A Library Service Model for Digital Data Support

by Susan Czarnocki [1] and Anastassia Khouri

Introduction

The analysis of numbers, beginning with raw research data and emerging with knowledge, is a vital skill. Its role in a solid education is well established, but it takes specialized skills and support systems to provide optimal conditions for data analysis to flourish as part of an undergraduate curriculum. As stated in an article elsewhere in this issue of IQ, data files have a layer of complexity that can make them more challenging to use than other information sources (Edelstein & Thompson 2005).[2] Data analysis is now routinely carried on almost exclusively with digital data, using some type of analytical software on a computer. Optimally, incorporating such analysis in an undergraduate curriculum involves the coordination of all those resources and the skills necessary to utilize them.

In setting up their courses, instructors take the assignment of required readings for granted. The support machinery and the skill sets are all in place: the library stocks the books, the students read the books and are rated on their ability to digest their contents.

But what if an instructor wishes to assign students a required data analysis project? Since the 1960s, completing an assignment using data has generally meant entering into the world of computers and software. How best to provide the data, computers, software and statistical and technical support that are involved in assignments that involve data analysis: this is a challenge for all colleges and universities. Over the years, this challenge has been evolving as the technology itself evolves.

This article examines the experience of McGill University in responding to the changing computing and data support environment.

Mainframe Era

Instructors who wanted students to do computer-based data analysis when mainframe computers were the only available option encountered many hurdles. Mechanisms had to be developed for students to get computer time, codes and passwords and to be initiated into all the intricacies of submitting, correcting and retrieving jobs using the abstruse job-control language of the mainframe. Of course, some students excelled, and produced reams of output. But generally, it was not an experience relished by the average undergraduate. By and large, it would only be those courses which were mandated to teach social science methods and statistics that would include a project involving analysis of primary data.

PC Era

The introduction of the personal computer has made the computer a much more widespread part of our environment. There is an expectation that an undergraduate needs to be able to 'use' a computer. But often 'use' means interaction at the level of an expensive typewriter and communications device, for example, sending emails or surfing the Internet. Numeracy may not play a role in using a computer. Undergraduate courses may be the first time that a student is expected to analyze a problem using numbers and graphs.

The personal computer has also dramatically changed access to data. In the 'mainframe era', data was being stored at the computing centre because it came on tapes which could only be read at the computing centre. With Internet connectivity, every PC, whether at home or on campus, can be a potential channel for acquiring data. The physical archiving of data can be 'anywhere' that there is an Internet connection. Likewise, the analysis tools can now be installed on computers anywhere. So, in terms of access to data and analytical tools, a huge revolution has occurred.

Supporting the Data-Access Revolution

Does this revolution, in itself, make the introduction of data analysis into undergraduate classes of 100-200 students a possibility? Can data analysis be widely incorporated into the undergraduate curriculum? Should it now be possible to promote use of data in any course which involves learning to make and challenge interpretations of numerical information relevant to a particular discipline? While the challenge of availability has receded, there is still the challenge of building up basic statistical literacy and software-related computer skills so that these are not barriers to a student's progress. Theoretically, just as data and software can be available 'anywhere', so should support. And in a sense, it is. The Internet does mean

that on-line software tutorials and guides can be available anywhere, through on-line help and a vast array of help-tools. But while these are useful for a portion of the students who have had greater experience with or aptitude for computers and number, many students are not prepared to confront that mode of learning about a topic which is essentially foreign to them. There needs to be local support to help "level the playing field" to some degree for those students who have not had such past experience if data-analyses projects are required. This implies opportunities for students to obtain training in the tools that are needed and to be able to have access to some assistance in the minutiae of software techniques, producing graphs, etc. which are the major sources of frustration for the uninitiated. There also needs to be sufficient access to computers and software so that access to them does not become an obstacle. Certainly, in terms of efficiency and reduced frustration levels for the user, these requirements are best met in a unit which can offer one-stop services for all of the elements of dealing with electronic data, from technical to statistical.

Activities of a Data-Support Unit

Support for large classes is, of course, just one of many forms of support that a unit can undertake. The creation of easily-usable datasets is more efficiently done once by a central service, than repeatedly by course instructors in several departments. Levels of support that were once of interest mainly to graduate students can now be applied to encouraging undergraduate use of electronic data. If they can receive minimal support in locating and manipulating data-sets they find more interesting, students doing individual research projects can go beyond those heavily-used standard datasets that have to be pre-prepared for instructional purposes. The local data support staff can create usable data sets on more esoteric topics, as requested, allowing the student to focus on analysing the data. In these interactions, the data-support team can offer some words of caution as to the pitfalls if a student has set his/her ambitions unrealistically high, and offer alternative suggestions.

Institutional Home for Data Support

In the world of distributed data and software, where is the best place to provide the assistance for being able to use what is now so readily available? As the data and tools have become much more widely diffused, it is not always evident where the support should be located for teaching how to use these tools. But without such support, it is likely that these data resources will remain underused.

The McGill Experience

In the 1970s and 1980s the researchers in McGill social science departments depended on the services of the university computing centre for support for their instructional and research computing services. There was no push to create a specialized data library for those departments, or centrally for the university. During that period, the Department of Economics operated a statistical consulting service composed of a statistical specialist with assorted graduate students as assistants. It focussed on assistance to professors with large research grants. Researchers were generating data from their own research, obtaining it from ICPSR, as well as buying it at high prices, often from agencies of the Canadian government. They tended to rely on the informal channels within and among departments for finding out what data might be available, for example, for graduate student research. Some graduate assistant time was used to make a catalogue of data-files on tape that were being archived at the computing centre. This catalogue was a simple text file, stored on a mainframe account, and although the file was 'public', almost no one knew of its existence. It was not generally distributed. Assistance on accessing data-files, or acquiring new ones via ICPSR was given to any student seeking such help, but this was done as a favour to other departments, not as a major mandate of the unit.

When the statistical consultant in the unit moved on to a position in the University Planning Office, the social science departments were spurred on to re-evaluate how computer services should be provided for research and teaching. They were able to obtain a budgetary allocation for the creation of the Social Sciences Computing Centre (SSCC). Susan Czarnocki, as the first manager of this service, was given a mandate to administer the training required to use the mainframe from the remote terminals in the SSCC, and to develop a set of student consultants who could assist others in resolving error messages, getting their printouts, etc. The computing centre had developed a number of procedures to provide support to instructional computing, but very few courses at the under-graduate level in the social sciences attempted to include assignments using data projects. The learning curve for using the mainframe was seen as absorbing too much energy away from the course content.

In terms of more general data support services, the SSCC manager was also given the role of ICPSR Official Representative (OR), and was handed the catalogue of data-files and a list of tapes at the computing centre. Students seeking to use data sets were told to speak to the SSCC manager, but offering of a full-fledged data-service never became part of the official mandate of the SSCC, nor of its expanded successor, the Faculty of Arts Computer Laboratory.

Getting the McGill Library involved

During this same time-period, across Canada, a wide-range of services were being developed for dealing with data access, reflecting the variation in budgetary and structural arrangements across universities. There were services at computing centres and services attached to Social Science departments. Those hired to offer these services often

met for the first time as ORs for ICPSR, or at IASSIST conferences and began to share their experiences as data support professionals.

The announcement of a sharp increase in the costs of acquiring the data files for the 1986 Canadian Census spurred a movement to greater collaboration. The result was a consortium formed through the Canadian Association of Research Libraries for purchase of that census data, with the University of Toronto Data Library taking on a central role in the reproduction and delivery of the data to the other members of the consortium. Through these negotiations, libraries across Canada were becoming engaged with the issues of support for data services at Universities. This first consortium agreement formed the precedent for expanding agreements with Statistics Canada, and the libraries became the channel through which the data was being released.

When Statistics Canada signed the Data Liberation Initiative (DLI) in 1996, McGill opted to retain the library as its channel for participation.[3] The Head of the Libraries asked Anastassia Khouri, who had been responsible for the computerization of the Libraries, to draw up a plan for a pilot-project data service. This lead to the initiation of the McGill Libraries Electronic Data Resources Service (EDRS) in the fall of 1997. The original remit of the service was to: support teaching and research in all disciplines; develop a core data collection and associated resources to support McGill research and academic programs that would complement the printed collection available in Government Documents and the Branch Libraries; provide access to resources and data via the EDRS web site where all electronic data resources services would complement the Library's current and future services.

EDRS and the Faculty of Arts Computer Laboratory (FACL) collaborated closely in an effort to support data usage across campus. This was a collaboration involving the two authors of this article. Anastassia would often provide students with data files, and then send them over to Susan at FACL so that she could make sure they understood how to use them. EDRS was a resource for locating, acquiring, archiving and retrieving data, and at FACL, users could obtain some limited one-on-one assistance for utilizing microdata. At that time, students could still choose to use the McGill mainframe, or personal computers, or both -- to complete their analyses.

Since 1997, EDRS has been designing web-based interfaces to promote access to data resources and services. An acquisition budget was allocated and space on a university server has been made available for archiving. We are currently aiming at the development of a data gateway offering access to the thousands of resources that are available on the web.[4]

During the 1990s, the Libraries also saw increasing changes in their mandate as they were moved from reporting to the Vice-Principal Academic to the Vice Principal-Information Systems Technologies (VP-IST). When a social scientist was named as VP-IST in April 2000, he was aware of the fragmented situation of data-support, and was interested in finding a better approach. Creating a unit that was a 'one-stop shopping' data service was a somewhat unconventional mandate from the perspective of many Librarians. However, the successful experience of developing the EDRS within the Libraries provided the impetus to develop that facility. It developed along the lines of other data services that were housed in Libraries around North America, and in some ways move beyond them to offer a wide range of support for research initiatives and professors interested in much broader inclusion of data utilization in their under-graduate courses. In 2001, Susan Czarnocki was transferred from the Faculty of Arts Computer Lab to join Anastassia Khouri in the EDRS. Since that time we have worked to develop a full range of services which aim to assist not only the heavy-duty data crunchers, but also to assist any instructor, with courses of any size, to incorporate assignments that utilize data for instructional purposes.

Social Science Data

Social science data is now a significant budgetary item for the Libraries. In previous decades, micro-data were often purchased by researchers for their own use, with only infrequent access to such data sets being made available to undergraduates. Macro-data tables, such as those contained in statistical yearbooks, were the main source of data for undergraduate research papers. The positioning of these volumes in reserve sections of Government Documents sections, or reference areas, meant that only the more ambitious students actually incorporated such material in their work. Now, the library is full of computers, and all the computers can access primary data sources in the area of international development and finance, United Nations data, and national archives and statistical offices around the world. In Canada, Statistics Canada has developed an interface to Canadian statistical and census data called "E-Stat" which facilitates the manipulation of macro-data down to the census tract level for a wide range of social and economic characteristics[5]. This has greatly widened the range of undergraduate courses in which the instructor can pursue the possibilities of requiring students to obtain and analyze numeric data.

Through EDRS, McGill also participated actively in the design and the creation of Sherlock, a shared infrastructure for bilingual access to and manipulation of numeric data. Sherlock was developed as a collaborative effort by Quebec university libraries.[6]

Current EDRS Service Model:

The EDRS service model has evolved since 1997 and will continue to adjust to the various changes in the delivery of

information services. The EDRS is part of a grouping of library services which have digital information as a major component: maps, electronic data and digital government Information. Historically, the library had maps and government documents as two separate services, operating in separate buildings. The location in separate buildings remains, but the direction of these services has been combined with the EDRS to form a unified administrative unit: The Government information, Maps, and Electronic Data Centre. The aim is to have the staff in each sub-unit have some familiarity with the software and problematics involved in the work of all three sub-units, in order to offer an integrated service to the increasingly frequent research collaborations involving GIS and government-generated digital data. The unit provides support for all research and instructional activities involving digital data. The description below focuses on the approach taken with respect to supporting use of data in undergraduate curricula and for graduate students.

The EDRS Facility

The EDRS occupies a large room near the so-called 'information commons' area of the Humanities and Social Science Library. The room is equipped with 14 computers for student use. Several of the major statistical software packages are available (SPSS, SAS, Stata, E-Views), along with full documentation. It also provides access to useful tools for electronic data handling such as StatTransfer and Adobe Acrobat. Support for group instruction is also provided, with projection facilities, etc. on the premises.

The Geographic Information Centre complements those facilities by having 6 GIS workstations with GIS software (Arcview and ArcGIS in addition to SPSS). The Government Information Service has an additional 8 workstations. All equipment has "cutting-edge" capabilities to allow easy manipulation, analysis, saving and accessing data and the associated full-text documentation. The Libraries also provide small and large electronic classrooms fully equipped with the necessary software. Those facilities, when not used for teaching, are available for student use.

Some copies of IMF print resources, such as World Bank Human Development Reports, are retained at EDRS, so that students can have a quick overview of the types of data that are available from these sources, or what coverage is available for a certain country, etc. before they start to access the data electronically. The print collection of ICPSR codebooks from years past is also housed at EDRS, along with codebooks and user guides for many of the widely-used Statistics Canada surveys.

Various help tools have been developed to help and support teaching and research. EDRS is developing a series of 'How to' web-pages: as an example 'How to Export a Table From a PDF File into An Excel Spreadsheet'.

Multidisciplinary Research and Instruction

Through combining the resources of those units dealing with digital data, it is able to respond more creatively not only to individual student or faculty research, but to large multi-disciplinary projects involving historical, socio-economic and GIS data. More and more, funding agencies are encouraging the development of research projects that mingle and merge across interdisciplinary boundaries. The combined unit is well-situated to support such projects and is being sought by researchers as a partner in their applications for such grants. Whether it is research on the fate of the aboriginal communities of the James Bay Cree, displaced by hydro-development, or ecological research at the McGill Biological research station in the Barbados, there is a need for a blend of digital GIS, numeric and official government information. Such collaborations make for rich research experiences which can now be extended in some cases down into upper-level undergraduate programs, such as the McGill School for the Environment, and courses on geography.

Support for Large Undergraduate Classes

Course web-pages: When requests are received for support for the use of data in a course, the EDRS staff will prepare a webpage tailored for the course, highlighting appropriate data-sources. We also offer to make a presentation based on the web-site during class-time, and encourage students to use the EDRS equipment and software to complete their assignments[7].

Software training sessions: EDRS staff provide a variety of training sessions and workshops on software, such as Excel, SPSS, Stata and SAS. Some sessions are linked to the fulfilment of a particular assignment, others are offered to any member of the university at the beginning of the academic year, using training spaces in the library.

Onsite consultation: Students seeking help can come to EDRS, work on an assignment for which EDRS staff has made an in-class presentation, or any other work involving numeric data and get technical assistance if they are having difficulties.

Support for Graduate students

Graduate orientation: EDRS staff make an effort to schedule sessions for graduate students in appropriate departments at the time of departmental graduate student orientation sessions. These presentations are aimed at demonstrating the wide array of research resources available, and indicating that the EDRS is available to help them track down elusive data resources as they begin pursuing their research ideas.

Class presentations: They also make presentations for graduate level classes, and provide sessions on using whatever software is being used for their projects.

Data acquisition: In some cases, graduate student research will stimulate the acquisition of resources not previously acquired. Being in close touch with the graduate students allows for a good gage of what resources are lacking.

The result of this approach has been to build support for several courses of 150+ students, across several departments, which are able to include assignments using major Canadian and international statistical data bases, as well as census data, etc. In the past year, for example, large lecture courses in International Development and in Urban Geography have had assignments requiring utilization of several of these complex electronic data bases. EDRS supports this process in several ways: EDRS staff work with the professors to package data resources specific to the course content; help to assess the feasibility of the data assignment; and receive feedback on how the students were able to complete the assignment. Assistance to the students involves EDRS staff members: demonstrating usage of these data-bases during a lecture period and helping those who come to the EDRS needing extra coaching, to improve their skills in the manipulation of electronic data. In the case of the Urban Geography class, EDRS staff prepared and presented 5 sessions demonstrating how to use Excel to fulfil the assignment.

EDRS also collaborated with a Professor in the History Department to make available original data on Lebanon that he had collected, as well as train his students to use SPSS to undertake some simple analyses.

Conclusion

As noted by Robin Rice in a report about 'Barriers to the use of numeric data in learning and teaching,' advances in information technology are creating new spaces for learning beyond the traditional classroom, and forms of teaching beyond the traditional lecture (Rice 2001).[8] The personal computing 'revolution' has made possible the speedy delivery of numeric and verbal information wherever the electronic network can be made to reach. This is a necessary condition for a great expansion in the ability of students to analyze real data in tacking real-world issues. But this is not a sufficient condition. For real analysis to result, much infrastructure needs to be in place and many skills need to be acquired. McGill University is experimenting with providing a unified administrative unit within the library system, for providing the infrastructure and skills-training for making data analysis happen, as part of the regular undergraduate curriculum.

The McGill University data service model has many resemblances to a number of similar services elsewhere, but with a special focus on an integrated service to users, combining data acquisition and retrieval with statistical and software assistance oriented to making students skilled users of data resources. We modify our approach to data dissemination and support as the prevailing campus technology-infrastructure changes. Resources are made increasingly available and we are following up every opportunity to augment the list of resources offered to our users. Many collections-development strategies are selected and various acquisition methods are applied. Our objectives are to augment the connectivity benefits of the library network, apply rigorously data and software licenses, to cooperate with our campus community and beyond it, and to coordinate our activities with multiple partners. The data world has been a model of cooperation and partnership which is encouraging the development of a true numeracy revolution.

Notes

[1] Contact: Susan Czarnocki, EDRS Centre, Repath Library Building, Room R-23, 3459 McTavish Street, Montreal, Quebec,Canada H3A 1Y1. Phone: +1 (514) 398-1429 / 398-4702. WWW: http://www.mcgill.ca/edrs/. Email: susan.czarnocki@mcgill.ca

[2] Edelstein, Daniel M. and Thompson, Kristi (2005), 'A Reference Model for Providing Statistical Consulting Services in an Academic Library Setting', Paper presented at IASSIST Conference, Madison, 2004.

[3] Data Liberation Initiative (DLI). Available at: DLI ref http://www.statcan.ca/english/Dli/dli.htm

[4] http://www.mcgill.ca/edrs/

[5] E-stat. Available at: http://www.statcan.ca/english/Estat/licence.htm

[6] Sherlock. Available at: http://sherlock.crepuq.qc.ca/public/anglais/sherlock.html

[7] http://www.library.mcgill.ca/edrs/seminar/health/nursing05.html)

[8] Rice, Robin (2001), 'Understanding Barriers to the User of Numeric Data in Learning and Teaching. 'IASSIST Quarterly. Vol. 25, no. 1. Pg.5-9.