

Data Services Assists Teaching and Research: Delivery of Data Services via the World Wide Web

Overview

So far in its history, the Data Services Division at UBC has played a passive role in the teaching and research that goes on at UBC. Patrons have come to us as the machine-readable data experts for many types of files in the different subject areas covered by the teaching and research efforts at UBC. There has been a lot of personal interaction with patrons as they struggle to learn various computer systems and access retrieval methods to get the data they want. With increased improvements in computer networking technologies that have taken place in the last few years, we are seeing patrons in a new light as their needs change. The UBC Data Services Division is thus becoming more proactive in the teaching and research efforts of our patron groups. We are changing our role in response to shifting patron needs; to offer services that will assist patrons towards self-sufficiency at information discovery and retrieval rather than being dependent on personal interaction with a data librarian. Such a move requires close contact with faculty and students to develop the tools that they need to achieve their goals. It also makes us a more active participant in the learning process that goes on in the classroom.

This paper will discuss some of the issues involved in providing access to data files over a computer network. It will be shown that there are different, identifiable user needs to searching out and acquiring data, and that we can employ mechanisms that can help us handle some of these needs. Specific examples of changes that have been implemented and are envisioned for the future at UBC will be examined.

Reasons for Change

There are two common themes in many Canadian public institutions: an external demand for increased service from patrons coupled with an internal desire from staff not to allow currently perceived levels of service decline. During the first three decades following World War II, there was a great desire on the part of governments and their citizens to grow, to build, to do new and innovative things. As a result, governments around the world went deep into debt, spending on and lending money for capital projects to build infrastructure like roads and schools and industrial systems, things that were believed to be needed to drive our

*by Brian Kroeker**

economies. Governments and the public alike became addicted to public debt, which left citizens with a society-wide attitude that they had only to ask for something, and the government would provide. As universities in Canada are mainly funded by the government, the idea of perpetual growth in services has been

reflected at university libraries as well. The situations which fueled some of the world-wide desire for growth have since shifted, leaving governments everywhere in the position of no longer being able to justify the current levels of public debt. Many members of the public though seem to be left with a desire for continued 'upward' growth and the constant provision of more services for which the existing funds are no longer able to provide.

Many institutions reacted initially to the trend of downward levels of funding by introducing new ideas about the use of technology in the workplace as well as methods of reorganizing operations. This was done in the hopes that greater efficiencies would lead to the maintenance of current levels of existing services, or even create opportunities for continued growth. The promise that technology and organizational restructuring by themselves would lead to continued growth has in large part failed to lead to desired improvements in overall levels of service in many academic libraries. Some of the tools that were deployed towards that goal however, have led to new services and ways of providing service. Important lessons that have been learned in the last few years from the organizational standpoint are:

- Know your patrons/clients/customers and what their needs are. Stay in touch.
- Know your available and potential resource limitations. Letting your patrons know as well helps them to understand your position better.
- Know what other institutions are doing.

We know now that continued growth of services will not happen in the current climate of downward or static funding levels. Less money and fewer staff means lower levels of service. Since patrons and service providers alike wish improved service, some improvements will likely be made in some areas, but the trade-off is that some existing services will be reduced or discontinued.

The last point to realize why institutions must change is that patron/client/customer needs change. Many public institutions have operated and continue to operate on the principle that they provide a service to the public which does not change unless the institution itself changes. This attitude ignores the importance of the patron's/client's/customer's evolving needs in providing service. At UBC as well as elsewhere, technology and new teaching methods are having an impact on how our patrons both currently use, and desire to use our services. We need to understand how patron needs are changing, so that we may change with them.

The remainder of this paper will outline what the staff of UBC Data Services are doing to bring about improvements in service for their patrons.

Identification of Patrons and Their Needs - Are We Being Effective?

Data Libraries in general are faced mainly with two different types of patrons, those who need to perform extensive research, usually over significant amounts of time, and those who need "just a few numbers" or facts. Faculty members and graduate students fall into the first category, along with professional researchers working for the media, law firms, and other information consuming companies. These researchers do not usually require significant amounts of staff time by the time they visit a data library as the academic way of life has taught many of them the art of self-reliance and asking detailed, direct questions. This group of patrons may need some initial training in procedures, but many of them are willing to jump through many hoops all by themselves to get what they want. If they can be pointed in the right direction, they will do the rest, so starting them off on the right footing is important.

Staff time can be consumed in large quantities by those patrons still being trained in the academic way of life, or for those who are not traditional students at all, i.e. members of the general public. Patrons who "just need a few numbers" have often been directed to an archive or library by someone more experienced such as a faculty member or former/current user of the institution themselves. This group of patrons most often encompasses undergraduate students, members of the media, and the general public. Members of this patron group may arrive on the doorstep of a data library with unrealistic expectations of service and end up consuming large amounts of staff time. This group needs to know exactly what the collections of the Data Library hold, as well as services provided. They also need to know where to go if the Data Library can't answer their needs, i.e. where is the next most likely place to go?

The UBC Data Services Division has always had a mandate to provide data services to the entire campus, but due to

staffing levels and availability of subject matter, the scope of the division's activities have always been focused on the Social Sciences rather than on every subject area UBC offers. Within the Social Sciences, printed reports in the subject divisions serve a growing number of students and researchers alike, and UBC's data-oriented library staff are too small in number to be familiar with the bulk of them, not working with them every day as others do. It is no longer possible, if indeed it ever was, for Data Services staff to effectively identify for patrons sources of information that they require or could use, especially print sources. As the growth of electronic sources of information continues, it is becoming apparent that Data Services does not have the staff to provide comprehensive knowledge of all subject areas under its mandate. Patrons need a single point of service for their subject information needs if this is possible.

The bulk of UBC's Data Services collections come from statistical bodies, both governmental and non-governmental. As well as raw, or primary research data, some of these data producing organizations also create and redistribute reams of machine-readable (static tables and reports) and print materials based on their raw data. In some cases their raw data are merged with the data of other organizations, leading to even more opportunities for analysis and reports. These printed and machine-readable report style materials are often more easily consumable information products than raw data products, which require specialized training to use effectively. The report style products are used by those patrons who wish to perform secondary analysis, i.e. to just "look up" what the statistical producers and others have decided would be of most common interest. Patrons have two needs in this sea of information products:

- They need experts who can guide them to the right sources of information for their needs.
- They need experts who are familiar with the various formats the information can come in and who can guide them in accessing the information.

How Has UBC Data Services Filled Patron Needs?

Within both patron groups identified in the previous section at UBC, Data Services staff members were answering many of the same questions over and over again. We wondered if we were performing an effective service here. It was a well thought of service, given the positive comments and praise we have received over the years, but was this the best way to serve our patrons? Could this time have been better spent, but still provide our patrons with what they required?

We turned to the technology of the WWW to answer these questions. We designed what we think is a fairly informative WWW site, one that answers many of those basic questions that we used to get over and over. We then

began directing faculty and students to this site when they had basic questions, instead of answering their initial queries for information. We also coordinated somewhat with faculty to make sure that this site contained the kinds of information they required for their research needs, and to pass on to their students, though this aspect requires more development. On top of this, we have educated our colleagues in other divisions about our WWW site and what it contains. The result after over a year of operation is that personal patron visits to the Data Services division are way down, but visits (hits) to our WWW site are very high, in the hundreds of hits per month range from UBC addresses alone. Feedback we have had from students and faculty indicates that people like this technology and are relying on our site to provide them with basic information. We feel that when people can read information when they need it, they will obtain a better grasp of what to do to get their data, as opposed to a data services staff member telling them what to do orally. When we didn't have the WWW site, patrons would often come into the Data Services Division time and time again for clarification and further understanding of procedures. This seems to happen a lot less frequently now.

Data Services used to be a low-profile reference point to which patrons were referred when they had already seen the alternatives of the specific subject areas. We have raised our profile over the years on campus however, primarily via our periodically printed newsletter, contacts with faculty members, and our WWW site. The downside of becoming better known is that we are finding that some patrons consider us the first point of contact for their information needs. Given the above identified problems in the section dealing with user needs and our effectiveness, maintaining a separate reference point for data services is now seen as undesirable. A more effective service for patrons should be service integration, which is discussed in the next section.

Lastly, patrons need access to data files and documentation on the campus network. UBC Data Services has made all of its holdings available on several campus unix machines through the use of the Network File System (NFS). NFS allows various machines on the campus network to share our unix machine's mountable disk. Patrons on the other unix machines around campus can then use our files as if they were locally mounted on their own unix machines. The NFS software is also available for the PC environment, but we haven't yet tried hooking our filesystem up to any PC's on campus. We haven't got the disk space for all of our files to be useable at one time in this fashion, but we can put up any file a patron wants usually in 24 hours or less by moving off or compressing other files that are not currently being used and swapping in the new file(s). A number of other institutions are using this approach as well.

This last technique allows patrons access to our data and documentation, but not user-friendly access. Since the unix machines attached to our unix machine through NFS are usually departmental machines and departments are unwilling to spend the people resources necessary to provide their students with access on these machines, undergraduate students generally have no access to our files from the departmental computers. There is a central campus machine that undergraduate students can get accounts on, but currently there is a fee associated with using it. We are using the WWW to allow all our patrons another level of access to our data files, one that bypasses the central campus machine. Details of our WWW access project is covered in following sections.

Integrating Service Points - Shifting Time To Other Needs

UBC, like many libraries, collects much printed and machine-readable information, but we have traditionally divorced their physical locations based on media types. If it was electronic, then it went to Data Services, if print, it went to the applicable subject area. This divorce between electronic and print media is becoming increasingly undesirable from the user's point of view, so we have decided that closer integration between electronic and print media in a given subject area should occur. It can be confusing for the patron who doesn't necessarily see why subject matter is separated into separate library divisions just because one set of material is in electronic form and one is in print form. The patron is more apt to want one-stop-shopping for all their subject needs, so if we can provide this convenience for the user at little cost to ourselves, then we should. By starting their search for information with the subject specialist librarian, if the patron does need to see the data specialist they should be somewhat educated in what they can expect from the institution's electronic and print collections, and what services can and cannot be provided to them. In short, data services staff will see more focused questions, and fewer of them.

A significant portion of our collection deals with electronic versions of materials already held in our Government Publications division in print form, as well as the raw data that go into the making of many reports available there, so this is a natural place for us to begin our integration efforts. There is a growing sense that the front line for library service to the general public, students, and researchers alike for all subject materials, whether electronic or print, should be shifted to the people in charge of the subject section, and away from the Data Services staff. We expect integration with other subject areas to continue as time progresses.

We have begun efforts to educate subject specialist librarians in the Government Publications Division about information products that mirror the print format, and at

least one reference librarian is being trained in microdata in conjunction with the Data Liberation Initiative, a cooperative project between Statistics Canada and research libraries across Canada. It is hoped that the people at the Government Publications Division will take over the basic reference questions from Data Services for government electronic information of all sorts, assisted by our WWW site, which will continue to grow and improve.

As already alluded to, our experience shows that a significant number of new patrons of our data services division often make the mistake of applying directly to us for information when they can be served by that which is already in print. This may be due at least in part to the popular myth that everything is now available "somewhere" in digital form, and one has only to sit down and easily gain access to the entire spectrum of human knowledge from the nearest computer terminal. It takes time to identify what subject the patron is inquiring after and to direct them to resources that might assist them. A shift away from Data Services staff to the subject specialist librarians providing initial contact information will probably get the patron directed to the resources they want, sooner. Even if the patron does need something in electronic form, they may benefit by noting what is in print form. Patrons with complex queries or patrons needing a different level of technical support will still be able to see a data services specialist.

Where To Reallocate Freed-Up Resources

What any individual institution can do for its patrons is a matter of investigation - asking a lot of questions. We have identified our users and decided that their teaching needs were not being met. We are in the process of implementing some of the changes mentioned above, in order to free up some people resources to focus on user requirements.

The following arguments and rationalizations hinge on the supposition that the World Wide Web is available to everyone at the institution, or very soon will be. The UBC campus is very well connected to the Internet and the campus is very nearly completely networked internally as well. This may not be the same state of affairs at all institutions everywhere, but we believe that UBC is close to the forefront of having everyone on campus able to access the WWW via what is currently a reasonably high-speed, high-bandwidth connection. It is the author's belief that currently less well connected institutions will be proceeding as we are proceeding in the years to come, with greater connectivity and WWW access for everyone as the cost of providing this service declines or is met by an internal reorganization of priorities.

If the Data Library is eliminated as a separate service point for reference, we will immediately free up some of the Data Services staff time as they will not be required to spend as much time with patrons as before, but the line-ups at the

main Reference Desk may get a little longer. The trade-off is that a fair bit of time savings may be realized by the Data Services staff. The question then becomes: How best then to make use of these freed up resources?

One use for the time of the former Data Services staff would be to coordinate projects within the institution to provide patrons with WWW access to electronic data files. This would require liaison with both research and teaching faculty, to see what sorts of things could be useful and what is required for themselves and their students. Many data files that data libraries deal with have either no interface programs that assist users in using them, or they rely on some kind of generalized software, like SPSS, SAS, or SHAZAM in order to extract desirable subsets of the original data files. Many of the data files housed in traditional data libraries could have surprisingly wide-ranging uses if access would become somewhat easier. Many surveys of a governmental origin, be they census data or surveys dealing with the health of the population or incidence of political opinion may be of interest to those scholars in fields traditionally outside the fields in which the surveys were taken. Non-traditional and traditional data users alike need easy access to the Data Library's data and documentation, and only through the WWW is this likely to happen, at least until something comes along to supplant the WWW.

Another proposition would be to assign the former Data Services staff to the function of WWW publication needs. This is not incompatible with the above goals of WWW access to the current data library machine-readable data files and documentation. The reasons for diverting resources to WWW publishing of library information are two-fold:

- The WWW is becoming an important method of mass dissemination of information in electronic form.
- The WWW is increasingly being used in teaching, for distributing information back and forth.

Many libraries would like to get resources directed towards having some sort of presence on the WWW, and the freeing up of Data Services staff from reference work should provide both badly needed computer expertise and people time in this critical area. Many libraries are facing the reality that they need a WWW presence, but that they do not have the resources to do a good job in this area. Most institutions depend on a very small number of dedicated, and often quite separated individuals to take the bull by the horns and do what they can as far as creating a WWW presence for their institution or division. What usually comes out of this approach is a patchwork solution, where each branch and division goes their own development way, some looking good, some looking poor, some updating their information frequently, and some, especially if a key individual leaves, not updating information for long periods

of time.

A group dedicated, or at least partly dedicated, to providing at least a coordinating/helping role in publishing on the WWW for the organization would go a long way towards helping out the organization. If well designed, WWW sites can be used for many types of general reference questions that patrons may have, thus reducing the waiting time in the line-up at the reference desk that may have built up temporarily when Data Services staff stopped doing reference work. Patrons can be directed to specific areas of a WWW site and be instructed to go through what they find there. This is not an argument for centralized control of all WWW publication for all divisions and branches within a library. The author believes that more of a guiding and nurturing role can be played by the WWW publications group, and that this will lead to better WWW sites and more dedicated involvement in developing the WWW sites by many more staff.

In direction of making better access to data library data and documentation, the Data Services staff at UBC have developed a prototype system for allowing access to data files and documentation in a friendly, easy to use manner, utilizing the WWW. We hired a student programmer for a summer and had him program a WWW data extraction interface to the data file and documentation for the **1993 Survey of Consumer Finances - Individuals Aged 15 Years and Over, With and Without Income** survey from Statistics Canada. This type of work is not necessarily new, others are involved in the same sort of enterprise, however we have gone to great strides to present the information in as friendly and user-obvious a manner as possible. We took the codebook as produced by Statistics Canada and marked it up in HTML. This was a fairly time consuming and tedious process, as the data vendor sent us the codebook in WordPerfect format and we had no way to use this file format directly on the WWW. There are now new and better conversion tools available to convert the popular word-processing formats directly into formats for the WWW, and this can only get easier. The newest and best tools preserve both tables and graphics from the word processor output, in HTML.

The patron sits down at any Internet accessible computer on the UBC campus with Netscape loaded onto it and selects a number of variables and cases from either the basic WWW form or the complete one. The basic form was envisioned to be used mainly by beginning students, and is used by instructors in courses introducing students to numeric survey data and it's analysis. Via the basic extraction form it is possible to only get a limited number of the most popular case selections, as defined by the teaching faculty, though full access to all variables is possible. The complete form of course provides access to all cases and variables, for researchers and others who may need access to more case selection possibilities than just

those provided by the basic form.

After a patron completes the data selection forms, a CGI-program on our Data Library server writes an SPSS job file to the filesystem in a special directory. Periodically throughout the day, the SPSS program is run if there are job files waiting for it in this special directory. When the SPSS job completes, the user is notified via e-mail, and the output of their job is written to the anonymous ftp directory of the filesystem. We decided to limit patrons to a certain number of requests per day, as well as a maximum overall disk utilization per user. So far, we have received no complaints regarding these limits, though they were chosen arbitrarily.

The feedback we have had from both teaching staff and researchers has been enthusiastic and very encouraging. We intend to do more. The author is now engaged in attempting to generalize the C code used in the system, as the current code is designed around the one survey that it supports. When completed, the system should be capable of adapting new survey files and their codebooks. These should be able to be added by a non-programmer. The concepts of resource classes and objects are being used in the coding to make other extraction systems besides SPSS work in the background for data extraction. A resource is in this context an extraction system, and an object is a data file and it's associated documentation. In theory and the author hopes in practice, the patron will not be able to tell how the actual objects are stored, nor what resources are being employed to act upon them.

* Paper presented at the IASSIST/IFDO 1997 Conference, May 6th May 9th Odense, Denmark