Data Library Services Review
The University of Regina
2006

The University Librarian initiated in June 2006 a consultation to review Data Library Services in the University of Regina Library. The findings of this investigation are based (i) on a series of interviews with academic and Library staff at the University of Regina, (ii) on comparisons with data services at peer institutions and (iii) on an assessment by the reviewer. On the basis on the evidence gathered, a series of recommendations address questions raised in the terms of reference.

Consistent themes emerged from the interviews about the need for Data Library Services on the Regina campus and about the operation of the existing service. In a nutshell, the data users' community on the campus confirmed their need for data library services both for teaching and research. Furthermore, they indicate their need for data will increase as the University expands and strengthens its research mission. The recent emergence of new institutes (for example, the National Summer Institute for Statistical and GIS Analysis of Crime and Justice Data), new centres (for example, The Environmental Research and Response Applications Lab) and new Canadian Research Chairs (for example, the Research Chair in Public Policy and Economic History) are signs of growth in areas where data are critical to the University's research mission. This community also expressed strong praise for the Data Library Services staff with several people offering testimonials about the high quality of service they have received.

The response from within the Library acknowledged the relevance of data services in an academic library but reflected a mixed understanding about the nature of these services and its place in the complement of Library services. From the perspective of this reviewer, misunderstandings stem from an earlier reorganization that introduced extensive changes based on a service model that viewed the campus as a homogenous undergraduate population. The specialized skills and services associated with data did not mesh with a move toward a single, generalist service-point model. While mainstreaming data may have been assumed under reorganization, reality revealed data services as a round peg that would not fit in a square hole. Data services fit better in a model that supports a heterogeneous community of users and incorporates specialist services, which are typically characteristics of a research university.

Several strengths of the existing Data Library Services became evident through this investigation. As already acknowledged above, the University of Regina Library has two very competent and skilled staff, Marilyn Andrews and Raye Quickfall, providing data services. They have worked in this area for many years providing stability for this specialized service, whereas staff turnover has caused problems for data services at other Canadian universities. This stability is especially an asset for data services because of the level of skills required for its on-going operation.

Another area of this unit's strength is in public services. The Data Librarian has a strong, collegial relationship with the users on campus. She is very knowledgeable of their research and teaching needs for data and keeps in close communication through a variety of methods, including newsletters, e-mail and personal visitations. The University of Regina Data Library Services has some of strongest ties with its user communities that I have witnessed across Canada. The Data Librarian also contributes to the Library's information literacy program by helping prepare trainers and by delivering specialized instruction. A recent example was her presentation in the National Summer Institute for Statistical and GIS Analysis of Crime and Justice Data.

The most significant challenges facing Data Library Services are related to boundary issues *internal* to the Library. As previously mentioned, the origin seems to stem from an earlier reorganization based on a service model that did not readily accommodate the specialization of data. Without dwelling on the causes of these boundary issues, which is beyond the scope of this review, their negative consequences can nevertheless be clearly observed. Foremost, the service has suffered from isolation. First, the staff are five floors from the main point of service, which has resulted in inconvenient, awkward or missed consultations between staff and users. At one point, the service was relocated to a branch library, which created another type of isolation. Secondly, a special data pod was constructed in the Information Commons to support some data products requiring authenticated access. The number of data products available through these workstations is a small part of the data collection, which has misled users expecting more data than what is available on these machines. Users have expressed confusion about the difference between data pod workstations and all of the other machines in the Information Commons. Third, data services at some point became viewed as a problem for technical services, including cataloguing, computer support and its Web presence. Without a widely shared understanding of data, its services became viewed as a 'problem child.'

Recommendations

The current level of Data Library Services in the University of Regina Library is appropriate for this institution given both the size of the University and its teaching and research mandates. One only has to look at the data services at Simon Fraser University and the University of Windsor, which were identified by several of those interviewed as peer institutions, to discover comparable staffing, collections and services in data. At the University of Regina, a core collection of licensed data products is maintained by a skilled staff, who provide quality reference and technical services to support the use of this collection. The level of services, as such, is not an issue from this reviewer's perspective. The following recommendations, however, address specific questions from this consultation's terms of reference and build upon the strengths of the existing Data Library Services while shoring up areas for which there are weaknesses.

A. A visible, publicly accessible point of service for data needs to be located on the main floor of the Library. Data services need to be on the "main street" of public services, which the entry floor to the Library has become as a result of an earlier reorganization and renovation. One space that seems particularly suitable for Data Services is the office adjacent to the computing consultants near the Library's entrance.

   Data are materials that can involve extensive mediation, which runs counter to the movement of increasing user-independence. This level of mediation arises from the level of skills required to work with these resources and from the licenses governing access to data, some of which explicitly stipulate mediated intervention by Library staff. While not all aspects of data services will become mainstreamed in the Library, this can be offset by placing data services on the "main street" of the Library's public services.

B. A solution is needed to move beyond the view that data are a "problem" for technical services. The current boundaries within technical services need to be more flexible in dealing with data. There are technical areas where data can be treated similarly to other Library resources, for example, incorporating product-level MARC records (see Recommendation C). There are other areas where the level of technical services needs either to expand to accommodate unique aspects of data or to enable the staff of Data Library Services to carry out these functions. For example, the software requirements to support data include statistical packages, such as SPSS and SAS. Computing support can either extend its coverage to include this software or it can grant system administrative privileges to Data Library Services staff to support this software on their machines. Another example is the Web presence for data services. Because Data Library Services function as the vendor interface for several licensed products, with DLI being a major resource, its staff need the flexibility of managing its Web pages to comply with licenses. Unlike other library database subscriptions on the Internet, the vendors of data resources do not always have direct pages to create the access link. These pages must be constructed locally and the Library's Web management practices must be flexible enough to allow for this.

C. Having MARC records for the Statistics Canada DLI collection included in the OPAC would enhance the discovery of DLI data products. Through the University of Regina's DLI subscription, the Library in Statistics Canada offers MARC records for DLI data products without additional charge. In consultation with the Data Librarian, the Cataloguing Unit should determine the fields from these records to incorporate in their local OPAC, including how patrons *locally* access DLI data.

   This recommendation is contingent upon the flexibility of the Cataloguing Unit to download and batch-process Statistics Canada MARC records. Copy-cataloguing DLI MARC records may be more easily handled through a university partner than Statistics Canada. This recommendation is not prescriptive of the method of

record capture and input but rather is directed at increasing discovery of DLI data through the incorporation of this collection in the OPAC.

The MARC records provide a product-level description and are only one type of metadata critical to data services. Unlike most other library resources, data demand accompanying metadata describing the detailed characteristics of the data. Without this level of detail, the structure and organization of the data will be unknown, rendering the data useless. Metadata at the data file level currently appear in multiple formats and standards, including print and PDF documents, record layouts and statistical software command files. The DDI standard is emerging within the field of research data as the best practice for preparing this level of metadata. Eventually, MARC records will be generated from DDI-compliant metadata supplied by data vendors. Because this development will be taking place over the next five years, there is no immediate recommendation regarding the use of DDI metadata. Instead, the University of Regina should be prepared to work within COPPUL and other consortia memberships to make use of DDI metadata as its availability increases.

D. The budget for licensing data needs to be rationalized within the Library's overall collections budget. Because each academic library operates within its own budgetary framework, offering a prescriptive recommendation would be futile. Furthermore, the budgetary practices for allocating electronic resources in academic libraries vary greatly to reflect institutional consortia participation, local licensing management and traditional disciplinary funding formula. Nevertheless, specific objectives can be met in building expenditures for data into the Library's collections budget. Data products are multidisciplinary by nature and disseminated according to the terms of annual licenses. Both of these characteristics should be guiding factors in determining a place for data in the collections budget. For example, the DLI license may appear in a system-wide electronic resources budget item because this membership supports multiple disciplines, while the budget item for the CFMRC TSX database may be assigned specifically to the Data Library because this database tends to be used by only one or two disciplines. The assigned management of data expenditures in the budget should also take into account the person (i) who is knowledgeable about the data product or license, (ii) who can assess the product's continued value for campus users and (iii) who has established a client relationship with the vendor. This determination should occur in consultation between the Data Librarian and the appropriate person managing the collections budget.

E. Senior management in the University of Regina Library should encourage and support their staff's participation in the annual DLI/ACCOLEDS training workshops. Through its membership in the DLI, the University of Regina Library can send as many staff as it desires to these annual three-day workshops that are substantially funded through the DLI national budget. No registration fees are assessed for participation. Yet to pay a commercial vendor, such as Statistics Canada or ESRI, the cost of this type of training would run close to $1,000 per

day per participant. In addition, DLI subsidizes the travel costs of one person from each member institution to participate. This combination of no registration fees and subsidized travel is an open invitation to academic libraries to receive continuing education in data services as a minimal cost (only lodging and meals and travel for anyone in addition to the person receiving the travel subsidy.)

Statistics Canada continually releases new data products and the best way for data services staff to stay informed is to participate in these DLI training workshops. These workshops not only provide information about new data products but also provide instruction in statistical literacy and teach skills in statistical and GIS software.

F. One final recommendation is related to Data Library Services but only tangentially. The Library has an opportunity to support an initiative to establish a Research Data Centre (RDC) on the Regina campus. Saskatchewan is one of three provinces that does not have an RDC. These Centres are secure data enclaves housing Statistics Canada confidential microdata. Strong interest was expressed during the reviewer's interviews in constructing either a full or branch RDC in Regina. The Library could become a partner in this initiative by volunteering to host the RDC in its central facility. The majority of RDC's across Canada are located in campus libraries. As central services for the entire campus, libraries house resources for all researchers and organizationally provide services supporting access to these resources. RDC's are simply another way of providing access to research materials, although in a highly controlled and secure environment.

Submitted by Charles Humphrey
September 6, 2006