resources and services by libraries and data archives exemplifies this kind of blurring of disciplinary boundaries. Using the multiple-points of access model (Stephenson and Kasianovitz, 2007) as a starting point, we then discuss kinds of practices/methods that are successfully shared across boundaries of institutional form or mission. We will discuss the role of institutional cultures and how they help or hinder. The role of libraries and data archives as joint service providers will be illustrated using the example of operations at both UCLA and UC Berkeley. We will provide a set of guiding principles on data services, training,

collaboration and participation in the professional data community, and the use of commonly agreed upon metadata standards. Finally, we will show the benefit in using methods that might start in one organizational setting and move across to another in terms of responsibility for and management of collections and services.

## 1. Introduction

University-based data services units have historically depended on multiple units to fulfill their basic mission. With respect to machine-readable data for the social sciences, data archives, libraries and computers centers have formed a triangle of loosely linked support environments for those engaged in quantitative research. However, the extent of their involvement and the degree to which they have provided complementary, overlapping and collaborative services has varied greatly over time and across academic institutions. The changing roles played by these organizations have been shaped by a broad array of factors, including changing technology, needs for specialized expertise, fit with organizational missions, financial support, turf and politics, and relations with data producers.

During the 1960's, when the organization of the field was in its infancy, data support activities were centered in data archives and survey research centers, which provided for centralized acquisition and access, technical and analytic support, and storage and preservation, but which served a relatively small clientele. In the 1970's and 1980's computer centers began to provide the technology infrastructure in terms of hardware, statistical software and programming expertise, while libraries provided a basis for

maintaining bibliographic control and reference services with published statistics.

Research and survey units enabled production as well as assistance in using publicly available raw data and documentation. Despite the disentangling of many of these roles, most interaction with users of social science data continued to occur in data archives or data libraries, but not usually within an academic library. The increasing availability of statistical software, the targeting of distribution of electronic census products at selected depository libraries, the development and refinement of machine-readable cataloguing standards, expansion of large publisher/consortia archives and spread of desktop computing technology set the stage for the explosive growth of machine-readable data in the 1990's and confronted data archives and libraries with new questions about their roles. The release of the 1990 census products placed libraries in a pivotal role for distribution as the 1990 US Census of Population and Housing was released as part of the Federal Depository Library Program (FDLP) on CD-ROM. The rising availability and use of ftp signaled the disappearance of the need for local storage of many datasets. Increasingly, data users had unmediated access to data products, while archives and libraries struggled with questions about data preservation, appropriate supports for potential and actual data users, and possibilities for collaborative projects.

Over time, libraries and data archives each evolved into complementary, and often collaborative, operations. Data archives gradually built collections of studies, census data, vital statistics, and administrative records to support a quantitative approach to social science research using secondary analysis techniques. Libraries, in contrast,

acquired published statistics, often from the same organizations, in printed tabulations or aggregate presentations of data, and later obtained these materials on microfiche and CD-ROM. Computer centers decentralized with the advent of the PC and laptop computers. Hardware and software provided was discipline specific; many universities built social science computing facilities.

From the beginning, at UCLA and UC Berkeley, data services to faculty and graduate students were almost always provided through the data archives. Faculty and undergraduates who could best use published statistics, tended to go to the campus library. Throughout the 1960's and 1970's quantitative research was carried out at the campus computing center via punched cards and magnetic tape storage media. In the 1980's PC's and laptop computing allowed users and support units to move to decentralized non-main frame work spaces. Within data archives or libraries there was a fairly clear division in resources obtained, services provided and clientele served, until the Internet and the technology it afforded to users and providers, blurred the boundaries of collection building, information access and user support.

The evolution of the Internet and the capability to easily create, use and transmit digital information has changed the way libraries and data archives can think about providing user support. At the same time the nature of secondary analysis in a quantitative approach to research has not changed. It is still the case that using data requires an understanding of quantitative research methods, knowing how to use a statistical analysis software package, and an ability to interpret results. It is still necessary to

review data collection instruments (such as questionnaires), codebooks, and reports on sampling and weighting. And those who assist users need these same skills. On the other hand, there are several new web sites permitting users to make tables on the fly, there are online analysis tools, and government information web sites provide access to easily manipulated tables in database formats such as Excel. This presents opportunities and challenges to libraries and archives in terms of the collections and services we provide.

The Internet has also made it possible for data to be collected and distributed by many more producers than in the past. Countries which did not have the FDLP infrastructure in the 1980's and 1990's are now able to share their research and government information through web sites. Globalization has had an effect on the ability of researchers to connect with each other and share ideas, resulting in transformative cross disciplinary approaches incorporating ecological and social frameworks. Evolving political dimensions and changing national borders as well as formation of larger regional governing bodies such as the European Union have also influenced the scope and focus of quantitative research activities. The breadth and scope of available data is greater than ever before. No one unit, data archive or library can hope to address it all. New models of shared responsibility need to be developed incorporating the strengths and expertise of all concerned. This is particularly true if we are to address the needs of the scholar engaged in cross-national research.

In this paper, we will frame the roles of data support by discussing the evolution of data support and the factors which shaped that support on our campuses since the 1960's, consider areas of overlapping expertise in libraries and archives, and introduce the multiple-points of access model which is an approach to facilitating campus users access and use of data. We subsequently discuss potential barriers to cross-unit collaboration and challenges to maintaining complementary services.

## 2. Data Services and Support at UCLA and UC Berkeley

### Worlds Apart: the 1960's

At UCLA, early data archiving and support activities emerged around the UCLA Political Behavior Archive (PBA), which held "[o]fficial statistics, election returns, and survey data for the United States [and] biographical and survey data on political recruitment and career patterns from many countries."<sup>1</sup> The PBA, founded in 1961, distributed data from ICPR, of which it was an original member, the Roper Public Opinion Research Center, and locally acquired holdings. The focus of the archive was on easily acquired data in strong demand, and supported the development of "easy to use programs for manipulating and analyzing information using the computing facilities of the Western Data Processing of UCLA".<sup>2</sup> This work was led by UCLA political scientist, Dwaine Marvick. In his proposal for development of a "social science archival data library" Marvick outlined plans to "exploit an archive's diversified holdings, to perform a

---

<sup>1</sup> Glaser, William and Ralph Bisco, *Social Science Information*, 1966 (5) 71. p 74.

<sup>2</sup> Bisco, "Social Science Data Archives: A Review of Developments", *The American Political Science Review*, 1966, 50 (1), p. 104.

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

dynamically determined sequence of statistical manipulations, [and] to work with a tailor made set of matrices derived from various sources within the data depository.”<sup>3</sup>

A Survey Research Center was formed at UCLA in 1964 and it housed an Archival Data Library, with which the PBA merged and was “charged with the tasks of gathering and making available a wide range of quantitative data having to do especially with public policy and policy-related decision-making processes.” Neither the PBA nor the newly formed archival data library offered consultative services; access to study documentation was provided on a “reference-room basis.”<sup>4</sup> The Institute for Social Science Research (ISSR) was established in 1974, and the Survey Research Center and the Archival Data Library were incorporated into this new research support unit.

In 1962, the Survey Research Center at UC Berkeley in association with the Institute of International Studies formally established the International Data Library and Reference Service (IDL&RS). This facility was designed to assist social scientists in obtaining, processing, and analyzing existing domestic and foreign survey materials. The facility also provided assistance in the collection of new survey data in this country and abroad, and provided facilities as a training laboratory for University courses of instruction.

Initially, the IDL&RS served primarily as a reference service, acquiring materials only when expressly requested to do so. However, under a grant from the National Science Foundation, the Library created an archive of survey materials collected from

---

<sup>3</sup> Marvick, Dwaine. Plans for the Development of a Social Science Archival Data Library in the UCLA Survey Research Center. Unpublished proposal. August 1967.

<sup>4</sup> Ibid.

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

developing nations of the world, and, more generally, obtained other materials related to international, comparative and area research. Several projects were included in this program in 1964-65. First, a continuing search was undertaken for existing data held by American and overseas research groups. Second, a survey was begun of overseas agencies in an attempt to ascertain their research capabilities. Third, a general review was initiated of the methodological and technical problems encountered in conducting survey research in developing countries. And fourth, procedures were explored to enable the Library to store, classify, search, and retrieve such information by modern data processing methods.

At the time when IDL&RS and PBA were established, data archives were in their infancy, and the network of peer institutions was still in formation. The number of data archives expanded rapidly during the 1960's<sup>5</sup>, driven by a number of factors. On the supply side, the expansion of survey data collection efforts by commercial market and survey organizations and university-based social research institutions, in combination with the increasing availability of machine-readable data released by governmental bodies<sup>6</sup>, resulted in a growing accumulation of data for potential secondary analysis. This growth was matched by increased demand from researchers and teachers, particularly those interested in comparative research. Finally, technological developments provided the handmaiden - computers - which enabled the expansion of

---

<sup>5</sup> In a 1967 compilation of social science data archives, of the 25 data archives which made up the membership of the CSSDA, 19 were started after 1960. (In Howard White(1974), p. 9).

<sup>6</sup> For example, the first release of Census microdata - the 1-1000 PUMS from the 1960 census - occurred in 1962. International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

data collection, its accumulation in a more readily accessible format, and secondary analysis.

Libraries were not unaware of the developments of data services and archives such as the IDL&RS and the PBA. At UCLA the Library had received funding from the National Science Foundation for a Center for Information Services (CIS) devoted to aiding in the development of expertise in managing “machine readable databases, whether bibliographical, numerical or full-text.”<sup>7</sup> “In 1964, the University [of California] authorized the establishment of the [Institute of Library Research](#) [ILR] ... in cooperation with the School of Librarianship [at Berkeley] and the School of Library Service at Los Angeles”<sup>8</sup> Along with the newly formed UCLA Archival Data Library, the ILR held a symposium in mid-January 1967 in order to “explore the library’s role in meeting the future demands of social scientists for quantitative information services. Should a modern library think largely in terms of reference services, bibliographic searches and documentation retrieval ...or must its staff also become familiar with the ways in which data are generated, stored, manipulated, transformed and analyzed by social science users?” Their “conclusions were that a new kind of information specialist –part librarian, part programmer, part social scientist – would be needed to staff these new ‘quantitative libraries’”<sup>9</sup> Nevertheless, although they had taken note of the emergence of this new form of information –social science data—libraries of the 1960’s did not maintain or

---

<sup>7</sup> Rowe, Judith and Mary Ryan, "Library Service from Numerical Data Bases: The 1970 Census as a Paradigm" in *Reader in Machine-Readable Social Data*, Howard White, ed., Information Handling Services, 1977. p. 199.

<sup>8</sup> See [http://sunsite.berkeley.edu/~ucalhist/general\\_history/campuses/ucb/colleges.html](http://sunsite.berkeley.edu/~ucalhist/general_history/campuses/ucb/colleges.html) University of California History Digital Archives, drawn from *Verne Stadtman's Centennial Record* (CR)

<sup>9</sup>Marvick, Dwaine. Plans for the Development of a Social Science Archival Data Library in the UCLA Survey Research Center. Unpublished proposal. August 1967.  
International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

support access to collections of social science data. Despite this single joint effort on the part of the ILR and the PBA to discuss common themes and issues, further collaboration between libraries and data archives did not take place for some time. The need for widely divergent kinds of expertise and non-traditional methods of acquisition, and the lack of computing and data processing facilities hampered libraries in their efforts to serve the social science community.

To the extent to which the boundaries between institutions with similar purposes are marked by differing technologies and tasks, a lack of circulation of practitioners and professional staff, and discrepant perceptions of critical problems and issues, the inclusion of "Library" in IDL&RS's name masked a wide gap between the campus library and IDL&RS. A shared commonality of interest in the acquisition, indexing, preservation and dissemination of materials for use of the academic community did not span the divide created by widely different media, sharply different skill sets needed for the processing and use of the media, concerns about the methods of production of content, and the nature of relationships with the direct producers. Although many functional roles were common to archives and research libraries, such as acquisition, dissemination of documentation, storage and maintenance, inventory, or retrieval, but took on unconventional dimensions in the data archive setting.<sup>10</sup>

---

<sup>10</sup> A summary of the functional roles of data archives by David Nasitir, the Director of IDL&RS, can be found in "Data Archives for the Social Sciences: Purposes, Operations, and Problems", *UNESCO Reports and Papers in the Social Sciences*, No. 26 (1973). A sense of how very different one role - that of acquisition - was for archives and libraries are detailed in a separate paper, "Stalking the Wild Data Set: The Acquisition of Machine-Readable Data Files at Home and Abroad", *Drexel Library Quarterly*, Vol. 13(1), January 1977. Howard White's dissertation, *Social Science Data Sets: A Study for Librarians*, (1974), uncovers the great dissimilarities in practice and capacity which underlie surface similarities in mission between archives and libraries of that period. International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

Acquisition, for example, relied on close ties to providers via personal relationships in the research community, frequently entailed long negotiations and foreign travel, knowledge of survey and sampling methods to evaluate the quality of the data collection, and technical expertise in computers and varied media. For international collections, media standards varied widely, and translation to a consistent format might entail copying of tapes with different densities or parities, or require the design of machines to translate information "from small punch cards with round holes to larger ones with rectangular holes". Similarly, retrieval entailed substantially different skills and infrastructure (including programming skills and mainframe computers) than did print materials, where access did not need to be mediated by machines, or from microfilm, where the technology was straightforward and the output relatively interpretable and unambiguous.

Other tasks had no counterpart in the world of the library. Once acquired, the data required cleaning in the form of reformatting, transformation, and consistency and completeness checks; creation of standards and conventions for treatment of missing data or measures comparable across different data collection efforts; or data processing and analysis, which might simply reflect providing users with data in the format and media which their local institution can support (punch cards, 7 or 9 track tapes, varying density or parities) or actual analytic work.

Initial archives grew up around content specialties, and were marked by the establishment of dual relationships: one with data producers and the second with

academics with strong substantive and methodological interest and knowledge about the production of the data. With the expansion of interest in secondary data and growth in the number of repositories, the relationship between repositories became key, as well. The Council of Social Science Data Archives grew out of a need to address that third set of relationships. It began as the "committee of eight" - a group of representatives of universities at Berkeley, Yale, Los Angeles, Chicago, North Carolina, Cologne, and Michigan, representing both ICPR and non-consortium members. With funding from the NSF, CSSDA soon incorporated involvement from twenty-five representatives from social science data archives (including two librarians). The CSSDA was active throughout the 1960s, but dissolved in 1970 as funding expired.

In short, the formation of data support services during this initial period, prompted by increases in data collection and technological change, coalesced around institutions deeply embedded in the data collection and research process which possessed the technical expertise, content knowledge, motivation and capacity to support those services for a fairly specialized clientele. These institutions formed complementary relationships with one another that largely excluded libraries, which lacked the expertise, familiarity and capacity to incorporate and support this type of holdings.

### The Thin End of the Wedge?

Although libraries were conspicuously absent in the provision and support of machine readable data in the 1960's, the release of the 1970 census summary data to selected

government depository libraries promised to change that. One of the activities of the NSF funded CIS project included the acquisition of and “practical experience with numerical processing by ‘experimenting’ with the tapes from the 1970 Census.”<sup>11</sup> David Nasitir suggested that if the release of the machine readable data from the 1970 census summary census tapes “form the thin end of the wedge [in academic libraries’ collections], a large number of sample survey tapes now held in archives may follow.”<sup>12</sup>

At Berkeley, a Census Service Facility was established as a cooperative effort between the Institute of Governmental Studies and the Survey Research Center, and provided a large set of standardized tabulation as well as performing customized tabulations to meet individual user requests. At UCLA, the 1970 census STFs marked the library's first foray into numeric data. They purchased the California tapes, programs and associated documentation from DUALabs, and placed those with the Campus Computing Network (CCN). Orientation and reference services were offered through the Public Affairs Service (PAS), and the PAS acted as a gatekeeper for access to the files held by CCN. No data processing and manipulation, printing, or customized services were offered; it was a service designed for “do-it-yourselfers”<sup>13</sup>.

These efforts met with varying degrees of success. At Berkeley, products included both broadly disseminated standardized publication and access to customized analyses and

---

<sup>11</sup> Carmichael, Robert L. Center for Information Services, Phase II: Detailed System Design and Programming, Part 4 – Development Scheduling and Planning, Phase IIA Final Report. Los Angeles: University of California, Los Angeles. 1971.

<sup>12</sup> Nasitir, David “Data Archives for the Social Sciences: Purposes, Operations, and Problems”, *UNESCO Reports and Papers in the Social Sciences*, No. 26 (1973).

<sup>13</sup> Rowe, Judith and Mary Ryan, “Library Service from Numerical Data Bases: The 1970 Census as a Paradigm” in *Reader in Machine-Readable Social Data*, Howard White, ed., Information Handling Services, 1977. International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

products. At UCLA, success was more muted. The UCLA Library's Public Affairs Service director, Mary Ryan was a key player in the campus acquisition of the DUALabs format of the 1970 US Census tapes. The service was intended to provide a limited amount of consultation and was meant to provide a general reference and orientation. Although UCLA was well aware of the extent of service and support provided at UC Berkeley, it was thought that duplicating this for UCLA researchers would be "needless and wasteful."<sup>14</sup> By 1980, Marilyn Nasitir noted UCLA's PAS "offers such a service [selection and acquisition of data files and codebooks, statistical analysis, and consultation and reference] only on a limited basis, restricted to parts of the 1970 census...for just a few hours per week", with services hampered by a wide physical separation between the library and both the computer center and ISSR<sup>15</sup>.

Factors external to the Library were also at play. As is true in any large complex organization, UCLA was not immune to the vagaries of turf, power and politics. The ISSR had at its helm a strong leader (Howard Freeman) with responsibility for ensuring the longevity of the Institute. Faculty support of ISSR depended in part on their satisfaction with research support, including access to and support in using raw data. Campus computing was also on the edge of major change. Formats for media storage made it possible to manage larger and larger file sizes, the development of easier to use statistical analysis packages and the need for staff to assist users of these statistical routines were factors so far outside the scope of the Library it is not surprising

---

<sup>14</sup> Rowe, Judith and Mary Ryan, "Library Service from Numerical Data Bases: The 1970 Census as a Paradigm" in *Reader in Machine-Readable Social Data*, Howard White, ed., Information Handling Services, 1977.

<sup>15</sup> Nasitir, Marilyn. "The Cataloging and Classification of Machine-Readable Data Files", *Cataloging and Classification Quarterly*, Vol 1(1), Fall 1980. The Haworth Press.  
International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

that the Library made the decision not to purchase any additional numeric data for the 1980 census.

Although data access and support met with less success in the library setting, substantial progress was made with respect to the traditionally library-based bibliographic mission and interest in machine readable data grew in library circles. Dodd (1977) notes that "[i]n June 1970, ALA/RTSD/CCS Descriptive Cataloguing Committee established an ad hoc subcommittee to study the bibliographic characteristics of machine readable data files and recommend methods of describing data files..."<sup>16</sup> and Rowe (1974) notes both the presence of the Machine Readable Data Files Interest Group for ALA's GODORT and the consistently high attendance at ALA sessions on data resources since 1972<sup>17</sup>. Other groups also recognized the need to persistently identify specific studies by incorporating a notion of authorship, publication and edition, through the use of the terms more commonly associated with data files (principal investigator, data producer, and version) and to determine the best sources for this information. These included the National Bureau of Economic Research Conference on the Documentation of Large Data Files, held in 1974, National Science Foundation's funding of the 1978 Airlie House Conference on Cataloging and Information Services for Machine Readable Data Files and the International Association for Social Science Information Services and Technology (IASSIST) first meeting in 1974, held in conjunction with the 8th World Congress of Sociology. The Anglo-American Cataloging

---

<sup>16</sup> Citation needed

<sup>17</sup> Rowe, Judith. "Facilitating Information Access: Interaction Between System Components (The Data Library and the Traditional Library)", *SIGSOC Bulletin*, V.6 (2,3), Fall/Winter 1974. International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

Rules Second Edition (AACR2), published in 1978 contains a section (Chapter 9) for describing data files and documentation. As Rowe noted “the format of AACR2 precluded the inclusion of the background materials necessary to understand the nature of MRDF.”, thus the production of Dodd’s Working Manual for Cataloguing Machine-Readable Data Files in 1976 and publication of her seminal work Cataloging Machine-Readable Data Files in 1982, provided clear directions, examples and information of use to data archivists, data producers and library catalogers for establishing bibliographic identity for studies produced by social scientists and statistical agencies.<sup>18</sup>

Machine readable data remained firmly seated in the data archive at Berkeley, and at UCLA, however the cataloguing and inclusion of their archive's data holdings in their respective library's catalog system began in the mid-1970s. As more data holdings and acquisitions become standardized, the archive also shifted from academic staff to professional staff. Increasingly data were acquired and shared via long-term formalized relationship with other archives or producers, and support for data users entailed local storage and preservation of data through the computer center and assistance with data identification, processing, and analysis. Growth in services and holdings partially continued, however, to rely on the development of contacts with data producers, as with the merging of IDL&RS holdings with a collection with a California focus (and a subsequent name change for the archive to the State Data Program) and the designation of the SDP as a member of the newly formed Census State Data Centers network in 1979.

---

<sup>18</sup> Dodd, Sue A. Cataloging Machine-Readable Data Files: An Interpretive Manual. Chicago: American Library Association, 1982.  
International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

By the early 1980's, the roles of the Berkeley library, data archive and computer center had assumed the configuration they would hold for the next decade. The data archive provided assistance in finding, acquiring and preserving secondary data for academic users, focusing on collections acquired from the Census Bureau, other data archives, and producers, held documentation and codebooks for its' holdings, and helped users understand the layout and content of the collection . The library catalogued the new holdings on an ongoing basis as they were acquired, and held documentation for widely used sources. The computer center held the currently active portion of holdings on tape, with non-active holdings and backup copies retained by the archive, and provided the necessary hardware and software needed to read and analyze the data, as well as the technical staff needed to support that infrastructure.

Despite the work done by Marvick and others at UCLA, using data in the computing and statistical environment of the late 1970's was not for the faint of heart. Although there was a significant survey research program generating data files, the management of the data received less emphasis for several years. The UCLA ISSR posted a job opening for a Data Archivist in 1977 and asked applicants to possess a degree in library science as well as background in one of the social sciences. This step came about in part because the social science faculty was increasingly frustrated with the complexities of using US Census and other data at UCLA. The proposed data archivist was to provide assistance to users of survey materials, in copying tapes and in the understanding of study documentation. The newly named Social Science Data Archive housed within the

ISSR was to function as a service unit supporting quantitative research, primarily secondary analysis, and to maintain a collection of social science data files and their documentation, coordinate the acquisition of additional data from a variety of sources, and provide access to publicly available data. Coordination with the technical and statistical consulting service at the main computing center was to be established. This step proposed by then Associate Director, Donald Treiman formalized the “archival data library” into the unit it is today, within the ISSR. When the Library made a decision not to acquire the 1980 US Census tapes, the ISSR agreed to do so; the 1970 Census tapes were transferred from the Library to the ISSR archive which agreed to provide all support and storage for the materials.

As data support services matured in the 1970's, separate but complementary roles for the library and data archives evolved on our campuses. The archives gradually built numeric collections of studies, census data, vital statistics, and administrative records to support a quantitative approach to social science research using secondary analysis techniques. Libraries acquired published statistics, often from the same organizations, in printed tabulations or aggregate presentations of data, provided cataloguing of data and, to a lesser extent, access to documentation about archive collections. Although domains of expertise differed, archive staff was increasingly drawn from pools of professionals, rather than academics, much like the libraries. These campus roles were embedded in larger domains and relationships, with professional library associations pursuing missions in documentation and bibliographic standards, archival involvement in data consortiums and new professional organizations like IASSIST, and the

involvement of both in dissemination networks for the Census and other Federal sources of data and statistics.

### Collaboration and Partnership: the 1990's and Beyond

In the early 1990's, the SDP at Berkeley was renamed the University of California Data Archive & Technical Assistance (UCDATA), and expanded its mission to include roles in research and data creation. In a partnership with the California Department of Social Services, UCDATA acted in both research and archive capacity in doing welfare policy research, laying groundwork for academic-administrative collaboration. Much of the work performed by UCDATA in this partnership lay in the collection, harmonization, anonymization, and dissemination as public use files data from administrative record systems from state and county agencies, and the linking of these data to survey data collected from participants in public assistance programs.

During the same period, and building on earlier work creating the Socio-Economic Environmental Demographic Information System (SEEDIS) and the Populations at Risk to Environmental Pollution (PAREP) project, Lawrence Berkeley Nuclear Laboratories began providing Internet access to 1990 Census products on CD-ROM. As a result of new Census dissemination of data on CDROM and the UCB Library's role in the Federal Depository Library Program, much of the data was provided by the Library, with additional juke boxes, CDs, and user support provided by UCDATA. In addition to access via anonymous ftp and NFS mounts, a MS-DOS menuing system was added,

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

and the 1990 LOOKUP software, subsequently ported to Census Bureau servers, was created to provide simple and powerful access capabilities to internet users.<sup>19</sup>

The relationships and circulation of staff associated with that project persisted after LBNL funding was discontinued. The jukeboxes, data, and applications were ported to Library servers and UC DATA facilities, and the historic data associated with the LBNL projects transferred to UC DATA. Subsequent collaborative efforts with the Library included the creation of the Social Science & Government Data Library (SSGDL), which provided software to access the 1990 Subject Summary Tape Files (SSTFs) and historic census data from SEEDIS.

The collaborative projects encompassed partners beyond the Library, as well, relying on linkages between UCLA and Berkeley, and with the CSM program at Berkeley to place collections of the Field Polls from 1956 to the present in SDA format and make them available for online access and analysis. Current partners also include the Census Bureau, via the provision access to non-public federal data under conditions that facilitate innovative research while safeguarding confidentiality in the secure Census Research Data Centers.

In recent years, the Library at Berkeley has pursued its own expanded role in data acquisition and support, establishing a Data Lab in the Library and pursuing an active

---

<sup>19</sup> See Merrill, Deane et al. "1990 U.S. Census LOOKUP: Mining a Mountain of Data" (<http://www.dlib.org/dlib/march96/merrill/03merrill.html>) and Merrill, Deane et al. "The University of California CD-ROM Information System" (<http://portal.acm.org/citation.cfm?id=205323.205334>) . International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

acquisition and support role in its own facilities. UC DATA and the Data Lab support users individually, but coordinate in development strategies and sessions and workshops designed to train librarians in supporting users of numeric data resources.

While the UC Berkeley archive ranged well beyond campus borders in its mission, the UCLA ISSR Social Science Data Archive became the de facto campus-wide data service unit serving the UCLA research community. Faculty conducting social research via quantitative methods came not only from social science departments but also from such diverse disciplines as public health, education, policy studies, law and management. Surveys conducted by the ISSR Survey Research Center were deposited with the archive. A key activity of the Archive was to support faculty and student researchers engaged in secondary analysis of comparative and cross-national studies, which was a well-established research approach at UCLA. In order to aid in identifying and evaluating potentially useful data and to support research using these materials, a collection of studies useful for cross-national work was built, resulting in the Intergenerational Stratification and Mobility File (ISMF).<sup>20</sup> In this endeavor, partnerships were formed with other data service units at universities in the US and Europe largely through membership in the Inter-university Consortium for Political and Social Research (ICPSR) and through the International Association for Social Science Information Services and Technology (IASSIST). An index to the ISMF was produced, initially in print format and is now a web based resource found at

---

<sup>20</sup>The work on the Cross-National Intergenerational Stratification and Mobility Index (ISMF) was undertaken by Harry B.G. Ganzeboom (Free University, Amsterdam), Donald J. Treiman (University of California Los Angeles) and Libbie Stephenson (University of California Los Angeles), with funding from the Netherlands Organisation for Scientific Research (NWO), Project H50.293  
International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

<http://www.sscnet.ucla.edu/issr/da/Mobility/mobindex.html>. Building the ISMF required using some professional librarianship techniques in identifying and evaluating resources, extensive search strategies, thesaurus development and indexing. Expertise from a data archive perspective included comparing raw data file content to documentation and assessing the long term preservation needs of the digital files and determining which files would or would not be physically acquired vs. referring the user to the originating archive.

In the 1990's the UCLA Library began to take a stronger interest in providing access to data as a new cohort of librarians was hired. Library information technology moved from simply providing a backbone to the library catalog and moved into tool development and direct information delivery, and library administrators revised the mission of the library based on the potential of the internet with respect to cataloging and metadata, collections, reference services and information access. Beginning with increasing amounts of government information being supplied as both published statistics and as data files and then with development of web based data access and manipulation tools, the opportunities for collaboration were evident. This became especially important as the volume of data available increased almost exponentially. Joint collection policies were developed along with some reference service guidelines. For the most part the roles for each unit are delineated by the format of the data; the ISSR archive supports and maintains raw data while Library efforts involve materials with value-added tools and published statistics in either print or electronic format. Joint work in outreach to faculty and students and in developing information and statistical literacy has become

well established. Librarians have attended training courses in establishing data services offered through the ICPSR Summer Program.<sup>21</sup>

While the 1960's were marked by domination of data support by archives, and the 1970's and 1980's saw the emergence of separate but complementary roles for archives and libraries, the period after 1990 has seen the increasing growth of collaborative efforts and joint projects, with each institution contributing resources and expertise. Staff increasingly co-mingle and circulate, and provide over-lapping services to over-lapping clientele. In the following sections we turn to ways in which these institutions support one another, can enhance one another's services, as well as considerations and obstacles to this mode of mutual support.

### 3. Collections and support in libraries and archives

There are two key overlapping areas of expertise which both libraries and archives can each offer; assistance in finding data for specific research needs and help to understand content and format of specific data. And, archives and libraries can also collaborate in obtaining data sets by sharing resources as well as building specialized collections.

Strategies for collaborative collections include drafting joint acquisition policies, and committing funding for specific purposes and for serendipitous purchases. Important considerations include:

---

<sup>21</sup> See <http://www.icpsr.umich.edu/cocoon/sumprog/course/0063.xml>  
International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

- Goals of the institution as a whole
- Research concentrations of faculty at the institution
- Technological capacity in terms of computing and software
- Curation and preservation requirements
- Expertise required for assisting users
- Access requirements –single user, multiple user, etc
- License and contractual conditions
- Cost –immediate and long term

New data useful in secondary analysis is made publicly available every year. While there are no accurate figures on the number of new studies released, Ekkehard Mochmann writes that “the annual growth of data archives varies between countries; the smaller archives grow with some 30-50 datasets per year, the larger ones have an annual growth of more than 300. Still the numbers represent only a small proportion of the potentially available datasets.”<sup>22</sup> Identifying available data for specific research projects is a skill librarians and archivists each can do and offer in their institutional settings. For example, ISMF data were acquired from many different archives and countries, all of whom shared their materials and expertise. When archivists knew of UCLA’s interest in building a collection of stratification and mobility studies they advertised and sent email when new materials were available.

---

<sup>22</sup> <sup>22</sup> Mochmann, Ekkehard and Paul de Guchteneire. *The Social Science Archive Step by Step*. Cologne: International Federation of Data Organizations, 2005. Accessed May 2, 2008. [http://www.ifdo.org/data/data\\_archive\\_workflow01.html](http://www.ifdo.org/data/data_archive_workflow01.html).  
International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

In most academic libraries, librarians have subject area responsibility such as government information, international, economics, education and so on. A collaborative approach to collection management can function between libraries and archives by sharing information about new materials, discussing the needs of faculty with which each interacts, and having a continuing dialog about the scope and breadth of subject area concentrations.

#### 4. User support and the multiple-points of access model

Using any information source requires some assessment of how useful it will actually be for a specific purpose. In the case of using social science data, this assessment first involves a review of the data collection instrument, and the actual item by item content of the data file itself. Users also need to review the literature based on the study. Once work with data begins users usually have technical and statistical questions. "The level of service [a] unit provides will depend on a number of factors: mission, user's needs, staff knowledge and training, access to technology (software, hardware, and support), support of administration, funds for collections and services, availability of services from other units on campus - and ... ability to collaborate with them." (Read, p. 63)<sup>23</sup> Jim Jacobs (1991) has suggested the range of services needed:

- Data file identification service
- Basic data file recommendation service
- Advanced data file recommendation service

---

<sup>23</sup> Read, EJ. (2007). Data Services in Academic Libraries Assessing Needs and Promoting Services. Reference & User Services Quarterly, 46 (3), 61-75. (See pp.62-63)

- Data file use advisory service
- Data extraction services<sup>24</sup>

There is no one model for data service delivery. Libraries and/or archives can implement services and develop collections according to the needs described above. For further examples, the Association of Research Libraries' SPEC Kit *Numeric Data Products and Service* provides examples from a number of North American libraries relating to collection policies, data services and access policies.<sup>25</sup>

There are a number of considerations in planning data services. These include:<sup>1</sup>

### Planning Data Services

Criteria	Data Archives	Libraries	Other units
<b>Institutional resources</b> - identify the potential players among campus units, including libraries, archives, survey centers, instructional technology, data centers, research units.	<ul style="list-style-type: none"> <li>-Archivists</li> <li>-Technical staff</li> </ul>	<ul style="list-style-type: none"> <li>-Ref Librarians</li> <li>-Library IT</li> </ul>	<ul style="list-style-type: none"> <li>-Survey Centers</li> <li>-Instructional Technology</li> <li>-IT</li> <li>-Research units</li> </ul>
<b>Clientele</b> - - includes students (upper and lower division undergraduates, and graduate students), faculty (social sciences, other departments and	<ul style="list-style-type: none"> <li>-Graduate students</li> <li>-Faculty</li> <li>-Campus wide</li> </ul>	<ul style="list-style-type: none"> <li>-Under grads</li> <li>-Graduate students</li> </ul>	<ul style="list-style-type: none"> <li>-Researchers</li> <li>-High end computing</li> </ul>

<sup>24</sup> Jacobs, JR. (1991). Providing Data Services for Machine Readable Information in an Academic Library: Some Levels of Service. *Public-Access Computer Systems Review*, 2 (1), 145, 154-157.

<sup>25</sup> Cook, Michael N. , John J. Hernandez and Shawn Nicholson. SPEC Kit on Numeric Data Products and Services. No. 263. Washington, DC: Association of Research Libraries. 2001. <http://www.arl.org/bm~doc/spec263web.pdf>

schools) and campus units (computing centers, instructional technology, administration)		<ul style="list-style-type: none"> <li>·Faculty</li> <li>·Campus wide</li> </ul>	<ul style="list-style-type: none"> <li>·Statistical support</li> </ul>
<b>Staff and expertise available</b> - includes archivists, librarians, technical staff, statistical consultants	<ul style="list-style-type: none"> <li>·Archivists</li> <li>·Technical staff</li> </ul>	<ul style="list-style-type: none"> <li>·Ref Librarians</li> <li>·Library IT</li> </ul>	<ul style="list-style-type: none"> <li>·Statistical consultants</li> </ul>
<b>Types of services</b> to be provided	<ul style="list-style-type: none"> <li>·Find and use raw data</li> <li>·Understand quantitative methods</li> <li>·Use of statistical procedures</li> <li>·In-depth understanding of data content/format</li> <li>·Data extraction and manipulation</li> </ul>	<ul style="list-style-type: none"> <li>·Use of online tools</li> <li>·Use of published statistics</li> <li>·Find and use articles, monographs and government information</li> </ul>	<ul style="list-style-type: none"> <li>·Statistical techniques and use of statistical packages</li> <li>·Programming assistance</li> </ul>
<b>Management and administration</b> of services - physical and virtual locations, subscriptions and licenses, areas of overlap, assignment of responsibility	<ul style="list-style-type: none"> <li>·Usually located in a research unit</li> <li>·Work with contracts and grants units</li> <li>·Work with local IRB</li> </ul>	<ul style="list-style-type: none"> <li>·Can house data service unit</li> <li>·Handle licensed and subscription based content</li> </ul>	
<b>Outreach and promotion</b> - classes, specialized programs, course support (web sites and courseware), cross unit and cross institution options	<ul style="list-style-type: none"> <li>·Individuals and classes</li> <li>·Workshops/seminars</li> <li>·Develop information and statistical literacy training materials</li> <li>·Provide course content to faculty and course web pages, moodle sites</li> </ul>	<ul style="list-style-type: none"> <li>·Same as Archive but focused on different formats, clientele</li> </ul>	<ul style="list-style-type: none"> <li>·Individuals and classes</li> <li>·Offer workshops, seminars</li> </ul>

<p><b>Collection building</b> - raw data vs. value-added, sources and suppliers, scope, approach, extent of unique holdings</p>	<ul style="list-style-type: none"> <li>-Raw data from research units, government agencies, individuals, consortia, local, national and international</li> </ul>	<ul style="list-style-type: none"> <li>-Value-added</li> <li>-Government information</li> <li>-licensed and subscribed content</li> <li>-Local, national and international</li> </ul>	<ul style="list-style-type: none"> <li>-Statistical packages and tools</li> </ul>
<p><b>Financial resources</b> - core funding, shared funds, grants</p>	<ul style="list-style-type: none"> <li>-Little or no budget</li> <li>-Consortia arrangements</li> <li>-Influenced by open access, data sharing</li> </ul>	<ul style="list-style-type: none"> <li>-Budget for data must compete with all formats</li> <li>-Licensing and subscriptions</li> </ul>	
<p><b>Preservation</b> - depending on the extent of unique holdings determine the type and level of archival activity</p>	<ul style="list-style-type: none"> <li>-Use of metadata standards – DDI</li> <li>-Preservation is part of core mission</li> </ul>	<ul style="list-style-type: none"> <li>-Cataloging and metadata standards</li> <li>-Reliance on provider for preservation</li> </ul>	
<p><b>Looking beyond</b> - participation in professional organizations, cross institutional collaborations, participation in local, state, national and international initiatives</p>	<ul style="list-style-type: none"> <li>-IASSIST</li> <li>-IFDO</li> <li>-CESSDA</li> </ul>	<ul style="list-style-type: none"> <li>-ALA</li> <li>-GODORT</li> </ul>	

At UCLA, there is now a desire to take a somewhat different approach from the way services are delivered at other institutions. Although the Library and the Social Science

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

Data Archive have worked together informally there is now a desire for formal coordination in collection management and service delivery. There are known gaps in user support; not all users know about the archive and not all users who need it are referred to the archive from the library. Therefore, our objective is that no matter where users start out in their research process, they can get at least some assistance, even if they are ultimately referred to another unit or person. The result is that we are developing a multiple-points-of-access model for services and acquisitions. Each unit will have its specialty and working together our strengths make for a comprehensive support network for users.

At Berkeley, the Library is pursuing a model of dual development, creating its own facilities for hardware, software, and expertise for support of data users. Data acquisition and facilities are well in place, but knowledgeable staffing and data expertise are current challenges. The data archive is assisting with staff training and, to a lesser extent, with direct staffing. The training and circulation of staff also provide a mechanism for identifying pools of expertise, knowledge about who and where to direct users to, and local resources. Like UCLA, the joint mission is to make sure that users are placed into contact with the support and resources they need, whether at the library, data archive, geospatial center, or other research organization, regardless of where they make initial contacts.

As stated in the wiki<sup>26</sup> on this operating model: “The basic goals include:

- Provide the best data services and resources possible

---

<sup>26</sup> See: <http://lurkingvariable.pbwiki.com>

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

- Develop in-depth knowledge, expertise and practices across campus units
- Collaborate in building collections relevant to the needs of our clientele in research and instruction
- Participate in information literacy initiatives; develop modules on statistical literacy
- Pursue development of access and data use tools, including DDI
- Support data use with GIS and data representation/display
- Use metadata standards to enhance data discovery

Having various points of service available for solving problems and sharing ideas has proved to be beneficial to all involved. Staff gain expertise in trying to answer questions as thoroughly as possible and are gradually able to assist with more and more complex inquiries. Since the data are accessible from virtually any venue with appropriate authentication, then having multiple physical assistance areas makes sense. Staff members in different campus units have expertise in the format, content and use of data from the point of view of their local users; they can contribute this knowledge to others. This produces an overall more in depth layering of knowledge and practices. Tools can be developed for use in a broad range of applications using the perspectives of staff from different units.”<sup>27</sup>

## 5. Barriers to collaboration

At UCLA the impetus for collaborating resulted from informal discussions among like minded librarians and archivists. But actually implementing the collaboration has had its

---

<sup>27</sup> Ibid.

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

challenges. Similarly, at Berkeley, the separation of units, presence of multiple data silos and localized support networks (e.g. in economics, demography, business, the geospatial center, and in research units), intentional duplication of efforts, and differing institutional cultures has made collaboration a slow process. Below are some examples institutions should address in building any data collections and services infrastructure:

In the wiki on establishing collaborative library/archive data services

([lurkingvariable.pbwiki.com](http://lurkingvariable.pbwiki.com)) each of the items are discussed in more detail.

- Getting administrative buy-in
- Determining scope and extent of service
- Collection building roles
- Curation and preservation responsibilities
- Financial arrangements
- Legal issues: licensing, copyright, human subjects protection
- Organizational culture
- Expertise, training and willingness of staff to participate
- Overall institutional mission

There can be some roadblocks in setting up data services. Depending on the institutional culture and internal environment these can be difficult to resolve. Issues of turf, ownership and political power plays are not inconsequential. There are some tools which can be used to try to identify and diffuse problem spots. First, carrying out a

SWOT<sup>28</sup> analysis can help to, as the acronym states, establish a base understanding of each units' strengths, weaknesses, opportunities and threats. And Kaarst-Brown's 2004 article on the Competing Values Framework will help to understand the operating culture of each unit.<sup>29</sup>

## 6. Challenges to maintaining differing specializations – potential directions

There are significant barriers to providing data services and data access remaining in the US. From an international perspective, it would be fair to say that the idea of collaborative, shared roles among data archives, libraries and data producers is less well envisioned outside North America. In many respects the environment outside North America more closely resembles the US in the late 1980's model of separate roles perceived by data archives, libraries and data producers. Challenges include rethinking physical collection development rationale vs. developing data discovery and access tools; national and international laws and legislation with respect to data ownership and respondent privacy; evolving methods of and opportunities for coordination among archives and libraries, implementation of standards for developing and operating trusted digital repositories, and in cross national data collection programs. The fates of the IDL&RS and ISMF collections illustrate the issues we all face.

---

<sup>28</sup>Sandra Nelson. (2001). The New Planning for Results: A Streamlined Approach - Workforms. American Library Association. (Includes templates and workforms for SWOT Analyses) Available at: [http://www.elearnlibraries.com/workforms/new\\_planning\\_for\\_results.html](http://www.elearnlibraries.com/workforms/new_planning_for_results.html)

<sup>29</sup> Michelle L. Kaarst-Brown, Scott Nicholson, Gisela M. von Dran, and Jeffrey M. Stanton, School of Information Studies, Syracuse University, Syracuse, NY 13244 LIBRARY TRENDS, Vol. 53, No. 1, Summer 2004 ("Organizational Development and Leadership," edited by Keith Russell and Denise Stephens), pp. 33–53

Maintaining cross-national resources such as the IDL&RS and the ISMF collections requires long term planning and obligations. Although interest in interdisciplinary quantitative research and cross national data collection of all kinds continues apace, collection building in the ISMF and the IDL&RS ended some time ago for a variety of reasons. Funding, institutional shifts, changes in personnel, researcher priorities have all had an impact. Technology also affects the longevity of specialized collections including activities associated with preservation, building virtual collections, and new modes of discovery and access. Whereas there can be compelling reasons to maintain physical collections of data files, it may not be as necessary as in the past. Data archives have well developed procedures for long term preservation of files and it takes next to no time at all to transfer a file from one computer to another. Rather than acquire files it may make more sense to build good search engines with links to sites where data are available. Standards for describing the content of data files have been adopted by major archives and the level of metadata available can be applied at the item level within a study. Search tools focused at this detail of metadata means that searching for useful data has the potential of yielding much more appropriate results than in the past, or more recently, than through the use of Google or other global search engines. Although they have not been well studied for efficacy, potentially promising alternative means for describing data file content through tagging are being explored in systems such as del.icio.us. Sites promoting the availability and visualization of data are being created outside the archive and library domains, including Swivel, Many Eyes and Numbrary.<sup>30</sup>

---

<sup>30</sup> See: <http://del.icio.us/>, <http://www.swivel.com/>, <http://services.alphaworks.ibm.com/manyeyes/home>, International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

Evolving national and international policies are exerting an increasing complexity affecting the ability of any one institution (whether library or archive or data producer) to support data access. Information access directives and emerging policies on ownership and data sharing, protection of human subjects, and responsibilities of data users have had some negative effects. Data producers increasingly require data users to complete onerous applications requiring legal guarantees about how the data will be used, stored, and protected. For the most part individuals cannot sign these contracts or licenses and they must obtain permission to do their work from their own research institutions as well as the data producer. Individuals without institutional backing cannot get access to these materials unless they are willing/able to take on such legal obligations. Different countries have different regulations; members of larger cross national alliances such as the European Union are affected by not only the laws of the member country, but by the larger body as well. Building cross-national collections means navigating a territory of still evolving legal policy and understanding data culture in terms of social, political and economic norms of individual nations.

Comparative social, economic, political and geospatial data are of great interest. In the past, researchers in the US traveled to other nations, collected data and brought it back to the US for analysis. Now, individual countries share their data via the Internet and the data collection activity is carried out by researchers in their native country.

Conducting surveys is expensive and there are challenges to researchers in studying

---

<http://numbrary.com/>

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

representative samples of populations. One way to address this is to collaborate with cross national data collection programs and there are numerous examples of such as the European Social Survey, the Eurobarometers, the International Social Survey Program, the International Adult Literacy Survey, and the Survey of Health, Aging and Retirement in Europe to name but a few. These surveys, survey results and publications based on use of the data are shared internationally and are potential additions to the ISMF index. Survey research continues to be conducted in separate countries as was the case with the IDL&RS collection. One outcome of this activity has been the establishment of new data archives in countries where they did not exist before. In turn, more archives has meant access to data collected by national statistical agencies, government agencies and by individual researchers is more readily available from many more countries. The South African Data Archive (SADA)<sup>31</sup> is one example of a fairly recently formed well managed archive. One can discover additional organizations by perusing the members of the International Council of Archives (ICA) which “has a global network of more than 1,400 institutional members in 190 countries [and] ... over 200 archivists and records managers as individual members.”<sup>32</sup> Not all of these are devoted to managing social science data but the growth in the number of well established organizations who are is continuing. At the same time, there is increasing emphasis placed on adherence to standards for operating data archives, preservation, and expertise needed. So how does this impact the maintenance of collections such as IDL&RS and ISMF? One aspect has been the ability of researchers to more easily

---

<sup>31</sup> See: <http://www.nrf.ac.za/sada/>

<sup>32</sup> See: <http://www.ica.org/>

collaborate cross-nationally in both data collection and analysis. And there has been some cross-national collaboration among archives and data producers. One measure of this collaboration is to look at the increase in membership in the International Association for Social Sciences Information Services and Technology. Attendance at the most recent IASSIST conference<sup>33</sup> included those from the countries of South Africa, Finland, Germany, Canada, US, UK, Scotland, Ireland, Denmark, Japan, Russia, Netherlands, Norway, Hungary, Romania, Sri Lanka, and Uganda.

There has been other international collaboration within the data community. The International Federation of Data Organizations (IFDO) was formed in 1977 “in response to advanced research needs of the international social science community.”<sup>34</sup> As stated on its web site the Council of European Social Science Data Archives (CESSDA) “is an umbrella organisation for social science data archives across Europe. Since the 1970s the members have worked together to improve access to data for researchers and students. The CESSDA Portal is a gateway to many kinds of research data and metadata, including sociological surveys, election studies, longitudinal studies, opinion polls, and census data. ... Users are able to locate datasets, as well as questions or variables within datasets, stored at CESSDA member archives throughout Europe.”<sup>35</sup> However this collaboration is largely aimed at the research community and data producing agencies; libraries have next to no role within either the CESSDA or IFDO organizations. And CESSDA itself has no mechanism for setting policy across nations

---

<sup>33</sup> See <https://www.stanford.edu/group/ADS/cgi-bin/drupal/>

<sup>34</sup> See <http://www.ifdo.org/>

<sup>35</sup> See: <http://www.nsd.uib.no/cessda/home.html>

as it currently does not have a legal status within the EU. CESSDA is developing a set of standards it will use to admit additional members (there are currently 20 member archives) largely from eastern European EU members.<sup>36</sup> CESSDA members have the support of the data community through organizations such as IASSIST; it remains to be seen whether having a legal status can aid in the development of laws and policies enabling data access or whether European libraries will eventually play a part in providing this access. It may be that similar nationally based member organizations need to be established for regions beyond Europe in order to build and/or expand global collaborative opportunities.

The experiences at UCLA and Berkeley share some characteristics. Early experiences show that, while apparent similarities in missions of acquisition, documentation, access and preservation might suggest that either archives or libraries could fulfill data users' needs, the expertise, network of contacts and resources required 'on the ground' necessitate looking beyond mission. Historical experience also shows that the kind of expertise, resources and networks needed will change - often dramatically. Many of these changes are the direct result of work libraries and archives do to build the infrastructure needed - through the creation and implementation of standards, the formalization of distribution networks, the evolution of professional organizations, and the creation of staffed physical facilities with the funding needed to sustain them. Other changes reflect technological progress which has vastly expanded the potential users of

---

<sup>36</sup> Based on notes from a presentation titled The CESSDA ESFRI project – Setting Up a One-Stop Shop for European Data given by Kevin Schurer and others at the annual conference of the International Association for Social Science Information Services and Technology (IASSIST), held in Stanford, California, May 27-30, 2008. <https://www.stanford.edu/group/ADS/cgi-bin/drupal/session/the-cessda-esfri-project-setting-up-a-onestop-shop-european-data>  
International Federation of Library Associations, Social Science Libraries Section, Satellite Conference, *Disappearing disciplinary borders in the social science library - global studies or sea change?* University of Toronto, Toronto, Canada, 6-7, August 2008

numeric data, the ubiquity of new data collection and sharing, increasing complexity of surveys and data collection processes, and the ease with which users can use - or misuse - data.

As a result of both of the sources of change, many of the resources and expertise needed to support users have either become routinized or embedded in the environment, while new domains of needed expertise have emerged. In general, acquisition has become a less specialized activity, the need for unique local collections has diminished, and many computing support issues have become less labor intensive. At the same time, the need for education and outreach has grown, as has the development of standardized metadata and importance of tool development. Left largely unchanged is the researchers' need to review data collection instruments, codebooks, and reports on sampling and weighting, and to understand quantitative research methods, know how to use a statistical analysis software package, and interpret results.

This evolution of the data service environment and the ongoing pace of change in that environment argues against either the monopolization of service provision by either archives or libraries found in the field's infancy, or even the fixed division of responsibilities characteristic of middle years; instead, a fully collaborative effort which draws on the skills, resources, networks, and experiences of both institutions is needed. The challenge is to evaluate what each institution can offer, bridge divides in institutional cultures, fit roles to meet institutional missions and constraints, and create

the sustainable relationships needed to appropriately support the broad range of users who are now our clients.

---

<sup>i</sup> See: <http://lurkingvariable.pbwiki.com>