

UNIVERSITY OF GUELPH

***DATA LIBRARY/CENTRE
(DLC)***

PROPOSAL

(Only portions of original document appear here)

Prepared for the Statistical Advisory Committee

September 1996

Prepared by

SAC - sub-committee on data library issues

Bo Wandschneider - Economics

Peter McCaskell - CCS

Doug Horne - Library

ACKNOWLEDGEMENTS	<u>v</u>
SUMMARY	<u>vii</u>
1. INTRODUCTION	<u>2</u>
2. BACKGROUND	<u>4</u>
2.1 What is a DLC?	<u>4</u>
2.1.1 Levels of Service	<u>8</u>
2.1.2 Considerations for a Computing Environment	<u>10</u>
2.2. What are the benefits of a DLC?	<u>11</u>
3. WHAT SHOULD GUELPH DO?	<u>14</u>
3.1 Interim Solution	<u>16</u>
3.2 Proposal 1 - Guelph Alone	<u>17</u>
3.2.1 Initial startup costs (Guelph alone)	<u>20</u>
3.2.2 Yearly Costs (Guelph alone)	<u>21</u>
4. Other Institutions	<u>23</u>
4.1 Calgary	<u>23</u>
4.2 UBC	<u>23</u>
4.3 University of Toronto	<u>23</u>
4.4 Carleton University	<u>23</u>
4.5 Western	<u>24</u>
4.6 Waterloo	<u>24</u>
4.7 University of Alberta	<u>24</u>
4.8 University of Victoria	<u>24</u>
4.9 Queen's University	<u>25</u>
APPENDIX A: Job Descriptions	<u>27</u>
Data Librarian	<u>27</u>
Data/Technical Analyst	<u>28</u>
Clerical	<u>29</u>
GTA	<u>30</u>

APPENDIX B: Conditions for DLI membership	31
APPENDIX C - List of files in DLI	33
APPENDIX D: Sample Job Descriptions from elsewhere	42
Data Resource Librarian	42
Senior Research Associate	45
Coordinator of Electronic Data Services	48

ACKNOWLEDGEMENTS

We would like to thank all those people who patiently described to us how they handle electronic data at their institutions. The experiences they shared with us were invaluable. We realize that it is sometimes difficult to find time but everyone approached was extremely generous and it is very encouraging to see this kind of cooperation.

Many of the ideas in this document come from discussion lists for the Data Liberation Initiative (DLI) and the Canadian Association of Public Data Users (CAPDU). Contributions come from many different fields from across Canada. Special mention should be made of the contribution from the ICPSR summer workshop put on by Dianne Geraci (SUNY Binghamton), Chuck Humphrey (University of Alberta) and Jim Jacobs (UC San Diego).

Background for this report was also generated from visits to the University of Toronto and the University of Western Ontario. Information gathered from an 1993 survey of CSS faculty, looking at their needs and uses of electronic information, was also used.

SUMMARY

At the request of the Statistical Advisory Committee (SAC) a working group comprised of Doug Horne (Library), Peter McCaskell (CCS) and Bo Wandschneider (Economics) was established to look at the issues related to gathering and disseminating electronic data at the University of Guelph. Although there have been significant changes in technology and long-standing awareness of data needs at Guelph, the real impetus was provided by the recent changes to the dissemination of Statistics Canada data, through the Data Liberation Initiative (DLI). It has been clearly stated that Universities wishing to participate must establish some formal procedure for acquiring and disseminating this information to the research and teaching community

The following document presents an interim solution and outlines two separate proposals. The first proposal describes a process in which Guelph would establish their own centre, while the second proposal expands this to include The University of Waterloo and Wilfrid Laurier University. The later proposal is consistent with other models being developed across Canada and it follows from other cooperative programs between the institutions. There are clear advantages in terms of costs and level of services associated with the joint proposal.

The working group strongly supports the idea of working jointly with the other two institutions. As the data culture in Canada evolves there will be a movement towards centralizing the dissemination of information in the form of one centralized point of service or possibly several specialized regional centres.. As an example, Great Britain has established five regional centres to create a 'virtual library' to disseminate electronic information. In Canada we are already seeing small regional groups of Universities sharing resources. A group comprised of Guelph, Waterloo and Wilfrid Laurier could form a strong base that could evolve into a regional centre in Canada.

There is however, concern with the immediate needs of the community. As such, the group recommends that Guelph proceed with their own DLC and when the details of a collaborative centre can be worked out we could simply role our own DLC into the joint one. No resources would be wasted and we would be developing much needed expertise in the interim.

In both proposals the Data Library/Centre (DLC) would be associated with the library, providing such things as identification, acquisition, storage, documentation, cataloguing,

system file generation and consultation. It would fall below what is termed a 'full' service facility. This implies that statistical consulting will be limited and archiving will not be undertaken to any great extent. The option to provide these in the future would be left open. Currently, the Ashton Statistical Laboratory and the CS Data Analysis group can assist in providing statistical consulting. The joint proposal would expand slightly on these services, moving more towards a 'full' service facility. The real benefit is that resources are shared and this will lead to a lower cost per institution.

The centre would house full-time staff, occasional contract workers, graduate students, equipment, the data collection and documentation. The initial set-up cost for the facility would be approximately \$160,000, while the ongoing yearly costs would be approximately \$79,000. The joint proposal would cost Guelph, \$57,000 to set-up and \$44,000 per year thereafter. It is important to note, these are not totally new expenditures but are partially a redistribution away from individuals or departments to the central DLC. There would also be a redistribution of funds from within, as staff and GTA's are seconded to the DLC. Taking into account the current financial situation it is going to be very difficult to solicit new funds. As such, it must be strongly emphasized that there is a redistribution as we re-prioritize our spending. Although there will be savings in terms of duplicated effort and decreased expenditures on data these are in no way easy to quantify. With the current ad hoc system of research and acquisition it is essentially impossible. Consideration must also be given to the fact that the University must remember that it has a mandate to build and improve upon what it does. As such there is a continual evolution in all areas, that will require administration to redistribute the University's priorities and resources.

As background information the document outlines the functions and benefits of DLC along with how other institutions are progressing. It comes to the conclusion that a DLC has become an essential part of any institution that hopes to excel in teaching and research. Therefore, a commitment by the University to restructure the current support system for disseminating information to researchers and students is essential.

It should be noted that benefits from increased and improved research, improved teaching and financial savings would be significant for at least four separate colleges on Campus, including CSS, ARTS, FACS and OAC. Individuals from across these groups indicated that the need for a DLC is not new, but has existed for quite some time. Although the initial proposals are driven by a need to disseminate data associated with the DLI it is likely that the skills and expertise developed in this task can be disseminated to other areas of teaching and research on campus.

1. INTRODUCTION

An earlier version of this proposal was first developed in the fall of 1993. Since that time there has been a great deal of discussion on and off campus about these issues. In 1993 there were very few institutions in Canada which had a DLC¹. Since then, there has been an explosion in the establishment of these facilities. Most institutions comparable in size to Guelph have either implemented or are in the process of implementing a DLC. Several reasons can be cited for these developments, including, significant changes in the availability of electronic information (DLI) and technology (for example the wide spread adoption of the Internet and tools associated with the WWW). It is a very exciting time to be a researcher, or student, who can benefit from this new freedom to access electronic information.

These issues are not isolated to Canada. Faculty, students and researchers at institutions across the world are being faced with major changes in the way they go about their everyday tasks of teaching, researching and learning. Society is continually showing a greater dependence on information. With the enormous advances in computer technology over the last decade there has been an increasing volume of data being disseminated electronically. A large portion of this data was previously not distributed and was not universally available, if at all. At the same time, another portion of data is being solely disseminated electronically. This presents new challenges in terms of information management²

This opens up many new and exciting opportunities for those able to take advantage of it. Canada has always been a great collector of information, developing strong central institutions such as Statistics Canada. The problem has always been that they haven't been great disseminators. The resources, in terms of skills and finances, needed to

¹ An informal survey conducted in 1990 at the University of Alberta showed that at this time there were already 11 Universities in Canada which had a policy in place related to data libraries or data resources. Some of these were very crude but the issues were being addressed. Information is currently being gathered to update this survey.

² The driving force behind these proposals is the need for an infrastructure to deal with the DLI. A great deal of information that used to be distributed to depository libraries is now distributed through the DLI and they have a clear set of conditions that must be met for participating institutions. (See appendix B). For a list of available files see appendix C.

access the information were very difficult to come by. The financial barriers are beginning to fall, as are the required skills. It is up to the research community to provide the resources to make this information available locally, at least in the short term. In the long run we may see a centralized dissemination service but this will take time as we define our needs and overcome the incompatibilities imposed by the different platforms across institutions. It should also be noted that there are opportunities for this institution to become a leader in the field of data dissemination. The infrastructure and human resources exist. They just need to be applied.

In order to take full advantage of this new information flow there needs to be a restructuring of the support functions provided to faculty, researchers and students. This requires a commitment of resources from the University. Whether these resources are merely a redistribution within the institution which may just reflect a healthy, natural and dynamic evolution of our priorities, or whether new resources are found, it is clear that without this restructuring the University can not hope to maintain its current position in terms of both research and teaching.

The rest of the document will outline what a DLC is, what benefits are associated with it, some options for Guelph and a summary of what other institutions are doing.

2. BACKGROUND

The following section will give some background information on the different options available for the establishment of a DLC. The term DLC can mean many different things to many different people. It is usually the case that budgets will dictate the level of service provided. There is no doubt a demand for **ALL** the services outlined below. The challenge will be to balance the costs and benefits within a DLC and across the whole University community.

2.1 What is a DLC?

A DLC can take many different forms depending on the priorities and budget considerations of the institution. In very basic terms it performs the functions associated with **the identification, acquisition and dissemination of machine readable data**. This electronic information can arrive in both hard and soft media. Hard media includes magnetic tape (in several formats), CD-ROM, and diskettes, while soft media refers to files accessed through ftp sites, web sites, bulletin boards, gophers, commercial services and such. In some cases DLC's go so far as to perform the task of converting hard-copy data to an electronic format.

Many of the roles associated with a DLC are extensions of those currently provided by libraries. As such, many DLC's are located within a University Library although this is not always the case. There are several examples of DLC's operating within Computing or Social Science Departments³. The trend is definitely towards having a DLC within the library.

The major difference between a DLC and a traditional library is that the information in a DLC is no longer in printed format and the staff responsible for the resource need skills not traditionally associated with librarians. The equipment needs are also significantly different and libraries have traditionally not budgeted for these items. This can lead to

³In very general terms, experience suggests that the bulk of users originate from either Social Science or Business Departments. Specifically, this includes Business Administration (Commerce), Economics, Politics, and Sociology, although there is increasing demand from the other Social Sciences (particularly Geography and their GIS applications), Arts, Family and Consumer Studies, OAC, Medical Sciences, Chemical Sciences and beyond.

significant problems on its own.

Figure 1 gives an interesting perspective on the current problems faced by traditional librarians. It represents, from a librarian's point of view, the paths followed by typical researchers when trying to locate a piece of information. Traditional libraries should be able to effectively deal with the questions in the unshaded boxes while they would probably have difficulty with those in the shaded areas.

Currently the library must answer NO to the question on ... *in computer readable form*, or use an ad hoc procedure to send the requester to individual(s) who have access to data sources on campus.⁴ These individuals come from a broad range of departments all across campus and many times one researcher needs to contact more than one source. This can become a very frustrating process.

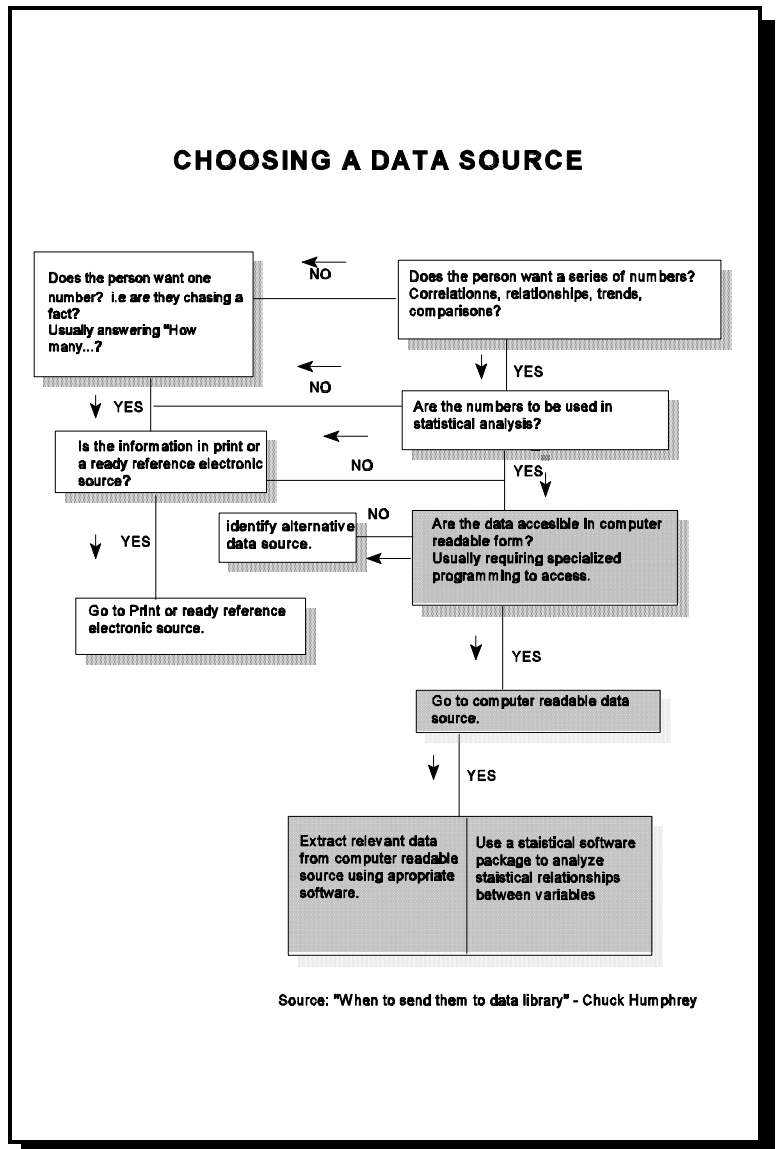


Figure 1

⁴It should be noted that this ad hoc referral is totally inadequate. The results are very inconsistent from one user to the next and there is no formal procedure for informing library staff where the information is located.

To fully understand the functions of a DLC it is easiest to give examples of the services that may be expected by data users and supplied by support staff. A series of tables will be presented outlining the following topics:

- i. **the general expectations of the users;**
- ii. **summary of how a library may meet these expectations;**
- iii. **details on reference services for data;**
- iv. **details on computing services related to data;**

After this there will be a summary of services presented in *Figure 2*, and examples of how they can be integrated into a DLC to provide a facility that offers various levels of service. The level of service will be dependent on the needs of individuals and the resources available.

User Expectations

- **Pre-Acquisition Services**
 - receiving requests for data
 - identifying source, format and costs
- **Acquisition Services**
 - determining appropriate medium
 - receiving and processing codebooks and data
 - record keeping and billing
- **Data Access Services**
 - informing users how to use and retrieve data and codebooks
- **Basic User Services**
 - help with interpretation of codebooks
 - explain formats such as system files, rectangular, cross sectional
- **Computing Services**
 - advice on best platforms for jobs
 - advice on code for reading data
- **Data Analysis (general)**
 - statistical and retrieval advice
 - interpretation of results
- **Data Analysis (expanded)**
 - handle requests for everything (ie. from input to final results)

General Library Services

- **Passive Referral**
 - respond to requests for data already in existence on or off campus
- **Active Referral**
 - promote collection through instruction, seminars, newsletters ...
- **Cataloguing**
 - create a central listing of campus data sources
- **Acquisition**
- **User Services (consultation)**
- **Archival**

Reference Services

- **File Identification**
 - respond to ... do you specifically have XYZ files?
- **Basic File Recommendation**
 - respond to ... do you have stats on income or ...?
- **Advanced File Recommendation**
 - respond to more detailed or obscure questions than above.
 - Usually related to specific survey questions.
- **File Use Advisory Services**
 - recommendations on what is and is not possible with specific files
- **Extraction Services**

Computing Services

- **Data Storage**
- **Copying and Subletting**
- **Data Retrieval**
- **Data Analysis**

There is an enormous array of tasks that are required by the users of MRD (machine readable data). The above tables only summarize these activities. Each item can be expanded to a multitude of tasks.

2.1.1 Levels of Service

Figure 2 summarizes the different functions into three groups, **collection care, user services and archives**. The type of DLC that is established is usually determined by the level of service provided. Service level can be simply grouped into primary, secondary and full. Generally speaking, primary is normally concerned with collection care, secondary adds user services and full deals with the archives; although it not always that clear-cut. An example would be the following:

Primary:

The primary, or minimal level of service would include the following. The DLC **collects currently held data resources** from across campus and **provides storage in a central location**. It also begins **acquisition from major sources** such as ICPSR and DLI, **provides access to codebooks and data**. There should also be a **hard copy inventory** compiled. Note that currently Guelph has not even met the primary level. Attempts have been made to store information centrally but due to limited human resources and an inability to express the needs of the user community only certain data sets have been supported and it has been extremely

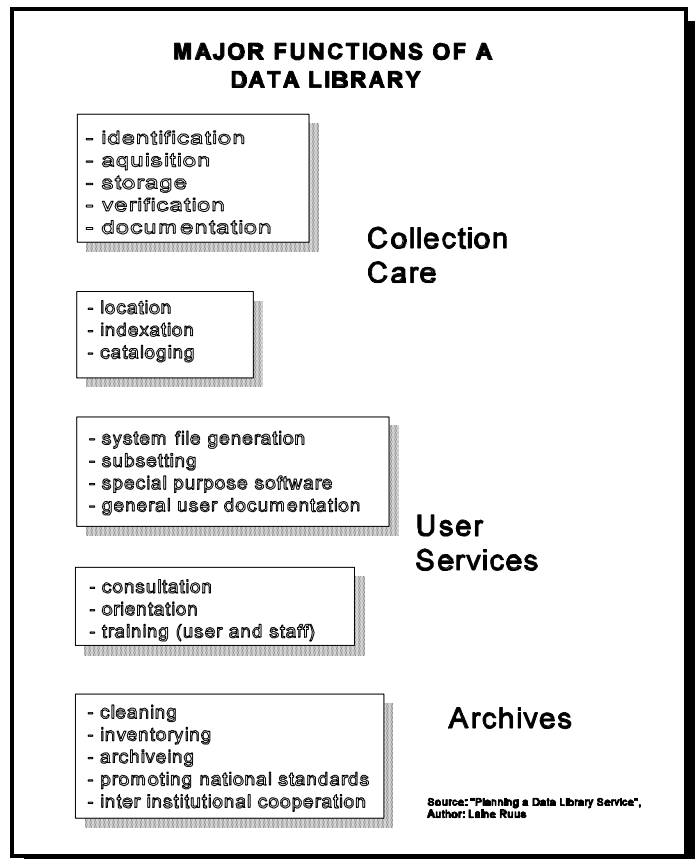


Figure 2

challenging to maintain what is already there. It should be noted that in terms of available hardware Guelph is already ahead of many other institutions and there should be no difficulty in implementing even this level of service. However, this is only a minimal level of service and it would not even begin to bring new researchers on-line, or would it take advantage of the new opportunities available to teachers. This still requires individuals to have an advanced understanding of computing. The last 20 years have told us that there is a large cost to obtaining this understanding. This in turn implies that not very many people do.

Secondary:

In addition to what is described above the DLC will maintain a close **liaison with research groups** on campus in order to **anticipate and order future data resources**. They will also keep **up-to-date information on other sources of data and other data archives**. Along with information on local holdings they will **publicize this information**, possibly in the form of a newsletter or a web site⁵. They should also consider providing: an **search engine for data holdings** that facilitates **key-word searching**; **develop or purchase interfaces** for accessing more commonly used data sources; **create system files** for accessing data less used or not suited for specialty interfaces. **Staff training** for traditional library staff, to keep them informed of what the DLC is doing, is also essential to running an effective DLC.

Full:

A full service DLC also **cleans and re-documents original data files**, provides **special purpose subsets**, including documentation, for such activities as teaching, **conducts seminars with experts in the field** as well as **in the classroom** covering such things as access to the collection, use of statistical packages and research methods. The DLC would also conduct **statistical analysis**, and provide **specialized consultation** on topics such as statistical methods, survey methodology, time-series analysis, cross-sectional analysis, etc. They can also be involved with **planning and conducting data collection projects**, which in turn

⁵ There is a sample web page available outlining some of the potential options for putting a DLC on the WWW (http://www.css.uoguelph.ca/ugdl/intro_ugdl.html). This is very preliminary and not all functions work. The retrieval example is driven by a test machine in CCS. It is the only machine with both SAS and HTTP. Several other institutions are using the WWW as a medium for establishing a DLC.

may lead to **setting up a true data archive**. True data archive in this sense means that the community beyond the University has access. In the extreme, researchers could submit projects to the DLC which would in turn conduct the empirical research.⁶

It should be noted that these are just general guidelines of what a DLC may look like. In reality most of them are a combination of the above examples. This document proposes that Guelph establish a DLC that falls somewhere between secondary and full service. Some of the tasks associated with the full service come at a large cost and are not likely justified for an institution of this size. If a joint proposal is undertaken between Guelph, Waterloo and WLU then the service could evolve to more of a full service facility.

2.1.2 Considerations for a Computing Environment

Any solutions proposed for Guelph should be tailored around our current computing environment. Guelph has a strong centralized computing environment built around powerful Unix servers. Any proposal should take advantage of this⁷. There are many different models out there but the general trend has been towards building DLC's around Unix platforms. This gives users **access to tape support, distributed files systems and large scale statistical power**. Small data sets can be handled on a Unix system, however personal computers still have problems working with large data sets like the CENSUS, SLID and LMAS. It makes most sense to have data stored and shared centrally and CPU cycles to be distributed in the same fashion.

Any proposal must provide mass storage with quick, multi-user access. CD-ROMS

⁶If this does occur there is usually a fee structure established where researchers can hire consultants from the DLC at a given hourly rate. Clearly this would only work on a small scale. If the research project is large, it is likely more efficient to hire researchers specifically for that project. It is also not usually considered the mandate of a DLC to perform these tasks, although they are usually requested to do this at some point.

⁷ It is very important that a DLC works closely with the central computing services people and the Ashton Statistical Laboratory. There are numerous examples across Canada where frustration levels between researchers and central computing providers has been very high. The lack of understanding between the two groups is surprising. Calgary is an example where the two groups work well together. Guelph has an infrastructure in place that could deliver first class services if the two groups have a better understanding of what the other does and what the priorities are.

and Unix tape drives are very slow and not well suited for a DLC. A centralised system that allows servers across campus to NFS mount file systems works very well. The current changes to our physical network structure also facilitates this solution. Access to powerful CPU's for running large statistical jobs is also essential, but need not be part of the proposal.

The following section will discuss the benefits of establishing a DLC.

2.2. What are the benefits of a DLC?

The establishment of a DLC can provide the University with many benefits. They can be summarized into three broad categories that include:

- **improved and increased research**
- **improved teaching**
- **cost savings**

The current situation at Guelph is best described by the chart in Figure 3. This can be used to help explain the potential benefits⁸. There are three distinct groups on campus that handle electronic data. These include CCS, the Library and the researchers / teachers (or user community). This latter group can be divided into group A and B, the distinction being that group B has funds to purchase data while the other does not. The result is that data flows into either group B researchers or the Library.

⁸It is clear that the situation is not this simple. There are many different types of researchers on campus. The simplification is used to illustrate the problem.

The flow of information between these groups is very ad hoc, as represented by the dashed lines. The information flow between the library and CCS has historically been particularly limited⁹. There is very little communication between the two groups of researchers, except indirectly between the other two groups. This leads to a situation of asymmetric information or 'haves' and 'have-nots'.

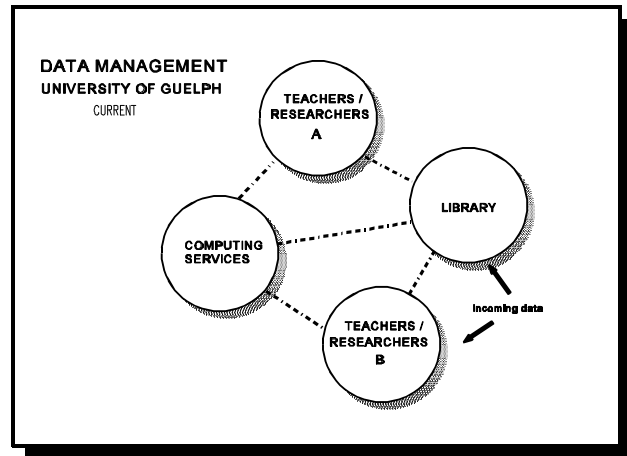


Figure 3

As a hypothetical example, researchers in group A may be considering a new project that requires they obtain one of Statistic Canada's data surveys. Assume it is not available from the DLI. They may find that they are unable to solicit the funds needed to purchase the data. As there are no formal paths for disseminating information they may not find out that group B researchers already have the data and are quite willing to share the information¹⁰. The result is that they either abandon the project or eventually solicit the funds to purchase it themselves. In the latter case the University ends up having two licences for the same data set.

This simplified example clearly shows that by acting as central disseminator of information and increasing the information flow a DLC can lead to **increased, and possibly improved research, improved teaching and an decrease in the waste of financial resources used for the purchase of duplicate resources.**

⁹ Part of the purpose of this proposal is to alleviate this. A DLC would be extremely effective in opening up the lines of communication by providing staff that have the skills capable of bridging the gap.

¹⁰ Data resources purchased or generated from grants such as SSHRC and NSERC belong to the University, not the individual researcher. This is an area of concern as not all individuals realize this. It is also normally a stipulation that generated data sets are deposited with a recognized Archive. It would seem that the University has some responsibility for establishing an archive for researchers to use.

There are several other examples of cost savings to the University through a DLC. As an example it must be assumed that the staff of the DLC will, over time, obtain a very specialized skill set that will allow them to process and present incoming data sources in a very efficient and effective manner. This implies that **the researcher will waste less of their own resources** in processing raw data sets and extracting the necessary information.¹¹ The **DLC can perform a large part of the processing once, and distribute it to the entire user community.**¹²

Along the same lines, the researchers or students will spend less time in their search for information, assuming the DLC has a central location. This also implies a potential **improvement in the quality and quantity of information that the user is capable of finding.** This in turn implies improved research and an improvement in the quality of work done by students.¹³

Another important point to consider is that **a portion of the potential user community will never use electronic data unless there is a DLC** that can provide them with the knowledge they need to get involved. There is a significant cost involved in developing the necessary skills to use electronic data. For some individuals this is a barrier to entry. The DLC simply decreases this cost and allows more people to get involved. There are numerous examples of faculty within within the CSS who have expressed research or teaching ideas but do not have the necessary skills to deliver them.

¹¹ Assuming the researcher has some tangible resources freed up after a DLC is established, it is an interesting problem as to what they do with these resources. It could be argued that they should be entitled to use them in other parts of their research to improve the quality of their work. The problem is that there is a redistribution of spending from the individual to the University. As mentioned, the DLC provides the service more efficiently, so there is a pareto improvement and the question is; should the individual use a portion of their benefits to compensate the University for establishing the DLC. What this compensation should be is difficult to calculate.

¹²Interfaces, such as the one used for CANSIM, allow the researcher easy search and retrieval options that are far less costly than what would be required to work with the raw data sets. The sample given on the UGDL web site shows the potential for something like the General Social Survey.

¹³Consideration should also be given to the quality of students that are graduating from Guelph. Clearly this is in the mandate of the University. Students entering the work force are faced with the reality that society is becoming increasingly dependent on information. The individuals most capable of dealing with this information are the ones who will succeed. The experience they could gain at a university having a DLC could give them a significant advantage.

Cooperation is another benefit from establishing a DLC. One of the concerns among the research community is the fact that Statistics Canada has significantly increased the prices of their products over the last few years. Products such as the CENSUS were priced beyond the reach of even the biggest institutions in Canada. To deal with the situation the CARL consortium was established to negotiate a group purchase of the CENSUS at more 'reasonable' prices. This has recently evolved to include more products, culminating in the DLI. Guelph was able to participate in the process and continues to participate through the library. The problem is that our participants do not have a clear understanding of the needs of the electronic data community, both in terms of what they want and what is required when the data arrives. **Guelph is also in a 'response' mode** in the sense that we wait and see what is going to happen next. We are not driving the research. **Preminent research does not just respond to what is out there but instead develops the questions and stimulates the collection process.**

A DLC could take a much more active role in future consortiums. There are enormous **benefits to be gained from cooperation between institutions.** The expansion of the Internet opens up endless options. Guelph needs to be committed to developing electronic data resources in Canada. We can not afford to sit back and wait to see what other researchers decide is worthwhile. We need to go out and express what our needs are. We need to drive the process at times. Cooperation within the University is also an important consideration. As mentioned in the hypothetical example, different groups on campus may be interested in the same data. The DLC could act as an intermediary bringing interested groups together **developing more inter-disciplinary research and teaching.** Purchases that could not be justified by single users may now be justified and, possibly, funds can be collected from various sources. The net result is an increase in data resources on campus.

The above examples only begin describing the benefits of establishing a DLC and as such they underestimate the potential benefits that can accrue to the University. To actually quantify these benefits is an almost impossible task. Suffice it to say that **the benefits are positive and significant.**

The next section outlines the costs involved in running a DLC using a specific proposal for Guelph.

3. WHAT SHOULD GUELPH DO?

The following section outlines an interim solution and two separate proposals. **Although the group strongly supports the idea of a joint proposal there is some concern about the immediate needs of the user community.** Any solutions developed under the 'Guelph alone' proposal could easily be rolled into the joint proposal once the details of a collaborative effort are worked out. Both proposals can be developed simultaneously without being detrimental to the overall objectives

A good deal of expertise and knowledge of our needs can be gained through the development of our own facility. This is similar to what is happening at the national level. There is potential for Statistics Canada, or some other centralized group, to take on the role of disseminator and developer of DLC type tools. Instead, we see that individual institutions develop their own facility at the 'grass roots'. It appears that they are trying to get a handle on what their needs are. Cooperation exists, but there is also a void in knowledge about what we should be cooperating on.

The working group also wants to emphasize that, under either proposal, there is a need for a physical location. Although the intent is to provide access to the information through many remote sites on the network there still needs to be one location that users identify with a DLC. At this site users should find access to computing equipment, data codebooks, expertise in using the data and expertise on retrieval systems.

The establishment of a facility such as this is a very challenging one. **There is a need to put in charge an individual who is highly motivated and has a very strong understanding of the needs of the user community.** The first few years are the most critical. If things are not put in place with long term considerations the costs may outweigh the benefits. The potential for putting a system in place that frustrates the teaching and research community is very real. Due to budget and staffing consideration there may be an incentive to second people on a part-time basis. The working group feels this would be detrimental to the process and suggest below that **individuals be seconded for a 2 year term in order to get the proper commitment and continuity necessary to establish a functional DLC.**

3.1 Interim Solution

The University is presently paying yearly fees to have access to products from ICPSR and DLI. **Users have already expressed interest in acquiring information from these services.** It is thus necessary to have an interim plan to meet these needs. The solution presented below is very ad-hoc and is designed to bridge the gap between now and the implementation of a true DLC. It is hoped that this can be accomplished within this calendar year. The individuals involved have volunteered to take on these responsibilities for a short period of time with the stipulation that their time commitment can only be small. Inevitably this will lead to problems if there is a large demand for data. This solution also falls well below that outlined in even the primary DLC. There are not enough resources to develop any retrieval systems, advertise, catalog or even keep adequate usage statistics. **If this interim solution is left in place for an extended period of time then the process is not being given a fair chance to succeed.** We would be better off not even attempting this as we would end up wasting valuable resources and causing a lot of frustration.

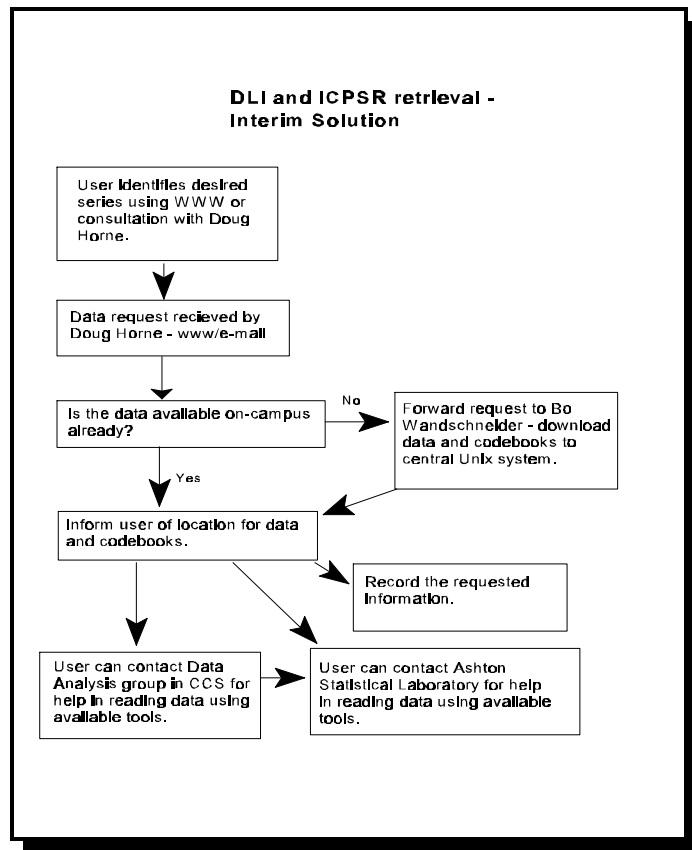


Figure 4

Figure 4 contains an outline of the steps a request for information would follow. An attempt will be made to expand the sample web page and

use this as the centre of a retrieval and distribution network. **All requests for data will be sent through Doug Horne in the Library.** The web page is designed to have fill-in forms for data and information requests. The section on form access to data will not be

expanded as it is very time consuming and there will need to be a commitment of human resources. Only data from the DLI and ICPSR will be processed. Once a request is recieved a check will be made to see if the data is already available. If it is available it will be located and this information will be passed on to the user. Generally the data will reside on the central Unix system and the user will be provided with a codebook (hard copy or electronic). It will be up to them to develop a program to process the data. They may then **contact Peter McCaskell of the data analysis support group in CCS or William Matthes-Sears of the Ashton Statistical Laboratory to get assistance in developing a program to read and analyse the data.**

If the data is not on-campus then the request will be passed onto Bo Wandschneider who will download the data from the appropriate site and upload it to the central Unix system. The current directory structure used by the Economics Department will be expanded and possibly moved to a more central location. If codebooks and sample programs are available they will also be obtained. The user will then be contacted and a similar proceess to that above can be followed.

There are several more details, and the 3 people involved are working through them as requests come in. This entire process should be considered a learning experience and good preparation for the actual DLC.

3.2 Proposal 1 - Guelph Alone

Guelph should establish a DLC that is associated with the Library. The DLC would maintain a collection of information that should be considered complimentary to the traditional library collection. **Researchers and students should be given a 'one stop shopping' option** when gathering information. The distinction between a book containing tables of data and a

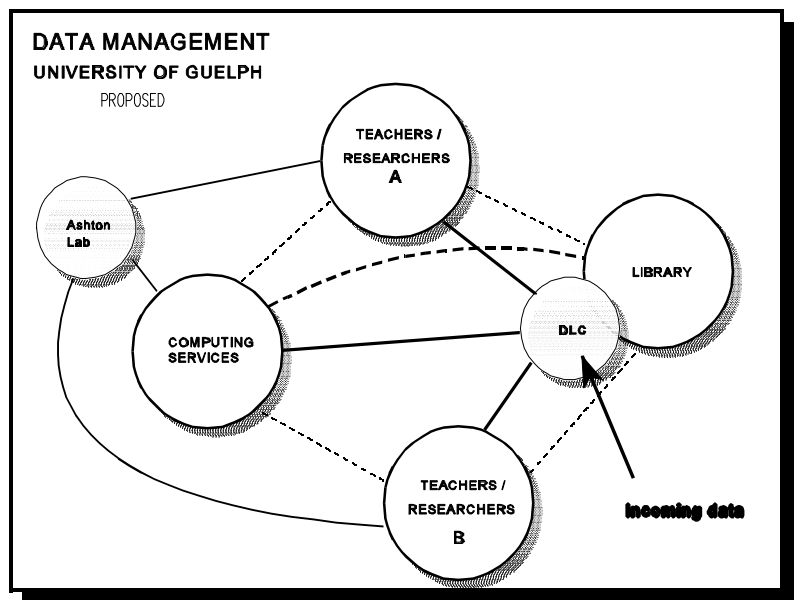


Figure 5

computer file containing data is essentially the same as the difference between a book and microfiche. The only real difference is in the skills required to use the different media. It also makes more sense when you consider what is happening in the world around us to bring computer skills to the library rather than library skills out to the computing people. The fact that the potential user community spreads over several colleges also suggests that a centralized facility is the best option and precludes models that set up centres within one or two colleges.¹⁴

The DLC should maintain autonomy having a physical location that users can identify with it. It is not a good solution to set up access from reference stations. There should be direct contact between DLC staff, students and researchers. There should also be a separate budget to run the facility. Specifically, there should be a room located on campus that has space to handle 3 (possibly 4) staff members, hardware, consulting and a collection of codebooks and data. Ideally, this room would be divided in two. One section would be devoted to staff workspace while the other would be devoted to consultation and access to the collection.

The CSS has offered space on the 8th floor of the Mckinnon building. In terms of size and location this room is very well suited for housing a DLC. The room is large enough to house staff, as well as providing room for consultation and equipment. It is also centrally located on campus, as well as being in the same building as the largest single block of users (CSS researchers and students). There is also easy access to networking facilities

The library is in the process of planning space for a DLC. Provided that the space is large enough this would also be a good location for a DLC. Having it close to the documents centre would be an efficient solution. One consideration may arise when considering the joint proposal. There has been concerns raised about the DLC being perceived as belonging to one of the three institutions rather than belong to the whole. Having a DLC in one of the libraries may add to this concern. **By keeping the facility removed, and on its own, it may give a greater feeling of being a facility equally shared and managed by the three institutions.**

The facility would be staffed by three fte's, contract workers and graduate students.

¹⁴Refer to Section 4 for some brief descriptions of models from other institutions.

The initial set-up of a DLC is a very difficult task, especially considering the field is new and information is limited. **It will need to involve someone who has a very strong commitment to making it work, and keen understanding of what is required.** Their mandate should be clearly to the DLC and not some other group. As such, the group feels that CCS and the library should make secondments to the DLC for an initial 2 year period. This would give the individuals time to get the process rolling, set some objectives, follow through on them, and set some long term plans. The demands during the first two years also suggest the need for full time secondments rather than part-time ones.

The three positions would include a **Data Librarian, a Data/Technical Analyst and a Clerical worker.** Refer to appendix A for a detailed job description of each position. There would also be an allocation in the budget that would allow for the hiring of contract workers as the need arises. This would be especially necessary during the initial set-up as all of the 'old' data is brought on-line. There may also be additional demands when projects such as the census are undertaken. The workload will clearly fluctuate within the year, as well as over the years. Staff in other parts of the library as well as some in CCS would also be trained to deal with users requesting information from the DLC.

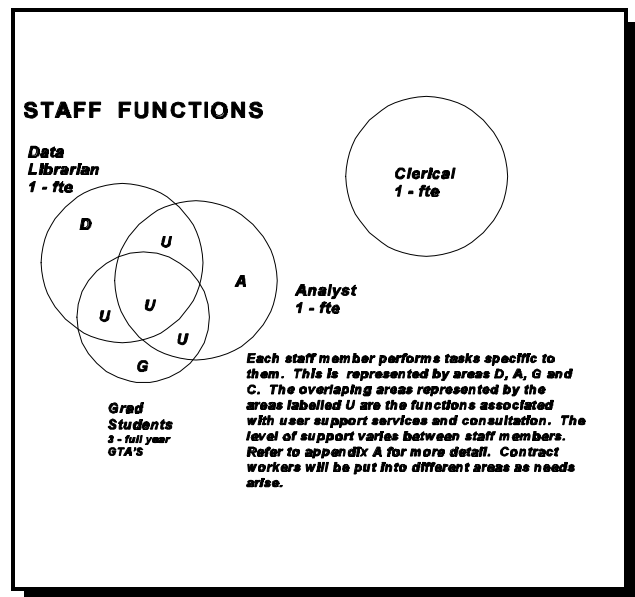


Figure 6

Aside from the contract workers these positions should be filled as secondments. **The Data Librarian would be a 2 year secondment from the library. The Data/Technical analysts would be a 2 year secondment from CCS. The clerical position would be secondments from both CCS and the library that would make up one fte. The graduate students would be assigned on 1 year contracts by the Colleges who use the facility the most.** As for the contract workers it is recommended that these be drawn from the high school co-op pool, the university co-op pool and work studies students, although consideration may be given to part time secondments.

The data librarian would be responsible for the operation and management of the facility and **would report to the head of the documents centre.**

3.2.1 Initial startup costs (Guelph alone)

The following section outlines the costs involved in establishing and running a DLC. They are just a rough estimate to get a feel for what is required. There has not been a great deal of effort put into actually costing this out in detail.

Startup Costs

Furnishings (desks, chairs, shelves, white boards bulletin boards...)		\$15,000
Equipment		
-Unix server (tape, 128 MB RAM, 30 GB HD CD-ROM stack)	\$75,000	
- 3 pentium systems for staff	\$10,000	
- 1 pentium system for consulting and users	\$2,500	
-1 laser printer	<u>\$3,000</u>	\$90,500
Software		\$10,000
Salaries (3 - 6 month contracts)		<u>\$45,000</u>
TOTAL		<u>\$160,500</u>

3.2.2 Yearly Costs (Guelph alone)

It should be noted that in the first year of operation some of the costs mentioned would be covered by the initial setup costs. In other words there would be no costs for furnishings, equipment and software during the first year as they are covered in the initial setup.

The total figure should also be put in perspective. These are not necessarily new expenditures as there is a redistribution of expenditures. As an example, researchers save resources by not requiring individuals to search for and process electronic data. As an example, assume that 10 departments, from the 4 colleges mentioned, each save the equivalent of one GRA per year by using the DLC¹⁵. If this GRA made on average \$12,000 per year the savings are \$120,000.

YEARLY COSTS

Salaries		
- staff - 3 - fte	N.A	
- 3 full year GTA's	N.A	
- contract workers	<u>\$20,000</u>	\$20,000
Furnishings (10% replacement)		\$1,500
Equipment (25% for replacement and upgrade)		\$22,500
Software		\$10,000
Data acquisitions		<u>\$25,000</u>
TOTAL		<u>\$79,000</u>

¹⁶ It should be noted that the savings does not necessarily accrue from GRA's.

The outlay for data acquisitions is also an expenditure that is already being made. It currently comes out of several budgets and the DLC is only bringing it all together. Some negotiations may need to be made regarding this redirection of funds.

Figure 7 shows how the new DLC fits into the University's computing environment.

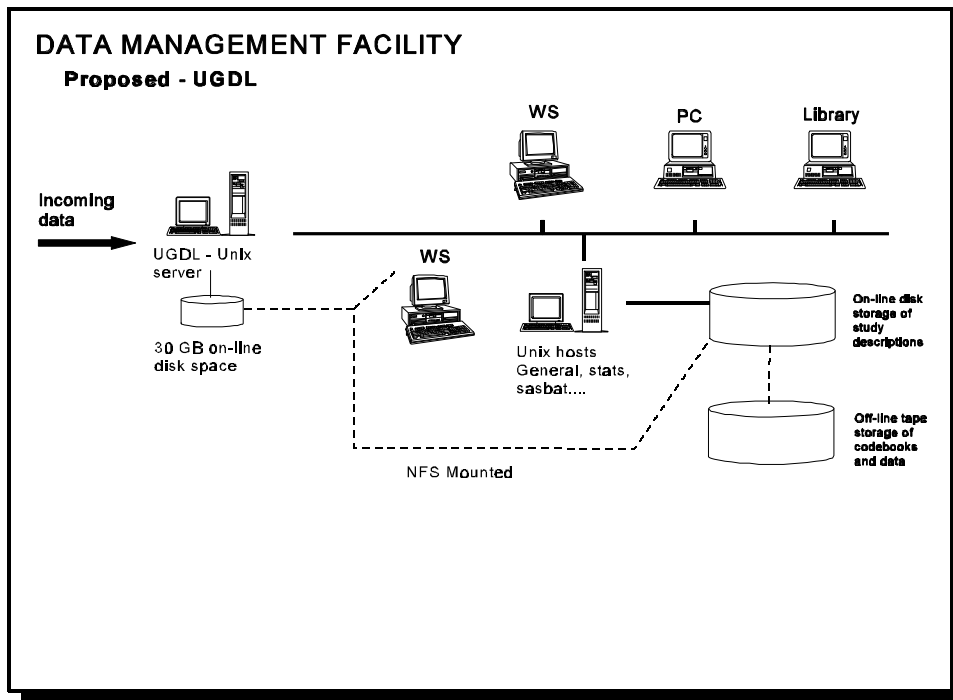


Figure 7

The significant data sets, such as the CENSUS would be stored on-line with the UGDL server¹⁶. Other data sets could be linked into the Universities HSM system so that they are backed off to tape when not used. Possibly a CD-ROM stack could also be installed. A limited amount of statistical analysis would be done on the DLC server. General processing associated with the web front-end and data preparation could be handled. Analysis by the end-user would generally be done from the central system run by CCS, after the data is moved over or linked by an NFS mount.

¹⁷ A careful evaluation of disk space needs would have to be undertaken.

4. Other Institutions

The following section summarizes what is happening at various other Universities in Canada. It is interesting to note the similarities in their experiences and their approaches to solving these problems. Most of the information is derived from reports from various sites to the CAPDU list. Some of the reports were not very detailed so the summaries are loose interpretations of what is being done (Bo takes responsibility if anything is said about an institution which isn't quite true...) Originals are available on request. Notice the weighting towards the western provinces, where a strong data culture is evolving. Summaries appear in no specific order.

4.1 Calgary

There seems to be a great deal of resources being put into their services. The emphasis is on developing search and retrieval using web access. Funding has come from teaching support and the private sector. They are developing a system called LANDRU which is also mentioned in reports from University of Alberta and University of Victoria. It appears that some joint ventures will be undertaken. Their web site is <http://www.ucalgary.ca/~libdata>.

4.2 UBC

They have experienced similar problems to us accessing large data sets after moving to Unix (tapes vs. disk access). They have 60 to 70 data sets on-line and share access to CANSIM. They NFS mount data for departmental machines, use SAS datasource and are also developing WWW/SPSS extraction engines. They also strongly support the notion of 'one stop shopping'. Their site is <http://www.datalib.ubc.ca/>

4.3 University of Toronto

Toronto has had a strong data library for some time. There is a traditional resource run through the library. There is also another group from EPAS developing front-end web-based systems, like the CANSIM interface that we use at Guelph. The EPAS site is: <http://datacentre.epas.utoronto.ca>.

4.4 Carleton University (and Ottawa)

Data services at Carleton have existed for 30 years. Currently they have 2 employees (one is the driving force behind the DLI). They also use Social Science students through work studies programs as well as always having a PhD student assigned to them. They are trying to work jointly with the University of Ottawa who recently appointed a PhD student to run their centre. They note that the introduction of the DLI has increased their workload and currently they have 30 GB of data on-line and expect more. One benefit of the DLI that they observe is the increase in the number of graduate students now requesting data and undertaking empirical research. Their web site will soon be at: <http://www.carleton.ca/~ssdata>.

4.5 Western

Western has a well established and fully funded Social Science Data library. It is a very nice model, but it uses a lot of resources and has a lot of staff. They also manage the computing labs and systems for the Social Sciences. It is beyond what we propose for Guelph.

4.6 Waterloo

Waterloo is looking at the same issues that we are. They are a little further ahead in general library WEB interfaces. They have also developed their own WEB front-end for the EPAS - CANSIM system. Some of their hardware/software problems are similar to our experiences. They have a web site at: <http://www.lib.uwaterloo.ca/>

4.7 University of Alberta

They are deciding on how to proceed. They have had some very experienced staff for some time and are currently adapting to changes in their internal computing environment and the introduction of DLI. They are interested in the search and retrieval systems using LANDRU, SAS/C or some other web based system. Currently they spend a lot of time subsetting data sets that users take away on diskette. They are having some problems delivering network access to their data. The move away from their old central computing systems has been problematic. They have a site at: <http://datalib.library.ualberta.ca>

4.8 University of Victoria

They are in the early stages, recently hiring a new person to run their service. They appear to be interested in sharing with UBC and Alberta.

4.9 Queen's University

Data services at Queen's have historically been run out of their Documents Centre and were established in 1982. They have recently established a Social Science Data Centre (see <http://satuffer.queensu.ca/docsunit/ab-ssdc.html>) . They have on-line access to Cansim and searchable indexes for all their holdings.

APPENDIX A: Job Descriptions

Data Librarian

1 - fte secondment from library (2 year term)

Skills:

- management and administration skills
- technical programming - computing skills
- extensive knowledge of statistics and statistical software
- extensive user support experience
- librarian skills
- research skills (preferably undertaking scholarly work)

Functions:

- general management of facility and its direction
- management of staff
- management of budget
- establish collection development policy
- prepare annual reports
- prepare funding proposals
- liaison between Library, CCS and user community
- report to Statistical Advisory Committee
- report to College IT committees
- liaison with outside data centres and suppliers
- promote national standards
- negotiate data purchase agreements (such as DLI)
- editor of newsletter
- provide user support and consultation
- staff training (incl. documentation)

Data/Technical Analyst

1 - fte secondment from CCS (2 year term)

Skills:

- technical programming - computing skills
- knowledge of statistics and statistical software
- understanding of research issues and experience with them
- consulting experience
- WWW/HTML experience

Functions:

- write interfaces and system files to access data
- perform subsetting and design user documentation
- handle the organization of any archiving functions, such as cleaning and inventorying
- provide user support and consultation
- staff training
- front line consulting
- design facilities to maintain statistics on usage
- design and maintain back-up (disaster) strategies

Clerical

1 - fte - secondment from CCS and library

Skills:

- general clerical skills
- cataloguing skills
- basic reference librarian skills
- basic computer skills to field general inquiries
- WWW/HTML skills

Functions:

- cataloguing acquisitions
- general secretarial functions (typing documentation ...)
- performing simple cleaning and organization of data
- handle billing and payment functions
- conduct evaluation surveys

GTA

3 - full year assignments from Colleges

Skills:

- statistical programming (SAS. SPSS....)
- some HTML/WWW experience
- knowledge of electronic data resources
- preferably PhD.

Functions:

- some programming and WWW
- front-line consulting with user community
- assignment to special projects (ie census)

APPENDIX B: Conditions for DLI membership

Responsibilities of Partners in the Project

I. Depositories

In order to participate in the project, depositories must undertake to:

- sign and respect the terms of the electronic depository agreement
- ensure that data obtained through this initiative neither be redistributed for use other than as stipulated nor beyond the depository membership and not be repackaged in any way for commercial use
- pay the DSP the annual subscription fee
- publicize the availability of datasets and databases
- designate an official DSP contact person in each institution who will be responsible for selecting appropriate datasets and databases
- select only data for which there has been a demand, or for which a request may be anticipated within a short time-span. This will minimize strain on the network
- identify and recommend to the DSP products for possible inclusion
- provide users with access to the data in as user-friendly a format as possible
- educate users regarding the content, file structure and potential uses of the data
- respond to reference enquiries
- publicize the guidelines under which data may be accessed
- track the use of data in terms of persons requesting the files, and where

publicly
funded
research
libraries

possible, track users, uses, publications and other end products

- collect bibliographic information regarding publications to be given to the National Library
- prepare a written report of project activities to the DSP annually

APPENDIX C - List of files in DLI

Updated Feb. 2 1996

General files

CANSIM

E-STAT

Census Area Profile Series (1986, '91, '96)

Basic Census Summary Tables "

Census of Agriculture Summary Files "

Labour Force Historical Review

Business Integrated Database (BID)

Health Indicators Database

Inventory of Statistics Canada's Questionnaires 1994 on CD-ROM

Small Business Profiles

Employment Change by Industry Aggregate, Business Size and Life Status

World Trade Database

TIERS (Trade) Database

The Social Policy Simulation Database and Model

Intercorporate Ownership on CD-ROM

Labour Market Activity CD-ROM (1986-1990)

Whole Farm Database Extraction System of Agricultural Statistics (ESAS)

Fleet Report, Inventory of Commercial Aircraft in Canada

Geographic Files

GEOREF (Geographic Information Reference System)

Digital Boundary Files (DBF)

Digital Cartographic Files (DCF)

Postal Code Conversion Files (PCCF)

Block-face Data File (BFDF)

Geographic Attribute File (GAF)

Street Network Files (SNF)

Skeletal Street Network File (SSNF)
Place Name Master File (PNMF)

Public Use Micro data Files (PUMFs)

Since a number of these files are done on a contractual basis, titles beyond 1995 are not yet known. They will, however, be included in the package. Some of the newer files have not yet been released, but are expected later in 1996.

It is our intention to make files collected from 1990 to the present available in year one. Older files will be gradually included where documentation is still available.

Census

Individual Files (1986,'91,'96)
Family Files
Household and Housing Files

Annual

General Social Survey
Survey of Consumer Finances
Survey of Labour and Income Dynamics (longitudinal)
Household Income, Facilities and Equipment (HIFE)
International Travel Survey (since 1990, quarterly)
- Canadian Resident Trips to the United States
- Canadian Resident Trips overseas
- United States Resident Trips to Canada
- Overseas Resident Trips to Canada (Excluding the United States)

1996

International Adult Literacy Survey - Canadian Data
National Longitudinal Survey of Children
Youth Smoking Survey
National Private Vehicle Survey
Canadian Travel Survey

1995

Survey on Smoking in Canada
National Population Health Survey (longitudinal)
Health and Activities Limitations Survey (HALS)
Aboriginal People's Survey

1994

Canada's Alcohol and Other Drugs Survey
Adult Education and Training Survey
Canadian Travel Survey

1993

Absence From Work Survey
Survey of Household Energy Use, 1993
Survey of Job Opportunities
Violence Against Women Survey

1992

Adult Education and Training Survey
Value of Wildlife to Canadians
Absence From Work Survey
Survey of Job Opportunities
National Graduates Survey 1990
Survey of the People Not in the Labour Force
Family Expenditure Survey (FAMEX)
Food Expenditure Survey (FOODEX)
Canadian Travel Survey

1991

Survey of Work Arrangements
Survey on Ageing and Independence
Survey of Job Opportunities
Current Population Profile
Absence From Work Survey 1991
Labour Market Activity Survey

1990

Adult Education and Training Survey
Health Promotion Survey
Absence From Work Survey
Labour Market Activity Survey
Canadian Travel Survey
Family Expenditure Survey (FAMEX)
Food Expenditure Survey (FOODEX)

1989

National Apprenticeship Survey
Survey of Literacy Skills Used in Daily Activities
Alberta Apprentice/Journeymen and Work Reduction Survey
Barriers to Advancement in the Public Service
Households Facilities and Equipment Survey
Survey Pacific Costs and Earnings
Survey of Job Opportunities
National Alcohol and Drug Survey
Housing Repair and Renovation Survey
Survey of Pacific Fishermen
Survey of Atlantic Fishermen
Absence From Work Survey
Labour Market Activity Survey

1988

Survey on Telephone Services
Health and Employment Survey
Survey of Apprentices and Journeymen in Alberta
National Child Care Survey
1988 Veterans Survey
Canadian Travel Survey
Survey of 1986 Graduates
Survey of Job Opportunities
Survey on Drinking and Driving
Shelter Cost Survey
Absence From Work Survey

Survey of the Importance of Wildlife to Canadians
Labour Market Activity Survey

1987

Fuel Consumption Survey
Canada Pension Plan Survivors Beneficiaries Survey
Canada Pension Plan Disability Beneficiaries Survey
Health and Activities Survey
Current Population Profile
Survey of Volunteer Activity
Survey of Apprentices and Journeymen in Alberta
Survey of Full-time Employees Concerning Part-time
Employment
Part-time Employment Evaluation
Ontario Child Health Follow-up Survey
Follow-up of 1982 Graduates
Survey of Job Opportunities
Absence From Work Survey

1986

Fuel Consumption Survey
Survey of Smoking Habits
Survey of Self Employed
Yukon Travel Survey
Health Promotion Survey (Montreal)
Survey of Apprentices and Journeymen in Alberta
Ontario Tourism and Motivation Study
Survey of Pacific Cost and Earnings
Census Awareness Survey
Survey of Pacific Fishermen
Canadian Travel Survey
Survey of Job Opportunities
Absence From Work Survey
Adult Training Survey
Survey of Displaced Workers

1985

Fuel Consumption Survey
Survey of Annual Work Patterns (1985)

Survey of Apprentices and Journeymen in Alberta

Survey of Disentitled Family Allowance Recipients

Airport Official Languages

Health Promotion Survey

Survey of Work Reduction

Survey of Job Opportunities

Survey of Educational Attainment

Survey of Maternity Leave

Absence From Work Survey

Survey of Atlantic Fishermen

Survey of Annual Work Patterns (1984)

1984

Fuel Consumption Survey

Survey of Union Membership

Travel to Work Survey

Survey of Job Opportunities

Vehicle Maintenance Survey

Canadian Transport Commission Survey

National Graduates Survey

The Canadian Health and Disability Survey

Transportation Survey For Special Care Facilities

Canadian Travel Survey

Survey of Job Opportunities

Post-secondary Student Survey

Vehicle Maintenance Survey

Family History Survey

Absence From Work Survey

Adult Education Survey

1983

Fuel Consumption Survey

Survey of Smoking Habits

Survey of Annual Work Patterns

Travel to Work Survey

Canadian Health and Disability Survey

Tourism Attitude and Motivation

Special needs in Public Transportation Survey

Survey of Job Opportunities
Absence From Work Survey
The Ontario Child Health Survey

1982 Fuel Consumption Survey
Survey of Annual Work Patterns
Travel to Work Survey
Education Survey
Current Population Profile
Canadian Travel Survey
Survey of Job Opportunities
Survey of Skills and Training
Survey on the Value of Wildlife to Canadians
Absence From Work Survey
Survey of 1981 Work History
Crime Survey (7 city)

1981 Passenger Car Fuel Consumption Survey
Smoking Habits Survey
Travel to Work Survey
Survey of Job Opportunities
Survey of Child Care
Absence From Work Survey
Annual Work Patterns Survey

1980 Passenger Car Fuel Consumption Survey
Current Population Profile
Travel to Work Survey
Survey of Student Finances
Canadian Travel Survey
Survey of Job Opportunities
Survey of Volunteer Workers
Absence From Work Survey
Annual Work Patterns Survey

- 1979** Passenger Car Fuel Consumption Survey
Smoking Habits Survey
Travel to Work Survey
Labour Market Comparison Study
Survey of Job Opportunities
Absence From Work Survey
Annual Work Patterns Survey
Greater Vancouver Crime Survey
- 1978** Travel to Work Survey
Canadian Travel Survey
Attitudes About Surveys
Survey of 1976 Graduates of Post-Secondary Programs
Hamilton-Wentworth Victimization Survey
Survey of Job Opportunities
Absence From Work Survey
Survey of Leisure Time Activities and Reading Habits
Annual Work Patterns Survey
- 1977** Smoking Habits Survey
Travel to Work Survey
Methods Test
Weeks Worked Since September 1976
Student Identification
Survey of Job Opportunities
Absence From Work Survey
Annual Work Patterns Survey
- 1976** Travel to Work Survey
Physical Recreation and Sport
Guns/Ownership Survey
Census Micro Match
Absence From Work/Annual Work Patterns
Income Screen Survey

Previous list - Compiled by:

Lynda Richardson Liaison Officer
Data Liberation Initiative

Voice: (613) 951-5904 Statistics Canada
Fax: (613) 951-1134 9-M R.H. Coats Bldg.
Email: richlyn@statcan.ca Ottawa, ON. K1A 0T6

APPENDIX D: Sample Job Descriptions from elsewhere

This is just an example of 3 different job postings recently placed in CAPDU. It is in no way complete but gives an idea of what sort of positions are being filled.

Data Resource Librarian

THE UNIVERSITY OF MANITOBA LIBRARIES
Invites Applications for the Position of
Data Resources Librarian

The Elizabeth Dafoe Library, one of the 11 library units within the University of Manitoba Libraries, supports the teaching and research requirements of the faculty and students of Arts, Human Ecology, Social Work and Nursing. Service is also provided to other members of the university community including off-campus students, and to the general public. The unit includes Circulation Services, Reference Services, Government Publications and the Icelandic and Slavic Collections. The library contains over 830,000 books and bound periodicals; 430,000 government publications; 606,000 microforms; 100,000 maps; and subscribes to over 4,000 periodicals. Approximately 320,000 items are circulated each year. The library performs the functions of collection development, reference services including on-line searches, plus orientation and library instruction. It also provides correspondence/off campus services. Staff consists of 10 librarians, 1 supervisor, 17 library assistants and approximately 8 F.T.E. part-time staff members. The incumbent reports to the Head of the Elizabeth Dafoe Library and the Head of the Reference Services Section.

Responsibilities of the incumbent are as follows. Provides direct service to users requiring machine-readable information, including the searching of appropriate sources for data, physical access to computer media, codebooks and technical specifications needed to access the data files. Acquires data files according to established subject profiles and on the recommendations of users and manages the data collection. Monitors the maintenance and acquisition of appropriate hardware and software to support the data resources service. Monitors developments in the use and application of data

sets in social science research. Other disciplines such as health-related fields may be included. Provides liaison with Computer Services, the Libraries' Electronic Services area, and appropriate bibliographers regarding the development, technical support and collection policy issues of the data collection. Provides liaison with consortia such as COPPUL and CARL regarding the acquisition and maintenance of data files. Promotes the data resources services within the academic community. Participates in the development and maintenance of the roles and objectives of the unit. Provides reference assistance to on- and off-campus patrons. Plans, prepares and presents orientation programs and literature. Is responsible for collection management in a specific area of the social sciences and of the reference collection.

Qualifications: A degree from an ALA-accredited library program. An undergraduate degree in the social sciences with a background in computing is preferred. Academic background in the social sciences with advanced training in methodology and research skills is highly desirable. Excellent communication skills and interpersonal abilities required to work well with people both individually and in groups. Knowledge of and experience with computing technology, including database software, statistical software (especially SPSS or SAS) and the management of files using a mainframe or microcomputer operating system (e.g., UNIX, MVS, MS Windows, DOS, etc.). Reference desk experience and/or classroom teaching experience are highly desirable. The successful candidate is expected to participate in professional development and relevant professional activities.

Effective date: May 1, 1996

Ranks and Salary Ranges: General Librarian: \$31,746 - \$41,410

This position has a two-year probationary period. Librarians enjoy academic status with possibility of promotion.

The filling of this position is subject to final budgetary approval.

The University of Manitoba encourages applications from qualified women and

men, including members of visible minorities, aboriginal people and persons with disabilities. The University provides a smoke-free environment, save for specially designated areas. In accordance with Canadian immigration requirements, this advertisement is directed to Canadian citizens and permanent residents.

Senior Research Associate

Senior Research Associate Social Science
Inter-university Consortium for Political and Social Research

Job Family: Professional/Administrative
Posting No: T-96-1657-LM
Job Title: SR RES ASSOC SOCIAL SCI
Grade: 12
Min/Max: \$ 33,300/ 86,300
Department: Center For Political Studies
Open/Close Date: 06/17/1996 // 06/21/1996
Job Class: 12880
Hours: 40.00

DUTIES:

Take a leadership role in the Inter-university Consortium for Political and Social Research (ICPSR) to create a new sub-archive focusing on data about substance abuse and mental health services; evaluate data collections for inclusion in the archive; direct and manage acquisition and archival processing activities; maintain working liaison with both data producers and data users; organize and provide support for training workshops and symposia in the uses of these data; write and/or edit all ICPSR publications related to data on substance abuse and mental health services, including an annual catalog of data holdings, advertisements and documents informing the research community about the characteristics of those data; assist in writing grant and contract proposals/renewals; write progress reports on existing grants and contracts; promote the use of the data for both research and teaching; prepare exhibits and attend relevant professional meetings to represent ICPSR and promote use of this sub-archive; participate in the representation of ICPSR and the sub-archive with the Substance Abuse and Mental Health Services Administration (Department of Health and Human Services), other potential funders of the sub-archive,

and the prime contractor, NORC; organize and attend meetings composed of scholars and agency officials; hire, train, supervise and evaluate a staff of 4-6 professional, technical and clerical employees engaged in archival data processing activities; funding for this position has been authorized for two years with renewal of funding possible; further information about ICPSR is available at <http://www.icpsr.umich.edu>.

DESIRED QUALIFICATIONS:

Ph.D. degree in a social science discipline relevant to the study of substance abuse and mental health services; demonstrated experience in proposal writing and fund raising; progressively responsible experience in a social science data archive or in preparing computer readable data for secondary analysis; experience in managing social science projects and supervising professional, technical and clerical staff; experience in processing large collections of social science data.

MINIMUM QUALIFICATIONS:

Masters degree or higher in a social science discipline relevant to the study of substance abuse and mental health services; two or more year's experience in computerized data processing using work stations, microcomputers and local area networks; demonstrated knowledge of social science data and research methodology; ability and interest in managing projects in the social sciences and in supervising professional, technical and clerical employees in data processing or data analysis; knowledge of and experience with social science software packages (such as SPSS or SAS) for processing of quantitative data; the ability to communicate clearly in both spoken and written professional English; excellent interpersonal skills.

Contact:

Present Applications For This Position To The Following Office:

Ann Arbor Campus Employment Services Office

Room G250 Wolverine Tower
3003 South State Street
Ann Arbor, Michigan 48109

(313) 764-6580

8 AM to 5 PM, Monday - Friday

The University of Michigan is an equal opportunity, affirmative action employer and encourages women and minorities to apply.

Note: the deadline for applications will be extended for two additional weeks beyond 6/21/96.

Coordinator of Electronic Data Services

POSITION: Coordinator of Electronic Data Services

AVAILABLE: Fall 1996

ENVIRONMENT: Emory University's General Libraries seek a Coordinator of Electronic Data Services who is eager to participate in the accomplishment of the following goal from the library's strategic plan: "Provide quality services that facilitate user access to a broad spectrum of locally maintained and networked information sources accessed remotely. User education in the skills required to access, manage, and assess information sources is important as is the goal of user self-sufficiency." The Coordinator of Electronic Data Services will work in a team setting with operational units of the General Libraries to develop and deploy numeric data information services and resources. A new facility, soon to be under construction, will enhance the opportunity for innovative approaches to the use of information technology and the development of services for the scholarly community in the electronic environment.

RESPONSIBILITIES: Within the context of a changing environment, this position will provide reference and research services as well as outreach and training for users of numeric data in machine-readable form. Working with other library staff, the position will develop numeric data as a major reference resource. The position will also be responsible for identifying and acquiring datafiles from ICPSR and other vendors and for selecting storage and access options including both hardware and software necessary to support distributed computing. Development of policies and procedures in support of the data collection and its services will also be a responsibility. The position will serve as official Emory representative to ICPSR and other relevant professional organizations. The position will report jointly to the Coordinator for the Social Science Collections and the Head of Reference Information, Consulting and Instruction.

QUALIFICATIONS: ALA accredited MLS or masters degree in the social

sciences or related discipline; knowledge of social science quantitative methodology; ability to work with all levels of researchers; good organizational skills; ability to interact successfully and communicate in a clear, knowledgeable and personable manner in both a team environment and on an individual basis; demonstrated proficiency with social science statistical packages (e.g. SPSS, SAS). Desired qualifications include: one or more years experience in a social science data library or related organization; knowledge of the UNIX environment and related programming languages (e.g. PERL); familiarity with networked environments and AIX.

EMORY LIBRARIES: The libraries of Emory University hold 2.3 million volumes and employ a staff of 263 FTE. In addition to the General Libraries, there are libraries for health sciences, law, theology, and Oxford College. The General Libraries, in conjunction with the Information and Technology Division, are currently embarking on a long term effort to develop a program of services focused on access to networked information resources. The Emory libraries are members of the Association of Research Libraries, OCLC, Research Libraries Group, Center for Research Libraries, the Coalition for Networked Information, the National Digital Library Federation, and the University Center in Georgia.

BEGINNING SALARY AND BENEFITS: Salary and rank dependent upon qualifications and experience. Minimum salary \$31,000. Comprehensive Benefits Package, including TIAA/CREF.

APPLICATION PROCEDURES: Submit letter of interest, resume and names, addresses and telephone numbers of three references to: Dianne M. Smith, Library Human Resources Officer, Robert W. Woodruff Library, Emory University, Atlanta, GA 30322-2870. Review of applications begins July 15th.

Emory University is an Equal Opportunity/Affirmative Action Employer and encourages women and minority candidates.

