Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

National Archives of Canada

Archives nationales du Canada

Canada

We build understanding

Final Report

National Data Archive Consultation

Building Infrastructure for Access to and Preservation of Research Data

Submitted by the NDAC Working Group to the Social Sciences and Humanities Research Council of Canada and the National Archivist of Canada

June 2002

Executive Summary

In October 2000, the Social Sciences and Humanities Research Council and the National Archivist of Canada established a Working Group of research and archival experts and asked them to assess the need for a national research data access, preservation and management system. After compiling extensive evidence for the need of such a service to support the knowledge creation work of Canada's social sciences and humanities research community, the Working Group now offers recommendations for the creation of a new national research data archival service. This service would have three core functions:

- Preserving research data that is complied by researchers, and preserving data compiled by government agencies, polling firms and other organizations that can be used by researchers to generate new knowledge;
- Managing the data held, including ensuring quality; selecting data for retention; developing and applying standards for metadata, authentication and security; and migrating data across technologies;
- Providing access to research data, including Web-based delivery systems, cataloguing services, user and depositor agreements to protect confidentiality and intellectual property rights, and connections to other data depositories around the world.

In addition, the Working Group recommends that a new National Research Data Archive Network undertake a number of other functions, including providing advanced training in data handling techniques, represent Canadian interests in the development of international data standards, promote data sharing as a best practice in research, undertake research in information and archival sciences, and act as a central hub and co-ordinating body for a network of data services in Canadian research institutions.

Digital information compiled for research purposes is playing an increasingly important role in today's knowledge economy. In many ways, data is the fuel driving innovation and our capacity to address complex social and economic problems. Although billions of dollars are spent each year collecting data, Canada lacks the necessary infrastructure to ensure these data are preserved and made publicly available. This limits the returns that can be made on our public investments in research and undermines good public stewardship.

Many of the building blocks necessary for the creation of a National Research Data Archive are already in place. University data services, high-speed transmission networks, legal and ethical guidelines and frameworks, potential partner institutions, various data depository and access portal initiatives, and an active data-producing research community already exist. The missing element is a preservation, co-ordination and management service.

Almost all developed countries have recognized the need for a national research data service, and some have more than a generation of experience in their operation. Canada is in a position to learn from this experience while developing a research data service that

fits our unique institutional and cultural context. We now have the technological capacity and expertise to create a "trusted system" that provides Canadians with an accessible and comprehensive service empowering researchers to locate, request, retrieve and use data resources in a simple, seamless and cost effective way, while at the same time protecting the privacy, confidentiality and intellectual property rights of those involved. The start-up infrastructure costs for this service could be funded through the Canada Foundation for Innovation. The annual operating costs for a comprehensive facility and network are benchmarked in the area of $3 million.

The Working Group offers three options for the creation of a National Research Data Archive Network:

1) Through federal legislation, create a National Research Data Archive Network as a modified version of a Separate Statutory Agency. This is the ideal approach to building a full-service, trusted agency, composed of a central data preservation and management facility and a series of access and service nodes located in research institutions. It takes full advantage of existing research infrastructure, has long-term stability, a direct connection to research data users and producers, and the capacity to represent Canada's interests in the development of international data standards.
2) Create a National Research Data Archive Network under the auspices of the Social Sciences and Humanities Research Council. This approach captures the characteristics of the first model, but does not require legislation. It benefits from a direct, immediate connection with researchers and established accountability and funding structures.
3) Create a Special Operating Agency within the National Archives of Canada. As a stand-alone division within the National Archives, this approach takes advantage of existing archival infrastructure and expertise. This has not been the preferred approach in other countries, because the core mission of a national archive and a national research data service are fundamentally different. Nevertheless, as a Special Operating Agency, the service could potentially have both stability and the capacity to develop a trusted research data preservation, management and access system.

As a next step, the Working Group recommends that SSHRC and the National Archivist create a Steering Committee to select the appropriate approach to setting up a National Research Data Archive Network, or research data archiving service, further define the characteristics and funding requirements for such a service, and promote its establishment.

Report Contents

Sections

Appendices

National Data Archive Consultation

Final Report

Building Infrastructure for Access to and Preservation of Research Data in Canada

1. Introduction

The network of institutions and agencies that make up the infrastructure supporting Canada's knowledge economy currently has a serious gap. Canada lacks a national agency to preserve, catalogue and provide systematic, efficient and convenient access to research data. This digital information enables researchers to substantiate existing knowledge, replicate and verify research findings and explore and create new knowledge. Effective access to, and use of, research data can play a central role in Canada's innovative capacity. The necessary infrastructure, however, must be in place.

In October 2000, the Social Sciences and Humanities Research Council and the National Archivist of Canada mandated a Working Group of research and archival experts to consult with the research and archival communities and assess the need for a national data archiving service or function. After completing this assessment, and compiling extensive evidence for the need of such a service, the Working Group investigated research data archives in other countries and explored possible approaches to building such a core research facility in Canada.

The Working Group now recommends the establishment of a Canadian agency to close this gap in the infrastructure of the Canadian knowledge economy – the creation and long-term, stable support of a National Research Data Archive.

Today, almost all research takes place in a digital environment. Complex multi-layered statistical databases, digital maps and images, and encoded texts are now commonplace tools for researchers. Although these resources have dramatically expanded the scope of research, and increased its efficiency, the institutional structures required to preserve, manage and make accessible that digital information have not kept pace. This situation undermines the innovative capacity of Canadian researchers and places tens of millions of dollars worth of highly valuable research data at risk.

To build a knowledge society, to foster innovation, and to deal with pressing, complex social, political and economic problems depends in large part on the discovery of knowledge through research. In order to be responsive and efficient, while incorporating multiple perspectives, researchers require access to, and sharing of, a wide variety of research data. For this to happen, infrastructure is necessary. Today, many elements are in place – university research libraries and data services, research support councils, high-

speed data transmission networks – but one vital element, a facility for storing, distributing and preserving research data is missing.

Good public stewardship demands that public investment in research data realise maximum returns. In order to maximise returns, research data should be used as many times, and in as many different situations, as possible. This can only happen if we put in place effective research data infrastructure. The cost of inaction not only puts our investments in science at risk, it undermines one of the core responsibilities of government.

2. What is Research Data Archiving?

Unlike many forms of traditional archiving, research data archiving is not about keeping records for legal, historical or cultural purposes; it is about meeting the needs of researchers operating in today's digital environment. The core mission of a research data archive is not to preserve the recorded memory of a group, organization or nation, but to provide a vital service to the research community.

Although there are many emerging institutional needs related to digital materials, the Working Group examined only the data access, management and preservation needs of the research community, mainly the social sciences and humanities. From this perspective, the Working Group defined the process of research data archiving as preserving, managing and making publicly accessible digital information structured through research methods with the aim of producing new knowledge. This process provides stewardship for those outputs of research that exist between initial information and published results. Acquisitions would include digital information produced by researchers and of interest to researchers, subject to the limitations of financial resources and retention protocols developed by research data archivists and the research community itself.

National research data agencies and archives in other countries provide a broad range of access, preservation, and management services to their respective research communities, including on-site and off-site storage, access to catalogues and data sets through the Internet, retention protocols, metadata creation, migration of data across software and hardware systems, training and developing international standards. In offering these services, they play an active and crucial role as information and knowledge brokers.

I see a National Data Archive as an institution that is trusted and recognized as having the Canadian mandate to preserve research data, to work with other governmental and non-governmental agencies in ensuring that their data management practices incorporate preservation standards, to work closely with other Canadian institutions charged with preserving Canada's heritage to guard against gaps in responsibilities, to co-ordinate and represent Canada in international research data exchanges and

in the development of related standards, to provide access to these data, to educate Canadians about the use of research data, to contribute to new research by helping create new data from archived data, to help safeguard privacy in Canadian society in light of massive amounts of stored digital information on individuals, and to conduct research and development into all aspects of data preservation.

Charles Humphrey, Data Librarian, University of Alberta, NDAC Working Group Member

---

3. The Need for a National Research Data Archive

As one of the Working Group members put it, an unprecedented firestorm is now incinerating Canada's digital research wealth. Although this may seem an overstatement, it is a deep-seated concern shared by many archivists, librarians and researchers around the world.[1]

Research information in digital form is extremely fragile yet capable of being collected in huge quantities. Today, we are only beginning to understand how to preserve and manage this information effectively. Although there are no easy short-cuts for dealing with such issues as media obsolescence, digital "rust", copyright, confidentiality, the creation of national and international standards, and the limitations of the current research culture, avoiding or ignoring them will prove costly in the long run.

In the initial phase of the National Data Archive Consultation, the Working Group sought input from a broad range of stakeholders who use, manage and produce research data related to the social sciences and humanities. The objective was to assess the need for a national research data archival service or function (see Appendix E). This assessment brought to light a number of structural gaps:

- Currently, there is no national institution preserving, managing and making research data publicly accessible on the scale required to support the Canadian research community. The National Archives of Canada does not have the resources to do so;
- University research data services have neither the resources nor the responsibility to act as nationally-oriented research data archives. Although they are struggling to fill the gap left by the absence of a national data archival service, university data services are, in general, only mandated to provide local patrons with access to readily-available data;
- The SSHRC Data Archiving Policy, which directs the researchers it supports to deposit their data with university data services, has not achieved its objectives. In fact, over an eleven-year period only 10 data sets have been deposited with the university data depositories listed in the SSHRC Guide. Although some researchers

[1] Numerous organizations are currently wrestling with research data archiving policies and structures, including the Library of Congress, the Economic and Social Research Council of the United Kingdom, the U.S. National Institutes of Health, the U.S. National Archives and Records Administration and the National Research Council, the International Council for Scientific and Technical Information, and the International Council of Scientific Unions.

are reluctant to share their data, it would be unethical for SSHRC to enforce this policy in the absence of a facility that would allow researchers to abide by the regulations;

- Canada has no co-ordinated voice in setting international research data standards, in metadata schemes such as Data Documentation Initiative, in tools for data access such as the Networked Social Science Tools and Resources (NESSTAR) project, and in collaborative international infrastructure projects such as the European Union Frameworks. As well, Canada lacks national representation on the International Federation of Data Organizations or participation in the initiatives of the Council of European Social Science Data Archives;
- One of the paramount problems researchers face today is difficulty in locating data relevant to their research. There is no 'union list' or catalogue of data sets held by data producers, distributors or other researchers. As a result, researchers may needlessly replicate costly studies, rely on anecdotal rather than empirical evidence, or use substitute data from other countries. Potentially, a national data service could place information about data sources, as well as the data itself, directly on the researchers' desktops, thereby saving time, money.

As well, a National Research Data Archive could serve fundamental needs by:

- Ensuring the authenticity of research data, a growing concern among both research data producers and data users. Authentication procedures embedded in the process of creation, transmission, receipt, use, maintenance and preservation of data files are the most effective way to ensure the authenticity of data over time. Currently, we have neither national standards of this kind nor any agency to oversee their application;
- Reformulating and articulating, at the national level, security standards that protect data adequacy and consistency. These standards should address: (1) methods for identifying data assets and risk-management procedures for assessing vulnerabilities; (2) identification of legal, statutory, regulatory and contractual requirements, including ethics guidelines and intellectual property rights; and (3) a set of principles, methods and procedures that organisations must follow to ensure the reliable creation, secure maintenance, confidential use and authentic preservation of their data.

If Canada were to build a National Research Data Archive, would it be used? Ample experience in other countries shows that data usage is growing in number of users and frequency (see Appendix C).

Among the top ten most popular data sets requested by users of the UK Data Archive in the 2000/01 fiscal year, four of these titles were from government departments, two were co-sponsored by government departments, and four were sponsored by a major research granting agency.

UK Data Archive Annual Report 2000/01

Two of the most common measures of activity levels of data archives are the size of their collections and the number of patrons whom they serve. For example, last year, the ICPSR at the University of Michigan added 1,835 data files to its collection, an eight-percent increase from the previous year. At the same time, it disseminated five thousand gigabytes of data to its patrons. During the same period, the UK Data Archive processed over 500 acquisitions and served 1,000 patrons who had placed 2,000 orders for a total of almost 9,000 data files. Over a three-year period, this was an increase of 2,000 data files delivered to users.

Several data archives record use statistics based on Web traffic. The Oxford Text Archive, for example, reported over 18,000 downloads of electronic texts during1999/2000. This electronic usage outnumbers Oxford Text Archive offline orders by a factor of 39. In addition to file downloads, the number of user contacts is also captured from Web statistics. For example, the ICPSR reported a substantial growth in patron contacts as a result of more users relying on the Internet for research and teaching. Over the past three years, during which more ICPSR resources were made available online, the agency reports an increase of more than one thousand gigabytes of data being accessed.

Data archives also maintain use statistics for other services. For example, the ICPSR training program consistently supports a yearly enrolment of between 500 and 540 participants. In another example, the Norwegian Social Science Data Service (NSD) maintains statistics about researchers' use of their service to investigate projects for legal compliance. NSD reports that this service has grown as much as 65% in a given year. Reference services usually maintain their own statistics. During 2000/2001, UK Data Archive staff fielded 332 post-order inquiries for assistance with data files, which represents just one aspect of reference services. During 2000 the Archaeology Data Service reported 174 total inquiries, with questions touching upon catalogue use, technical assistance, and general archaeology information. The History Data Service received approximately 480 general reference inquiries during this same period. As well as reference support, the Oxford Text Archive provided technical assessments for 125 grant applications.

Another statistic used by some data archives is the volume of licenses for software that their service develops and distributes. For example, NSDstat, which is developed and distributed by NSD, is licensed to approximately 2,000 institutions in Norway and 200 organizations internationally. While an exact number of individual users per license is unknown, experience indicates that several individuals have access to NSDstat through a single copy of the license.

Larger data archives also record statistics about their international activities. For example, the German Central Archive for Empirical Social Research (ZA) reports that they consistently have 50 international scholars each year doing on-site research with data at the ZA EUROLAB. The ZA also integrates the data and documentation for a number of international projects, including the International Social Survey Program for 38 countries and the Eurobarometers for the European Commission.

Overall, data archives that offer comprehensive services (including training, software development, and online access to data files) demonstrate significant use by researchers of a national and international scope. In every case, this use is growing.

There are many reasons to share data from NIH-supported studies. Sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new data sets when data from multiple sources are combined. By avoiding the duplication of expensive data collection activities, the NIH is able to support more investigators than it could if similar data had to be collected de novo by each applicant.

National Institutes of Health (US),
Policy Statement on Sharing Research Data

---

4. The Building Blocks of a National Research Data Archive

Over the past several years, the Government of Canada has taken major steps towards building a comprehensive and coherent research infrastructure and research support system in Canada. Measures such as the creation of the Canada Foundation for Innovation and the building of CA*Net3 have gone a long way towards filling existing gaps. One of the few gaps remaining, however, is a facility or institution with the responsibility for ensuring preservation of, and access to, research data. Nevertheless, Canada already has many building blocks for this agency in place.

University Data Services – Perhaps the most important of these building blocks are the existing university data services. Although limited resources prevent them from acting as full-service agencies, the university data services have the potential to be nodes of a National Research Data Archive. This has been strengthened enormously through the experience of the Data Liberation Initiative, where librarians and data archivists from 66 universities have come together to form a consortium to improve access to Statistics Canada data. These dedicated professionals remain in close contact with each other, sharing best practices, information about data sources, ways to improve services for their clients, and the latest advances in technical capacities and standards. With sufficient resources, university data services could form a comprehensive, nation-wide network of contact points for researchers who wish to access research data collected by others, deposit data they collected themselves, seek training in advanced statistical and data handling skills, and obtain advice on how to conform to data standards and best practices. Perhaps more importantly in the long run, the network of university data services personnel could act as a feedback system from users, helping to shape and improve the services provided by a National Research Data Archive, and ultimately the knowledge created by Canada's researchers.

Canadian Archival Institutions – As with university data services, those Canadian archival institutions with a specific research mandate offer other potential nodes in a National Research Data Archive network. They exist in local, regional and institutional environments, either as independent entities, as part of a parent institution, or within municipal, provincial and federal levels of government. Furthermore, they exist in many communities that do not host universities. While Canadian archives have, until recently, dealt primarily with non-digital records, their community infrastructure, descriptive standards, best practices, extensive experience with privacy protection and copyright, etc. all provide a firm basis from which to develop the knowledge and skills to participate in a national research data network.

International Representation – Although lacking national authority, some university data services staff currently provide one of Canada's principal connections with numerous international bodies and agencies charged with the management of research data and the establishment of international standards for metadata creation, data sharing, and preservation. The creation of these standards, agreements and common practices are vital in a scientific world that increasingly works beyond national borders. Employing their experience and expertise in a co-ordinated effort will mean that Canada's interests are represented when key decisions, with long-term implications, are being made.

Data Transmission Infrastructure – Connecting university data services is CA*Net3, and soon, CA*Net4, the ultra-high speed national optical data transmission network, built by CANARIE Inc. Now linking all of Canada's major research institutions, CA*Net3 provides the extensive pipeline necessary for the nation-wide distribution of research data. The huge capacity of this network allows for the rapid, efficient and reliable transmission of very large, complex data sets. This is crucial for the future. Research data sets are increasing in both size and complexity at an amazing rate.

Management Frameworks – The management frameworks for the use of research data are just as important as the digital pipelines and access nodes. Because of the sensitive information contained about individuals, social science data in particular must be managed within a comprehensive ethical framework, as well as Access to Information and Privacy legislation. The Tri-Council Guidelines on Research Involving Humans provides one of these frameworks. These guidelines spell out in general terms the principles by which a National Research Data Archive should treat privacy and confidentiality. Along with the university-based Research Ethics Boards, we have both the rules and the institutional capacity to ensure that information on individual citizens is protected. These Boards determine the conditions under which sensitive data can be deposited and released and so constitute a built-in, first stage screening process for a National Research Data Archive.

Research and Development – In our rapidly developing digital world, many aspects of handling research data are done without sufficient knowledge. Ensuring the quality, authenticity and security of research data are examples. A National Research Data

Archive will be positioned to capitalise on the knowledge emerging from cutting-edge research in this field, including, for example, the SSHRC-funded InterPares project.

Partner Institutions – Various institutions can play an important role in the operations and services of a National Research Data Archive. Both the National Archives of Canada and the National Library of Canada have, over the years, developed significant expertise with their respective records and in the transition of those records to electronic form. Storage environments, descriptive standards, physical and logical format migration, and protection of copyright are just some of the areas where knowledge could be shared and joint projects undertaken.

Research Data – The central building block of a research data service is the research data itself. Not all research data sets should be preserved, of course. Some will be of limited use beyond the project for which they were collected; some will contain personal identifiers that cannot been effectively removed; some simply re-produce data collected elsewhere. Determining what should, and what should not, be preserved, however, lies at the core of archival science, and is critical to an effective partnership between researchers and data archivists.

The existence of plentiful research data is not in question. In the first phase of the consultation, the Working Group determined that SSHRC-funded researchers produce, on average, some 400 data sets each year. Since SSHRC is able to support only a fraction of the Canadian social sciences and humanities research community, the total number of data sets produced each year could be three or four times this number. This does not include those data sets produced by natural scientists, health scientists or research engineers, but it is not unreasonable to estimate that some 4,000 to 5,000 are produced annually, all of which are supported by public funds. Although impossible to know in precise detail, this represents a public investment of tens of millions of dollars annually.

Government Research Data – The Working Group's investigations of data archives in other countries revealed that, in the social sciences, government-produced research data are often more widely used than data produced by researchers themselves. One valuable role for a new agency would be to provide a preservation facility, catalogue and access conduit for government collected research data. The Working Group heard testimony on numerous occasions that accessing such information is, at best, difficult and time-consuming, and at worst, impossible. Yet, it has been estimated that departments such as Statistics Canada, HRDC, Health, Natural Resources, Environment, Justice and many others spend upwards of $1 billion annually on collecting data. Finding effective and efficient means for researchers to utilise this data is a matter of good public stewardship.[2]

Preservation Services for Other Research Agencies – Today's information technologies greatly facilitate our ability to access, manipulate and apply digital information to research questions of fundamental importance to Canadians. However, the long-term preservation of digital research materials is one area, from both a technological

[2] Canadian Global Change Program, Data and Information Systems Panel, "Data Policy and Barriers to Data Access in Canada: Issues for Global Change Research", (Royal Society of Canada, 1996), p.7.

and institutional perspective, that has not kept pace. In the research world, the current emphasis is on compiling and providing access to information, predominantly through the Internet. Inter-agency cataloguing and preservation services are often considered of secondary importance or ignored altogether. The Canadian Institute for Health Information, the Canadian Centre for Justice Statistics, the Canadian Information System for the Environment, GeoConnections, and the recently announced Community Social Data Strategy of the Canadian Council on Social Development all provide excellent data access systems, but lack a well considered, adequately supported, long-term data preservation strategy. One of the most important roles that a National Research Data Archive can play is providing the preservation services and expertise for these, and many other, research data access initiatives.

Publicly funded research should require that the data generated, research instruments employed, design used and sampling frameworks etc. be archived and made available for other researchers. This would be very important to activities such as fostering collaborations, longitudinal studies, replication studies, comparative studies, creation of 'normative' question designs in certain areas of inquiry, and secondary analyses. Transparency, accountability and responsibility would be encouraged by requiring the archiving and access to data. Further, consideration of such data should become a more central attribute of planning 'new' primary research -- less re-inventing the wheel and more imaginative and creative work might result.

Questionnaire Respondent

---

5. Towards an Agency Model: Lessons Learned in the International Arena

The Working Group examined all existing national research data archives focusing on the social science or humanities (see Appendix C). This investigation included face-to-face interviews with data agency directors, comparative analysis of policies and regulations, examination of services, mandates, budgets and governing structures. Chief among the lessons learned are the following:

- Many countries have long recognized the need for a research data archive to assist and support the work of the research community. Several of the data archives examined have been in existence for 30 years or more;
- Although many services of a research data archive, particularly those related to access and training, are best distributed among a number of locations, for reasons of economy, practicality and effectiveness, preservation, network management and standards development functions are best performed within one facility;
- No two research data archives are the same. Each was established within a specific national or disciplinary context that reflected the particular needs of the research community it serves. They range in size from small, disciplinary specific, limited service organizations to large, multidisciplinary, full service, internationally networked, R&D focused, national institutions;
- Successful research data archives are directly attached to a country's research infrastructure, rather than to its archival community. They are characterised by a

service orientation that emphasises access to, and preservation of, the most useful data for research, rather than capturing records of the past;

- Research data archiving is a complex and highly technical business. Successful data archives employ dedicated, professional data experts and place considerable emphasis on training the next generation of research data managers. Developing highly qualified personnel serves the needs of both the research community and many other areas of the public and private sectors that have to deal with large volumes of data;
- There is a direct correlation between the funding stability of a research data archive and its success in supporting the research community. By its very nature, archiving is a long-term enterprise. The most useful data archives are those that are assured of their continuing existence;
- Although research data archiving requires long-term funding commitments, the institutional costs are always only a very small fraction of the costs of data collection;
- Building trust with both users and producers of research data is vital. If users cannot rely on the timely and efficient delivery of high quality data, and if depositors are not convinced that their intellectual rights and the protection of their participants will be upheld, no one will trust or use the services provided;
- The most successful data archives have both institutional independence and flexibility. They work in close co-operation with numerous government departments and universities but are not dependent upon any particular one for financial stability or decision-making. Independence is necessary to ensure that the data access needs of the research community remain the first priority, rather than the record keeping needs of government departments or traditional cultural archives. Flexibility is important for the adoption of new technologies and the ability to respond to the changing needs of researchers.

The Working Group's detailed survey of 36 institutions produced three generalised approaches to preserving and providing access to research data. Each represents the organizational characteristics of today's national data archiving services:

- **A small scale, specialised topical data archive**, usually hosted by a university department, with limited data handling capability, employing off-the-shelf technology. Clientele are often restricted to one, or a small group, of research disciplines, and annual operating budgets range from between $200K to $400K.
- **A medium sized, agency-based data archive**, whose parent organization is usually a national research institute or government department. Often located on a university campus to better serve its core research clientele, these archives base their mandate, and subsequent collection activities, on that of their parent agency. Services are moderately extensive, and staff sometimes take leadership roles in relevant national and international organizations. Annual budgets range from $500K to $1.5M.
- **A comprehensive research data archive**, servicing a wide variety of communities, including academic researchers, NGO and government policy analysts, public archival agencies, and individual citizens. Often established through legislation, such data archives are recognized as a national institutions responsible for the general principles and specific duties outlined in their founding Acts. Through one or more physical locations, and extensive use of the Internet, a comprehensive range of

services are provided, often including specialised training, educational outreach, technical support and R&D. Data management capabilities are extensive and often developed in-house. Such agencies have established working relationships with other national institutions and government departments, and staff are often leaders of international associations and actively engage in international data exchanges. Annual budgets range from $3M to $6M.

Benefits of Depositing and Archiving Data:

** Reinforces open scientific inquiry;*
** Encourages diversity of analysis and opinions;*
** Promotes new research and allows for the testing of new or alternative methods;*
** Improves methods of data collection and measurements through the scrutiny of others;*
** Reduces costs by avoiding duplicate data collection efforts;*
** Provides an important resource for training in research;*
** Ensures the safekeeping of data;*
** Allows owners to avoid the administrative tasks associated with external users and their queries;*
** Fulfils grant obligations regarding making funded research available to the research community;*
** Enables researchers to demonstrate continued use of the data after the original research is completed.*

Inter-University Consortium for Political and Social Research Web Site, University of Michigan

---

6. Core Principles and Assumptions

The Working Group concluded that a National Research Data Archive should operate according to a set of core principles. The overall objective should be to create a "trusted system" that provides the research community with an accessible and comprehensive service empowering end users to locate, request, retrieve and use data resources in a simple, seamless and cost effective way. Such a system should follow these core principles:

1) A National Research Data Archive should support the creation of knowledge by being an integral part of the research process and should aid discovery and decision-making in Canada, including the formation of public policy, by preserving and making accessible sources of evidence;
2) Whenever possible, access to research data should be as open as possible and free of charge;
3) Ensuring confidentiality, privacy and the protection of human research participants should be paramount in all operations;

4) Data collected with the use of public funds should remain publicly available, subject only to conditions of fair prior use by the depositor and the ethical and legal provisions under which the data were collected.

The Working Group heard on numerous occasions, and from many authoritative and experienced sources, that establishing trust is the key factor in building a successful research data access and preservation system. This can only be accomplished if the institution's users and depositors know that the archive is an integral part of their research processes, that it will provide useful services, and that it will add value to their work. Moreover, the data service must support and actively up-hold established regulations and guidelines regarding protection of confidentiality, privacy and intellectual property. Most importantly, in order to be a trusted system, a new agency must have long-term stability, both in its institutional structure and financing. This is one of the hard lessons learned by many data archives around the world. The source of mandate, governance, accountability and a stable, long-term commitment to providing the necessary financial resources determine success or failure.

7. Options for Canada

Drawing on these lessons and consultations with the research community, the Working Group concluded that Canada would be best served by an agency with the following general characteristics:

- A comprehensive mandate derived from, and responsive to, the needs of a wide variety of stakeholders;
- Dedication to society and the individual as the core subjects and scope of the target data;
- A service orientation that emphasises both preservation and access;
- Protection of privacy and confidentiality as a core element of its operating principles;
- The ability to process data according to international standards, engage in international data exchanges, and represent Canadian interests in international negotiations;
- The capacity to conduct advanced research and development in archival and information sciences;
- Application of the latest information and communications technologies to maximise access to research data while reducing the time and cost burdens on researchers;
- The capacity to educate and train both the producers and users of research data and the next generation of data management professionals;
- Established, on-going working relationships with other national agencies and organizations, such as the National Archives and National Library, as well as extra-governmental agencies such as CANARIE Inc.;
- Institutional memberships and other formal data exchange agreements with major data archives outside Canada, such as the ICPSR in the United States and the European CESSDA network;

- Public funding, on a long-term sustained basis, as its principal source of support. This could be supplemented by the sale of value-added data products and consultation services to for-profit organizations, but should not constitute core funding.

The Canadian context, however, shows that a National Research Data Archive must also have the following specific traits:

- A fully bilingual service;
- Access to research data produced by all levels of government, while respecting federal and provincial jurisdictional boundaries in areas such as education and health;
- Respect for, and assistance in developing, Canadian intellectual property, copyright, privacy and confidentiality legislation, regulations and guidelines;
- Close working relationships with major Canadian data producers such as Statistics Canada and provincial statistical agencies;
- Use and support of existing research infrastructure, research support services and funding support programs, including existing university data services and research libraries, the research support councils, the Canada Foundation for Innovation, the Canadian Centre for Justice Statistics, the Canadian Institute for Health Information;
- Interest in research data from both the social sciences and humanities, and, where appropriate, the natural sciences, health sciences and engineering.

Data archiving involves the long-term commitment to the resources, expertise, and public service required to ensure perpetual access to data files, to describe and document the files, and to provide access to and intellectual control of those files. One of the reasons why researchers may not be excited about this issue is that it is difficult to find out what data have been collected. It only makes sense to use economies of scale and centralize the resources required for an enterprise of this magnitude.

Questionnaire Respondent

A Canadian National Research Data Archive should meet data preservation and access needs, as well as push the boundaries of information and archival science. It should build on existing research infrastructure while learning the lessons provided by a generation of data archiving experience in other countries. Most importantly, it must be successfully adapted to fit the Canadian social and institutional context, while meeting the public need for accountability and effective governance.

In exploring how a National Research Data Archive could be created, the Working group examined existing federally-funded university research centres, sought advice from the Privy Council Office and used the guidelines provided by the Treasury Board's *Framework for Alternative Program Delivery*. The Working Group considered six possible options and discussed each in detail; reviewing institutional and governance structures, requirements for start-up and long-term stability, and both strengths and weaknesses from the perspectives of data users and producers (See Appendix D).

The Working Group first explored the option of creating a new division or Special Operating Agency within an existing national institution such as the National Archives or National Library of Canada. While the mandate of the National Archives is broad enough to extend to unpublished research data, its current level of funding could not support a move into such a new area of service, while it simultaneously responds to the government-wide challenges of information management in the era of e-government, the transition of its records into electronic form and the extensive digitization of its existing holdings. Furthermore, the failure of an earlier attempt to create a data archives division within the National Archives (1973-1986) suggests a disjunction between the broad cultural preservation role of the National Archives and the specific service role that a National Research Data Archive would be called on to play within Canada's research infrastructure. These differences extend from acquisition strategies to available staff expertise, current descriptive practices and the needs of clientele.

The National Library of Canada does collect a limited number of research data sets that meet the definition of "publications". These are, however, a small sub-set of the research data sets requiring preservation in Canada. As with the National Archives, preserving, maintaining and providing access to the two institution's current holdings do not require the extensive knowledge of quantitative research methodology, statistics and advanced computing skills necessary to meet the needs of those who would use a National Research Data Network. The Working Group believes that the unique requirements of the research community, and the research data they use, could marginalize the activities of a research data archive within these existing institutions, thus undermining the long-term stability needed for success.

Another option that the Working Group examined was the creation of what the Treasury Board refers to as a Public Partnership. This involves establishing an agency as a partnership between federal and provincial levels of government. Although this route has certain interesting aspects, it does not lend itself to the building of direct connections with the university and non-governmental research communities. As national data archives in other countries have learned, this is a crucial element in building a trusted system.

A Separate Statutory Agency or Departmental Corporation has many of the characteristics necessary for a robust, full-service and effective National Research Data Archive. It would be a permanent institution secured by legislation. It would be an element within the policy framework of the Innovation Agenda, focused on research, capacity building, stewardship and international competitiveness. It would have clear lines of authority and accountability, and a ministerial champion. Funding would be secure, stable and from a single source. Like Statistics Canada, it would have the potential, and the means, to develop a reputation as a "trusted system", and could have official national representation status in the international arena. The one important element missing is a direct connection with the research data user community.

The final option discussed, a University-Based Centre, has this direct, immediate, on-site connection. With such a facility, a sense of ownership, operations and policies would be in the hands of the associated university members. It builds on existing data services,

expertise and technology infrastructure within universities; it could use a hybrid centralized/de-centralized system, where the Centre takes care of preservation and data set processing and the associated members act as local facilities for access to data, deposit of data, on-site advice, and training activities; scope of the agency is scalable and could include NSERC and CIHR areas of science. Finally, digital archival research activities would take advantage of proximity to university-based information science researchers. The principal weakness of this option is that it lacks long-term stability. A second weakness is that it would not necessarily have the authority to act as a national voice in the international arena.

After examining and discussing all these options in detail, the Working Group concluded that the nature of the Canadian federal system of government, new communication and information technologies, the particular characteristics of the research community, and the emerging needs of Canada's knowledge economy, present a unique opportunity for institutional innovation – the creation of a hybrid agency that combines the stability of a separate statutory agency and the user community connections of a university-based research centre.

8. Recommendations for the Implementation of a National Research Data Archive Network

In order to build an effective national research data archiving service, one that best meets the needs of Canada's knowledge economy, fosters innovation, builds on the strengths of existing infrastructure, ensures effective public stewardship and gives Canada a voice in the international arena, the Working Group recommends that the Government of Canada undertake the following:

- Legislate the creation of a National Research Data Archive Network as a modified version of a Separate Statutory Agency;
- Require that this agency report to Parliament through either the Minister of Industry or the Minister of Canadian Heritage, or – preferably – a combination of the two, in an accountability structure similar to that of agencies such as the Climate Change Secretariat;
- Enable this agency to operate at arm's length, in the same manner as the federal research support councils;
- Allocate operating funds directly to both the central facility and the nodes by annual vote in Parliament, within the regular federal budget process, or alternately, flow funding through participating federal research support councils, as occurs with the Networks of Centres of Excellence.

Regarding the structures and operations of a National Research Data Archive Network, the Working Group further recommends:

- That the new agency develop a comprehensive service network, with a central facility responsible for data management, standards development and preservation and a

series of nodes, located within university research data services and other institutions responsible for providing access, depository, training and consultation services for researchers. It is suggested that institutions wishing to become nodes form a consortium to seek initial infrastructure funding from the Canada Foundation for Innovation and various provincial matching-fund research agencies;

- That a Management Board be created to govern the National Research Data Archive Network, composed of representatives from the various regions of Canada and the various stakeholder groups that manage, use and produce research data;
- That the agency develop, over time and in response to the identified needs of the research community, a suite of research data access, management and preservation services;
- That the agency develop the capacity to further our knowledge and understanding of information management sciences, ethical and legal frameworks, knowledge management practices, and promote a culture of research data sharing within the research community;
- That the agency enter into formal co-operative working relationships with other national institutions such as the National Archives and the National Library, and data access and preservation agreements with major data producers such as Statistics Canada and provincial statistical agencies;
- That the agency be given the authority to act on behalf of the Government of Canada in international negotiations related to research data management standards and common practices.

Although the Working Group is convinced that the model outlined above would place Canada at the forefront of data archiving and information science, and would substantially increase the competitive advantage of the Canadian research community, the members are also aware that the best or ideal solution is not always the most practical or feasible. With this in mind, we suggest two alternative routes to establishing a National Research Data Archive.

1) **A SSHRC National Research Data Archive Network** – following the approach taken by the Economic and Social Research Council in the UK, this would involve establishing a university-based facility and network under the auspices of SSHRC. The agency would be accountable to the SSHRC Board of Directors. It would have the same management and network structure, and range of services, as the option outlined above. Such an agency would not require enabling legislation, since it would fall within the research support function of the SSHRC mandate, but it would also lack the long-term stability that legislation provides. It would benefit from a direct connection with the research community, as well as from SSHRC's working relationships with major data producers such as Statistics Canada. Conceivably, it would be scalable to include all areas of scientific and humanities research, and could take advantage of SSHRC-funded research in information and archival sciences.

2) **A Special Operating Agency within the National Archives of Canada** – although on the surface this may seem to be the most logical route for establishing a National Research Data Archive, it should be noted that the Working Group heard very few

voices recommending this course of action. Moreover, the investigation of research data archives in other countries revealed that only one – the Danish Data Archive – is directly attached to a national archive, and anecdotal evidence suggests that this arrangement is having a detrimental effect. Nevertheless, the creation of a Special Operating Agency within the National Archives could provide a simple solution. A Special Operating Agency would be able to draw on the archival experience of the National Archives staff, use existing facilities, as well as technical and administrative infrastructure, and have the stature and authority to act as Canada's voice in the development of international standards and practices. As a Special Operating Agency it would have a degree of autonomy within the management structure of the National Archives, while still being accountable to the National Archivist. This would provide greater stability than that of the now defunct Machine Readable Archives Division. The most significant disadvantage of this approach is that the agency would not have a direct, immediate connection with the research community, either through its management structure or through the university data services. Although this could be built, the agency would still have to exist within a federal government body whose core mission is to preserve the national memory and the records of government, not service the data needs of researchers.

Researchers are in agreement that the infrastructure to allow for sharing of research data is long overdue in Canada and that we need to have a coherent infrastructure to collect, document, share, and preserve digital research data. In particular, it is critical to reduce the high costs of data collection and make files available for secondary analyses.

Submission from the University of Calgary, Office of the Vice-President (Research)

---

9. The Cost of a National Research Data Archive Network

The amount of funding required to establish and maintain a national research data service depends on the size and scope of its operations and on the range of services it provides. The key consideration is to define the minimum level of funding that would be required to provide an adequate level of services. Too small a funding base would not only restrict the range of services that could be provided but might threaten the continuation of funding. There is a danger that if the agency had to exist on too low a level of funding it would become too narrowly focused on a limited number of disciplinary areas or types of services. This in turn might arouse resentment from the users not being serviced and thus jeopardise the continuation of funding.

Funding for a National Research Data Archive Network -- both the central facility and its nodes -- should come through the Federal Government of Canada. Although supplementary funding can be secured through other routes, such as R&D grants, the sale of value-added data products and charges for speciality consultation services, general agency operations should be funded this way. This is the only effective means to ensure that the research data archive serves all Canadians, across all regions, has long-term

stability, meets the needs of a broad range of researchers and research data producers in academia, government, NGOs and the private sector.

A detailed costing of a National Research Data Archive Network, along the lines recommended here, is beyond the resources currently available to the Working Group. The international study of existing archives, however, provides a solid benchmark for the levels of funding necessary to provide certain levels of services.

In current Canadian dollars, and once fully operational, the low end of an annual operating budget for a full-service research data agency is approximately $3 million. As pointed out by the Irish Data Archive feasibility study, approximately 40% would be devoted to acquiring, processing, cataloguing and preserving data, while the remaining 60% would be spent on processes involved in servicing user needs.[3]

Initial infrastructure costs would depend on a range of factors, including the location and size of the central preservation and processing facility, the number of nodes that join the network, the distribution of specific functions between the nodes and the central facility, and the overall capacity and complexity of the computing hardware. If the agency were to be attached to Canada's research institutions, rather than the National Archives or other federal government department, infrastructure funding could be sought through the Canada Foundation for Innovation and the various provincial matching-fund agencies.

The services provided by the network, and therefore its operational costs, could be scaled up over time as both deposits and usage grows. This has been the usual route taken in other countries. The volume of data held does not significantly affect operational costs, since the price of digital storage is declining rapidly. Rather, the experience in other countries is that data handling, management, and value-added services grow as the research community uses the services and becomes aware of its real and potential benefits.

10. Next Steps

In order to initiate the process of creating a National Research Data Archive Network, the Working Group recommends the following:

- The SSHRC Board and the National Archivist create a Steering Committee to initiate the implementation process, including seeking the support of AUCC, CARL, HSSFC and other relevant organizations, enlisting ministerial sponsorship if enabling legislation is required, and securing participation of stakeholders to further develop mandates and organizational structures;
- This committee should establish appropriate contacts with Justice and Finance Department officials, and officials from other relevant departments and agencies, to begin the process of implementation and further develop the operational details of the proposed new agency;

[3] The Data Archive, University of Essex, "The Irish Data Archive Feasibility Project", 1997, p.49.

- This steering committee should be given the responsibility for developing selection criteria for the central facility and nodes of the National Research Data Archive Network;
- The committee should also advise on criteria for the composition of the Management Board, if one is called for;
- The Working Group strongly recommends that these activities begin as soon as is feasible in order to make the case that a National Research Data Archive Network can play a central role in furthering the Government of Canada's Innovation Agenda.

Appendix A
National Data Archive Consultation
Working Group and Resource Group Members

Chair -- Dr. John ApSimon, Science Advisor to the Deputy Minister, Environment Canada

Dr. Alexandra Bal, School of Image Arts, Ryerson University

Dr. Paul Bernard, département de sociologie, Université de Montréal

Professor Gérard Boismenu, département de science politique, Université de Montréal

Mr. Ernie Boyko, Director, Library and Information Centre, Statistics Canada

Ms. Sue Bryant, Treasury Board Secretariat, Senior Project Co-ordinator, Public Key Infrastructure Secretariat

Professor Joanne Burgess, département d'histoire, Université du Québec à Montréal

Professor Joseph Desloges, Department of Geography, University of Toronto

Professor Luciana Duranti, Department of Library, Archival and Information Studies, University of British Columbia

Ms. Fay Hjartarson, Director, Government Information Holdings Officer, Strategic Policy and Planning Branch, National Library of Canada

Mr. Douglas Hodges, EPMS Project Manager, Information Technology Services, National Library of Canada

Mr. Charles Humphrey, Data Librarian, University of Alberta

Professor José Igartua, département d'histoire, Université du Québec à Montréal

Professor Ian Lancashire, Department of English, University of Toronto

Professor Matthew Mendelsohn, Department of Political Studies, Queen's University

Dr. Michael Murphy, Director of the Rogers Communications Centre, Ryerson University

Dr. Frits Pannekoek, Director, Information Resources, University of Calgary

Mr. Michael Ridley, Chief Librarian, University of Guelph

Professor Geoffrey Rockwell, School of the Arts, McMaster University

Dr. Fraser Taylor, Department of Geography, Carleton University

Ms. Wendy Watkins, Data Librarian, Carleton University

Consultation Managers:

Ms. Yvette Hackett, Electronic Records Officer, Government Records Branch, National Archives of Canada

Dr. David Moorman, Senior Policy Advisor, Social Sciences and Humanities Research Council of Canada

Appendix B
Terms of Reference

Context

In June 2001, the National Archivist and the SSHRC Board of Directors accepted the Phase One Needs Assessment Report of the National Data Archive Consultation, including the recommendation that the consultation move to Phase Two. The SSHRC Board also asked the Working Group to undertake activities to increase awareness of data archiving issues amongst the research community and to explore the possibility of bring new partner organizations into the consultation process. In addition, both the National Archives and the SSHRC Board requested revisions to the original Terms of Reference for Phase Two.

Terms of Reference for Phase Two

In Phase Two of the consultation, the Working Group, with the assistance of the Resource Group, are asked to examine the following questions, and report back to the SSHRC Board and the National Archivist:

- Given that a new national research data archiving system or function is recommended, how should it be implemented, what functions should it perform and what institutional form should it take?
- How can a data preservation and access system or function best take advantage of emerging information and communication technologies to increase its efficiency and effectiveness?
- What should be the scope of holdings managed by a new system or function?
- How should a new system or function deal with ethical and legal issues, such as confidentiality, privacy, consent, intellectual property, copyright, and epistemological concerns, as well as issues such as authenticity and security of data?
- What is the most appropriate working relationship between a new system or function and existing agencies such as the National Archives of Canada and the National Library of Canada, particularly in relation to government-produced data? How should duplication of responsibilities and services be avoided?

Upon completion of its investigations, the Working Group are asked to submit to the lead agencies a full report on all matters, along with recommendations for action by the Council, the National Archives, universities and other relevant bodies.

Appendix C
International Models of Data Archiving Services

One project of the Research Sub-committee of the National Data Archive Consultation was to investigate existing models of international data archiving services in the social sciences and humanities. A template of functions and services was constructed by the Sub-committee to guide the collection of information from thirty-six archiving and data services that represent the leading international organizations of their kind (see Appendix A for a complete list of these institutions). Detailed information was harvested from each institution's Web site completing as much of the template as possible from these online sources. This research was followed by e-mail correspondence with each institution to confirm the accuracy of the information gathered and to fill in details that were not available on these Web sites. A total of twenty-two institutions participated in both phases of this research, while information from only the Web was gathered on the remaining fourteen services.

Twenty-eight of these institutions are dedicated to the social sciences while the other nine serve the humanities. Five of the services in the humanities (ADS, VADS, HDS, PADS, and the OTA) are specific subject-providers for the Arts and Humanities Data Service in the United Kingdom. Most of the organizations are national, if not international, in scope. Services whose focus seemed particularly regional were not included because of our specific interest in national-level data archiving services.

The report below presents a typology of organizational models based on the investigation of these thirty-six institutions. The description of these models is followed with a summary of the characteristics used to construct this typology.

A Typology of Organizations for Data Archives

A typology of organizational models for data archives was developed from the findings of this study. Based on the information from thirty-six existing international organizations, three generalized models have been identified that summarize groupings of characteristics amongst these institutions[4]. As generalizations, these three models may not describe any one particular organization, but are intended to represent general categories of organizational models. Like an arithmetic mean, the value of the mean may not be any of the actual values upon which it is computed. Nevertheless, the mean is taken as a representative value for the batch of numbers upon which it is derived. Similarly, the three models below represent organizational characteristics of today's international data archiving services. While no single organization may be completely described by one of the three models, the typology offers a fair summary of the current mix of organizations.

[4] An additional organizational model was identified during the deliberations of the Research Sub-committee. Although this model was not evident among the institutions of this study and furthermore was deemed to be inappropriate by this Consultation for a national data archiving service, its description is presented in Appendix C for the sake of completeness.

Model A: The Topical Data Archive

This type of organization has a narrowly focused mandate to serve a specific field of study or a few closely related disciplines. The location of the service is in a university setting and is often attached to an individual academic department. The origin of the mandate is likely to come from the host-university and the financial base for the service dependents on in-kind support from the host. Additional financial resources are derived through contract work and user fees. The complete budget will fall between $200,000 and $400,000 Cdn.

The organizational structure consists of one unit staffed by up to five full-time employees and an equal number of part-time student employees. The level of data processing to prepare materials for its archive and for dissemination is kept to one or two steps. The formats in which materials are received are limited to a few widely accepted practices to minimize the extent of processing required to accession products. Access to the collection is provided through the Internet and the service's dissemination strategy is dependent on individual users retrieving products directly online. Off-the-shelf technology is used to support the operations of the unit.

The unit is the only one of its kind in its country and is recognized by other institutions as the national location for this service. Over time, the staff accumulate expertise in this specialized service and eventually receive international recognition for the skills that they develop.

Model B: The Agency-based Data Archive

This type of organization receives its mandate from the agency within which it originated. This parent agency is likely to be a national institute or federal department with a strong research mandate. The parent organization may choose to locate its data archiving service on a university campus to maintain close contact with those researchers with whom the agency has a working relationship. The data archive's mandate is part of the larger mandate of the parent organization. Institutional funding serves as the primary source of revenue for the service, although its financial base may be supplemented with cost recovery charges or user fees. The complete budget will range between $500,000 and $1,500,000 Cdn. Some operational expenses will be absorbed by the parent organization, such as the overhead for administrative records and management, and consequently, will not be specifically identified in the data archive's budget.

The organizational structure consists of two to three units staffed by 10 to 30 full-time employees. The number of part-time employees depends on the proximity of the service to a university and the availability of student labour. Three or four processing steps are taken to prepare data and documentation for its archive and for subsequent dissemination. Materials are received in a variety of formats and converted to the archive's standards. These standards are published and available on the service's web site. Access to the collection is provided directly online as well as through a mediated service, which is

required because of restrictions placed on products to ensure confidentiality. One staff member is dedicated to negotiating and establishing privacy assurance on all data products. Reference services are offered to assist patrons in locating data and to field inquiries about products.

While this archive may not be the only one of its kind in its country, it is recognized as the authoritative service for the type of research directly associated with its parent organization. The staff develop a level of expertise that is acknowledged nationally and internationally. They take leadership roles in national and international organizations relevant to their areas of expertise and to the parent organization's research interests. The senior managers of the service participate in international projects and exchanges relating to common data and research interests, including the development of new technology and standards.

Model C: The Comprehensive Research Data Archive

The mandate of this data archive arises from a variety of communities sharing a common interest and concern. These communities include academic researchers, policy analysts in various levels of government and non-governmental agencies, civil servants charged with preserving public information, individual citizens interested in access to information, and participants in the Knowledge Economy who create and use research data. Through legislation, a mandate is articulated to serve the interests of these communities. The data archive is recognized as a national institution responsible for the general principles established in legislation.

This service may have more than one physical location. For example, it may choose to locate some of its operations on a university campus and other operations in close proximity to major data producers. The primary source of revenue for its operations is institutional funding, although it supplements this source through contract work with major data-funding agencies and producers in the public and private sectors. The complete budget ranges between $3,000,000 and $6,000,000 Cdn. Some operational expenses may be contracted with other organizations, such as a university where a branch operation is located.

The structure of the data archive consists of five or six major units that support administration, archives/collections, technical support, research & development, reference services/user support, and educational outreach. These functions require a staff of between 60 and 100 employees, with up to 20 of these positions begin part-time. The archives/collections unit is responsible for the selection of materials regardless of discipline or field of study. Critical to the selection process is an assessment of the research value of a data product. Up to six processing steps are utilized to prepare data and documentation for its preservation and access.

Employing its own standards, the data archive receives and processes materials in a wide range of formats that are subsequently converted to archival standards. Details about these standards are not only easily accessible from the service's web site but workshops

about these standards are conducted for data producers as part of the organization's educational outreach. A sub-unit within archives/collections is dedicated to issues of privacy and intellectual property to ensure that these rights are protected while also ensuring that valuable research data are preserved. Conditions of access are negotiated and managed by this sub-unit.

A reference services/user-support unit exists to help patrons access the collection. Some of the data will be available unmediated online through the Internet. This unit will also support access to other data that are restricted or have conditional requirements on their use. Educational workshops will be conducted to prepare users to work with data in the collection and to promote data literacy skills generally.

This service is recognized as having a national mandate for archiving research data. However, it works closely with other national agencies, such as the National Library, National Archives, and other archives in the country, to coordinate and ensure the preservation of research data. The staff of the data archive are experts in their field, which is acknowledged nationally and internationally, and are leaders in national and international associations connected with their profession. Staff also engage the data archive in international projects and data exchanges on behalf of their country. Its research & development unit is internationally recognized for its adaptation and use of technology to push back the frontiers in the preservation of and access to data.

This concludes the three generalized models for data archiving services summarized from the characteristics of thirty-six existing institutions. The following section reviews the findings upon which these generalizations were drawn.

The Research Findings

Below is a summary of findings from this research. See Appendix B for a complete listing of the template of functions and services used to direct the gathering of information for this project and Appendix D for the details collected about each institution.

Physical Location

The majority of these organizations (25 in total) are located within a university, or a department in a university. Four of the remaining services exist within a larger social science institute. These include the SDA, SDB, SADA, and SIDOS. The rest of the archives are independent organizations (TARKI, ICSSR, CIS, DDA, SA, ADP, and WISDOM).

Legislated Mandate

Out of the 36 organizations, only six reported that they operate under the mandate of some form of legislation. However, each of these services offered different interpretations about what constitutes a legislated mandate. The CIS in Madrid operates

under a Royal Decree from the Spanish government. This is a smaller organization that also functions as a survey centre for the country. A larger organization, such as the ICPSR, has its own constitution and bylaws, and directly serves a paid membership of data providers and users. A few of the organizations fall under the legislation of their host organization. For example, the ESSDA's statue was set out and approved by Tartu University, where the archive is located. The United Kingdom Humanities data service providers under the AHDS are currently developing Service Level Agreements within an umbrella organization that will define responsibilities and provide clear objectives.

Governing Body Structure

Twenty-two institutions indicate some form of governing body structure. Out of these 22, ten have a governing or advisory council made up of social scientists, researchers, or institution representatives (universities, members, etc.) with specific interest in their service's data. The number of advisors on these Councils ranges between 6 and 35.

The remaining twelve organizations mention a larger entity or consortium that governs their services. For example, ZA is governed by the German Social Science Infrastructure Services. The Research Council of Norway is the governing body of the NSD, whose governance structure appears to have other responsibilities and organizations on which to focus. Another example includes the Slovenian ADP, which is governed by the Ministry of Education, Science and Sports.

Agency Source of Mandate

With regard to the agency providing the source of mandate, eight organizations report to a separate institute and eleven to the university in which they are located. Some institutions receive their mandates from both their host university and a separate institute. For example, the ISSDA's mandate is set out jointly by the University College Dublin and an independent institute. Most of the service providers of the Arts and Humanities Data Service receive their mandate from the AHDS itself, although some of them do pre-date the AHDS (for example, the Oxford Text Archive). Other organizations, such as the ICPSR and SADA receive their mandate from consortia or institutes for social research. Only two organizations noted that they receive their mandate directly from the federal government: the ICSSR in India, and the CIS in Spain. TARKI functions as a stand-alone organization with shareholders, and is therefore independent of any government agency. The Linguistic Data Consortium (LDC) in the United States appears to be accountable only to its membership. Eight organizations do not have an agency-based source of mandate.

Formal and Informal Relationships

In our follow-up with organizations, definitions of 'formal' and 'informal' relationships tended to be interpreted by the respondents even though our Research Assistants offered definitions. Formal relationships were described as agreements among institutions or individuals regarding data sharing, exchange, or the provision of services, while informal relationships emphasized associations with others without contractual commitments, such

as a membership in IFDO or CESSDA. However, some organizations persisted in defining their membership in IFDO or CESSDA as a formal relationship. For the purposes of this report, memberships in another organization will be counted as informal relationships.

Most archives noted several formal and informal relationships with other organizations or individuals. Many said they maintained relationships with universities, university departments, different levels of government (usually on a national level), international data archive organizations, and other national data archives, including paid memberships with other archives. Only CETH said they did not have formal relationships with any other organization.

The most commonly reported relationships (22 in total) were with other national institutes or data archives. Ten other institutions said that they maintained formal relationships with university departments. Seventeen noted formal relationships with international organizations, and six mentioned government or government departments. The ICPSR, LDC and SSDA-ANU are the only ones maintaining a membership structure with users.

A number of informal relationships were also mentioned. Most of these (17 in total) seem to be regional in focus and of this group, four specified connections with educational institutions. Twenty-one organizations noted that they are members of CESSDA; fifteen belong to IFDO; and four identified their membership in the ICPSR. As an example of another international affiliation, nine institutions mentioned association with the International Social Survey Program.

Budget Summary

Seventeen organizations confirmed their budget information. These institutions generate revenue through a variety of sources, including grants, in-kind support from host institutions (usually universities), contract work, user fees, and some cost recovery.

Institutional Funding:
Institutional funding, defined as support from the organization's host institution, is the primary source of funding mentioned by most organizations (29 in total). Of these, twelve organizations receive institutional funding that make up at least half their budget. These include FSD (100%), SSJDA (100%), ESSDA (100%), SIDOS (90%), DDA (90%), ZA (80%), SA (90%), CIS (79%), ISDC (70%) and SSD (65%), HDS (50-60%), and UVA (50%). On the other hand, VADS, NSD, and the ICPSR report operating with only 30%, 23%, and 11%, respectively, from institutional funding. Thirteen organizations receive this type of support but did not indicate the proportion that this is of their overall budgets. Three organizations, all service providers for the Arts and Humanities Data Service, listed 'in-kind' support from their host institutions.

Granting Agencies:
Twenty-two organizations receive grants for their work, in varying proportions. PADS receives 100% of their budget from grants; ADP, VADS and ADP reported 70%, NSD

56%, and ADS, UVA and HDS 50%. The ICPSR generates 36% of its revenue from grants. Four other institutions generate 15% or less of their budgets from grants. Thirteen organizations did not confirm a percentage based on grants, while nine said that they do not receive any grants.

User Fees:
Sixteen institutions charge user fees. Of these, seven did not specify the proportion of their budgets based on this source. The remaining group noted that user fees make up a small proportion of their revenue – usually under 5%. Only the ICPSR (31%), the ISDC (20%), and the NSD (21%) received a larger percentage of their revenue from user fees.

Contract Work:
Eleven institutions receive revenue from contract work. The ADS generates 50% of its income from contracts, which is the highest percentage reported for this source amongst these organizations, followed by the ICPSR at 20% and the CIS at 16%. SIDOS, SSD, SA and ADP generated 5% or less of their revenue from contract work. Four organizations receive money from this source, but did not indicate what percentage it is of their total budget. Sixteen organizations receive no revenue from contract work.

Other Sources:
Only five organizations reported other sources of revenue. These include Hungary's TARKI, which generates revenue from investments and stockholders. The ICPSR notes 2% of their revenue comes from 'indirect cost recovery,' while ADP and SSD (20% and 15%, respectively) cite one-time grants from their governments as part of their budgets for a specific year. The UVA receives some funds from an endowment. The great majority of organizations (31 in total) do not receive income from other sources.

Categories of Expenditures:
The major categories of expenditures listed by organizations were Staffing/Salaries, Administration/Operations, Archives, Web Development, Publicity, and Technology. "Travel/Conference" was mentioned by one organization. Some of those receiving in-kind support from their host institution (such as the ADS, PADS, UVA, SA, and other service providers of the AHDS) did not list operational services as part of their expenditures. Most respondents did not go into great detail about their expenditures listing only up to five categories of budget items. The UKDA and the SIDOS are the exceptions, listing more than five.

Total Size of Budget:
Seventeen institutions reported total annual budget figures. The CIS operates with the lowest budget, at just over $9,000 Cdn, followed by ADP from Slovenia at $42,000 Cdn. The largest budgets are held by the ICPSR ($5.7 million Cdn.), the ZA ($5.03 million Cdn.), the UKDA ($3.5 million Cdn.), and the DDA ($1.397 million Cdn). Budgets of the other organizations range from $958,000 Cdn (SIDOS), $838,000 Cdn (FSD), to $565,000 Cdn (ADS). VADS reported their budget as "confidential." The SA defined their budget as 'shared' within NIWI, and could not provide a single figure.

Staff

The largest organization is ICPSR, with over 100 staff. Those with 50-100 staff are ZA, NSD, CIS, CIDSP, and the UKDA. ICSSR and SDB are staffed with between 20 and 50. Six organizations have between 10 and 20 staff (SSD, FSD, DDA, SSDA, SSJDA, and TARKI), and the rest are under 10 full-time equivalent staff. Ten of these employ less than five staff.

Organizations were asked to categorize staff roles under three functions: Administrative, Technical, and Professional categories. *Professional* was defined as archivists, data librarians, or similar professional designations. *Administrative* includes managers, directors, or other staff that focus on the operations of the service. *Technical* captures those involved in technological aspects of the organization and its archive holdings.

Most respondents noted fewer than ten staff dedicated to any of these three functions. The SA said, as with their budget, that being a part of NIWI meant that they could not isolate the number of staff dedicated to administrative or technical functions. Several organizations did not provide numbers for one or two of these functions.

Institutions were also asked to indicate the number of their staff who is dedicated to 'preservation' and 'access' functions. 'Access' includes anything regarding user support, membership outreach, or other initiatives related to users and data. 'Preservation' focuses on procedures and operations to maintain the contents of its data archive.

Most organizations reported fewer than ten staff are dedicated to either preservation or access. Eighteen archives did not respond to this question.

Temporary staff are employed by thirteen of the thirty-six institutions. Seven archives in this group employ under ten temporary employees on an annual basis, while six employ more than ten. Four organizations do not employ any, and nineteen archives did not respond.

Organizational Structure

Over half of the institutions surveyed exist under 'flat' structures, with seventeen noting they are designated as a single unit of operation. Larger organizations such as the UKDA and the ICPSR have fifteen and five units, respectively. SIDOS, a medium-sized organization, has six different organizational units. ADS, CIS, and NSD maintain four units (although ADS noted that their units are not 'formally defined'). Five organizations maintain three units each (SDB, TARKI, HDS, LDC, ZA), while six institutions have two units (ADPSS, SSJDA, DDA, FSD, SA, CIDSP).

The most frequently-stated unit division titles are Administration, Archive/Collections, Technical, Research, User Support/Services, and Education. Some categories that received a single mention were External Projects and Methodology, as well as Eurobarometer by SIDOS.

Scope of Archive

Twenty-eight organizations focus on the social sciences, five on history (including SSD and UKDA, which also collect social science data), five on fine arts, and one on linguistics. These organizations receive data from a great variety of sources, including multiple sources. Many (26 in total) have a primary focus on researchers, academics, or other data 'creators'. Government departments or other public sector organizations contribute data to sixteen organizations. Thirteen utilize commercial resources. Four either generate their own data or receive it from an institution within which they are located, and seven receive data from other organizations of a national scope within their country. Only two noted that they collect data using an international exchange partnership (ICPSR, ISDC).

Data collected are usually national in scope and focus on the country in which the organization is located. One notable exception is the LDC, which collects linguistic data from around the world, and another is the ICPSR and the ISDC, which are involved in international data exchange.

Subject areas are varied. For example, the service providers for the Arts and Humanities Data Service focus on one or two different subjects (e.g. PADS, ADS, VADS). Most of the social science institutions collect in a wide range of fields compared to the humanities archives, which seem to have a narrower collections mandate. A few organizations collect across the social sciences and humanities. For example, the Swedish SSD collects data in both the social sciences and history.

Twenty archives have written acquisition policies with sixteen of these policies being accessible online. Seven archives do not have formal policies.

Eighteen organizations publicize their procedures for data depositors online, in varying amounts of detail. Of the seventeen that do not publish deposit procedures, three of these (CETH, UVA and CIS) are humanities services focusing on projects that arise within their organizations.

Twelve organizations house topical sub-archives. These include both social science and humanities archives. Sub-archives are divided by various criteria: qualitative and quantitative research methods (FSD), literature by language of origin (UVA), source format (linguistic categories at LDC) and subject (e.g. ICPSR, ESSDA, etc.). The UVA had the highest number of sub-archives, totaling fourteen, while the remainder had between two and five. Fourteen organizations responded that they do not maintain sub-archives.

Depositing Data

Fifteen organizations reported information about the number of processing levels that they employ after data are deposited. Of these, eight specified exact procedures. The ISDC in Israel and the ICPSR in the United States use four to five levels of processing. The NHDA in the Netherlands uses nine, ZA uses eight, and the FSD up to three. Smaller

organizations, such as CETH, accept most formats, and their procedure involves simply accessioning the dataset. The SSD in Sweden also differentiates between full and partial processing, depending on the quality of the dataset deposited. Twenty-one organizations did not state the number of procedures they use, but the step of accessioning datasets is assumed.

Procedures most often mentioned are: checking for errors within the data, verifying anonymization, creation of or checking documentation, conversion to accepted format, standardization of markups in XML or SGML, creation of metadata or finding aids for dataset, backups, and finally, data release.

Many archives accept a wide range of formats, particularly the humanities archives preserving text-based data. Eighteen archives publish their accepted formats online for depositors, while ten do not publish this information.

Twenty-four archives also publish on their web sites, or as hard copy documents, the formats to which they convert data. CETH, at Rutgers University, maintains only XML-marked texts, which are in a standard format. The most frequently mentioned formats include SPSS (for social science data archives), PDF and TIFF (for codebooks or other documentation), XML, DDI Standard (Data Documentation Initiative), ASCII and MS Word.

Other descriptive standards mentioned most often include XML/ASCII, PDF, DDI , and RTF. Twenty-two archives have their documentation descriptive standards published online for depositors. Twelve organizations do not publish this information online, and the UVA and CETH do not use descriptive standards.

Primarily, social science archives are concerned with anonymization of individual-level data. Sixteen archives have a written policy that outlines their anonymization procedures and priorities; most humanities archives did not specify this as a priority (other than the History Data Service and the Archaeology Data Service for some kinds of data).

Sixteen archives have a written policy regarding off site storage of data, while twelve do not. Most common formats for off site storage mentioned include RAID, CDs, diskettes and backup tapes. Two of the five service providers for the AHDS (ADS and PADS) mentioned that their policies for off site storage are in progress and under discussion with the AHDS. This implies that all AHDS service providers may be coordinating this function centrally with their umbrella organization.

Several archives noted that their disaster recovery procedures are linked to off-site storage. This includes CIS, CETH, FSD, NHDA, ZA, and the SA. No archive noted a complete written policy, although the ICPSR, ISDC, ADS, and PADS said their policies are ‘in progress.’ Thirteen archives responded that they do not have a policy in place.

User Groups and User Services

Thirty-three organizations note academic audiences as a major source of patrons. Seventeen reported the proportion of academic users being between 80 to 100%, four from 41 to 80% and two between 1 to 40%. Twelve institutions acknowledged that academic users are part of their audience but did not give a percentage.

A lesser number agreed that other public users were amongst their audiences; ten organizations estimated the percentage of other public users being between 1 and 40%. Ten additional institutions did not give a percentage of users, but did list public individuals or groups amongst their users. Nine archives do not have public users.

Twenty organizations note private or corporate users amongst their audience, but most of these were under 40% of their users. Of these twenty organizations, twelve did not give proportions compared to other user groups, and thirteen said they have no private users. Only the CIS in Spain said 52% of their users are private or corporate.

With regard to services, all organizations publish a web site, with varying levels of information and detail. A noted difference between large and small archives is that the smaller organizations offered fewer services, and definitely fewer services online. Twelve organizations offered reference services for users, and of these, three are service providers to the Arts and Humanities Data Service. Six archives do not offer a reference service. Fourteen offer reference service through contact with their staff, via e-mail or phone. Most organizations operate their reference service like a call center, with service offered through contact with their staff.

The majority (twenty-two) of the organizations provide some kind of training through individual workshops or a more extensive 'summer school' program (like the ICPSR). The CIDSP's training program is primarily for their temporary staff, who are PhD students at their institution. All of the organizations from AHDS provide workshops that appear to be coordinated through their umbrella organization.

Publications created and disseminated by the organizations include newsletters, annual reports, CD ROMS of data, publications on data and data archiving, and in the case of the Arts and Humanities Data Service, "Guides to Good Practice" for humanities data preservation and archiving. As mentioned previously, all of the organizations publish at minimum a web site; all but four create other publications.

Fourteen organizations offer data extraction services. Some of these, such as the humanities archives at CETH and ADS, consider data extraction to be the realization of their own projects. Others, such as the CIS, do their own surveys and extract data from these sources. Thirteen institutions reported that they do not extract data for patrons.

It appears that most national archives offer a mix of direct and indirect access to their datasets. However, direct access to data files seems to be a current priority for many

national archives. Twenty-five reported that they provide at least some direct access to their collections. Some of these are through another coalition or organization. Twelve do not offer any direct access.

Direct access appears to be what the majority of archives are working toward, if they do not offer it at this time. Twenty-two offer mediated access to data, while nine do not provide access in this way. Of these nine, the majority are humanities archives.

Service summaries reflect a wide range of priorities in terms of what these organizations offer users. As well as the services discussed in the above paragraphs, most archives maintain library functions, and many publish an online catalogue of their collection for users to search. Some of the unique services reported include data analysis (NORC, SSD, SA, ICPSR), software development (NSD, LDC, ZA, ICPSR, CETH), dataset building (ISDC), the provision of fellowships for study (ICSSR), 'data search' services in other archives (VADS, SSDA), and the creation of independent projects within institutions (CETH). The CIS creates and disseminates their own surveys, and NORC will design questionnaires and other data collection tools. Some also participate in international projects and data exchange, such as the ICPSR.

Organizations such as the ICPSR and the UK Data Archive are very comprehensive in the services that they offer users. Archives that are medium to large in size also tend to be involved in international projects and data exchanges (LDC, ISDC, SSD, ZA, NESSTAR members). Smaller archives with fewer services, include ADPSS in Italy, SSJDA in Japan, ISSDA in Ireland, and CIS in Spain. The activity of these organizations tends to stay within a national focus.

Monitoring Usage

Two of the most common metrics used to monitor the activity levels of data archives are the size of their collections and the number of patrons whom they serve. For example, in the 2000/2001 fiscal year the ICPSR added 372 data titles to its collection, which was an eight-percent growth rate from the previous fiscal year. These new titles contained 1,835 data files or a four-percent increase from the previous fiscal year. Regarding patrons served, the ICPSR disseminated five thousand gigabytes of data to its patrons. During the same reporting period, the UKDA processed over 500 acquisitions and served 1,000 patrons who had placed 2,000 orders for a total of almost 9,000 data files. Over a three-year period, this was an increase of 2,000 data files delivered to users.

Several archives record usage statistics based on Web traffic. The number of data and documentation downloads is typical of this type of measurement. The OTA, for example, reported over 18,000 downloads of electronic texts during the 1999/2000 fiscal year. This electronic usage outnumbers OTA offline orders by a factor of 39. In addition to file downloads, the number of user contacts is also captured from Web statistics. For example, the ICPSR reported a growth in patron contacts as a result of more users relying on the Internet for research and teaching. Over the past three fiscal years during which

more ICPSR resources were made available online, they report an increase of more than one thousand gigabytes of data being accessed.

Data archives also maintain usage statistics for other services. For example, the ICPSR training program consistently supports a yearly enrolment of between 500 and 540 participants from a variety of disciplines and countries. In an another example, NSD maintains statistics about researchers' use of their service to investigate projects for legal compliance. NSD reports that this service has grown as much as 65% in a given year. Reference services usually maintain their own statistics. During the 2000/2001 fiscal year, UKDA staff fielded 332 post-order inquiries for assistance with data files, which represents just one aspect of reference services. The ADS reported 174 total inquiries during 2000, with these questions touching upon catalogue use, technical assistance, and general archaeology information. The HDS received approximately 480 general reference inquiries during this same period. As well as reference support, the OTA provides technical assessments of grant applications. In the 2000/2001 reporting year, OTA provided consultation on 125 grant applications.

Another statistic used by some data archives is the volume of licenses for software that their service develops and distributes. For example, NSDstat, which is developed and distributed by NSD, is licensed to approximately 2,000 institutions in Norway and 200 organizations internationally. While an exact number of individual users per license is unknown, experience indicates that several individuals have access to NSDstat through a single copy of the license.

Larger data archives may also record statistics about their international activities. For example, the ZA reports that they consistently have 50 international scholars each year doing on-site research with data at the ZA EUROLAB. The ZA also integrates the data and documentation for a number of international projects, including the International Social Survey Program for 38 countries and the Eurobarometers for the European Commission.

Overall, data archives that offer comprehensive services (including training, software development, and online access to data files) demonstrate significant usage by researchers of a national and international scope.

Evaluative Functions

Fifteen organizations track the use of their data through publications of researchers, often through requests to researchers to keep the archive informed about new work that arises from the data that they have provided. A number of these note that, while they wish tracking was mandatory, it is 'not enforced.' Twelve organizations do not track publications, but two of these (PADS and ICPSR) said that a tracking tool is ‘in development.’

Even fewer organizations track citations that refer to their collections. Only six report that they perform this function. One of the humanities service providers for the Arts and

Humanities Data Service, PADS, notes that this function is in development and should be in place within two years. The ICPSR does publish a citation guide on their website for researchers to use.

Thirteen institutions track usage of the data they provide directly to users via the web. Several archives note that they simply track the hits on their direct data website and not the usage of the data in a formal way. These include ICPSR, OTA, PADS, and NHDA. The Linguistic Data Consortium in the United States only tracks usage if the data depositor requests it. Thirteen archives report that they do not track data usage.

Maintaining an intellectual property policy is a major priority for twenty-seven organizations, which is reflected the procedures they have written to protect creators' or researchers' rights. Privacy policies seem to be more of a concern for social science archives than ones in the humanities. In fact, three humanities archives (particularly PADS, CETH and UVA) noted that they do not see the need for a privacy policy regarding their data.

Technology

Responses to questions about technology were varied. Not all organizations responded to this part of the questionnaire. Consequently, information was often gathered from web sites, annual reports or other publications. For those that did respond, storage capacity for the collection ranged from 4GB (SIDOS) to 83GB (NHDA). Archives supporting a larger storage space include CETH (70GB), ICPSR (with high speed RAID disk storage), and the SSD (50GB). The History Data Service, as part of the UK Data Archive, shares storage space with its parent organization, and PADS noted that their systems will soon move under the administration of their parent university.

Security protocols were mentioned by only two organizations, both of which noted that their systems were protected by a firewall (in the case of the SA, by its parent institution).

The most frequently mentioned platforms are Windows 95-2000, UNIX, NT4 servers, Linux servers, and Sun 4500 Enterprise servers. Most of the organizations note that they use open-source software, which permits them to customize programs to their needs. Many mention using Microsoft Office for staff computers. RAID storage was mentioned by a few institutions for the purpose of backup. Text archives, such as UVA and CETH, also employ scanning programs, such as Photoshop for image manipulation and OmniPage Pro for OCR text scanning.

Some archives use commercial databases, including Dbase, Data Perfect, Filemaker Pro, FoxPro, Oracle, and Access. The most frequently mentioned statistical packages and utilities include SPSS, SAS, and StatTransfer. Several organizations note the use of XML (including DDI) as a coding convention, particularly in the humanities archives.

Data Services and Archives in the Social Sciences and Humanities Included in This Study

Acronym	Name and Country
ADP	Data Archive for Social Sciences, Slovenia
ADPSS	Data Archive for Social Sciences, Italy
ADS	Archaeology Data Service; part of the Arts and Humanities Data Service, UK
CETH	Center for Electronic Texts in the Humanities, USA
CIDSP	Information Centre for Socio-Political Data, France
CIS	Centre for Social Research, Spain
DDA	Danish Data Archives
ESSDA	Estonian Social Science Data Archives
FSD	Finnish Social Science Data Archive
HDS	History Data Service; part of the Arts and Humanities Data Service, UK
ICPSR	Inter-University Consortium for Political and Social Research, USA
ICSSR	Indian Council for Social Science Research
IRSS	Howard W. Odum Institute for Research in Social Science, USA
ISDC	Israel Social Sciences Center
ISSDA	Irish Social Science Data Archive
LDC	Linguistic Data Consortium, USA
NHDA	Netherlands Historical Data Archive
NORC	National Opinion Research Center USA
NSD	Norwegian Social Science Data Services
NZSRDA	New Zealand Social Research Data Archive
OTA	Oxford Text Archive; part of the Arts and Humanities Data Service, UK
PADS	Performing Arts Data Archive; part of the Arts and Humanities Data Service, UK
SA	Steinmetz Archive, Netherlands
SADA	South African Data Archive
SDA	Sociological Data Archive, Czech Republic
SDB	Social Data Bank, Greece
SIDOS	Swiss Information and Data Archive Service for the Social Sciences
SSD	Swedish Social Science Data Service
SSDA-ANU	Social Science Data Archive – Australian National University
SSJDA	Social Science Japan Data Archive
TARKI	Social Research Centre Inc., Hungary
UKDA	UK Data Archive, United Kingdom
UVA	Electronic Text Center, USA
VADS	Visual Arts Data Service; part of the Arts and Humanities Data Service, UK
WISDOM	Vienna Institute for Social Sciense Documentation and Methodology, Austria
ZA	Central Archive for Empirical Social Research, Germany

Alphabetical Listing by Name of Institution

Data Archive for Social Sciences, Slovenia (ADP)
Data Archive for Social Sciences, Italy (ADPSS-Sociodata)
Archaeology Data Service; part of the Arts and Humanities Data Service, UK (ADS)
Center for Electronic Texts in the Humanities, USA (CETH)
Information Centre for Socio-Political Data, France (CIDSP)
Centre for Social Research, Spain (CIS)
Danish Data Archives (DDA)
Estonian Social Science Data Archives (ESSDA)
Finnish Social Science Data Archive (FSD)
History Data Service; part of the Arts and Humanities Data Service, UK (HDS)
Inter-University Consortium for Political and Social Research, USA (ICPSR)
Indian Council for Social Science Research (ICSSR)
Howard W. Odum Institute for Research in Social Science, USA (IRSS)
Israel Social Sciences Center (ISDC)
Irish Social Science Data Archive (ISSDA)
Linguistic Data Consortium, USA (LDC)
Netherlands Historical Data Archive (NHDA)
National Opinion Research Center USA (NORC)
Norwegian Social Science Data Services (NSD)
New Zealand Social Research Data Archive (NZSRDA)
Oxford Text Archive; part of the Arts and Humanities Data Service, UK (OTA)
Performing Arts Data Archive part of the Arts and Humanities Data Service, UK (PADS)
Steinmetz Archive, Netherlands (SA)
South African Data Archive (SADA)
Sociological Data Archive, Czech Republic (SDA)
Social Data Bank, Greece (SDB)
Swiss Information and Data Archive Service for the Social Sciences (SIDOS)
Swedish Social Science Data Service (SSD)
Social Science Data Archive - Australian National University (SSDA-ANU)
Social Science Japan Data Archive (SSJDA)
Social Research Centre Inc., Hungary (TARKI)
UK Data Archive, United Kingdom (UKDA)
Electronic Text Center, USA (UVA)
Visual Arts Data Service; part of the Arts and Humanities Data Service, UK (VADS)
Vienna Institute for Social Sciense Documentation and Methodology, Austria (WISDOM)
Central Archive for Empirical Social Research, Germany (ZA)

The Volunteer-based Virtual Data Archive/Service

A fourth type of organizational model was identified during the deliberations of the Research Sub-committee. This additional type is characterized as a Virtual Organization (VO)[5] existing on a distributed, wide-area network and staffed by volunteers sharing digital resources. While none of the thirty-six organizations included in this study embody this model, some projects outside of this research were identified as exemplifying a VO. For example, the Gutenberg e-text project is supported by a group of volunteers who key books, for which copyright has expired, into computer files. These files are subsequently submitted to a central site on the Internet where the electronic versions of these texts are available for dissemination. The Open Source community also behaves like a VO. Volunteers develop software tools that are shared and maintained by this community.

The Working and Resource Groups did not endorse the VO model as a recommended institutional model for a Canadian national data archive. First, none of the findings from the Sub-committee's research shed any light on the VO model. Secondly, the Groups' debate about this model raised some serious concerns. There was however support to have the recommended institution perform research into potential uses of the VO model for some of its functions. This investigation would be conducted as part of the research and development agenda of this new Canadian institution.

The points below summarize the advantages and disadvantages of the VO model discussed by the Groups. This information documents the completeness of the models considered by the Consultation.

1. What are the disadvantages of a Virtual Organization Model?

1.1 Quality Control. A volunteer organization complicates ensuring adequate quality control over the operations essential to the service. This is particularly true of very complex tasks. The greater the complexity, the more difficult to ensure quality control in a volunteer setting.

1.2 Comprehensive Service. A volunteer organization does well those things in which the volunteers are interested and from which they benefit. Therefore, such an organization is unlikely to provide a comprehensive service that performs not only the interesting tasks, but also the tedious ones. One might envision the results of turning over the preservation of the Census to genealogists. Pockets of material might be well preserved while other areas where volunteers do not have as great an interest would likely disappear through neglect.

---

[5] For a discussion of virtual organizations, see "The Anatomy of the Grid: enabling scalable virtual organizations," Ian Foster, Carl Kesselman, and Steven Tuecke, **The International Journal of High Performance Computing Applications**, vol. 15(3), 2001.

1.3 Level of Expertise. Expertise in a volunteer organization is likely to be diffuse and consequently, remain at an amateur level. This could be particularly detrimental regarding tasks that require specialized training for which the skills are not widely held.

1.4 Accountability and Responsibility. A volunteer organization is not answerable to others and difficult to manage because volunteers cannot be expected to do tasks in which they are not interested. Typically those who do the volunteer work have disproportionate control over what is done precisely because they choose to do only those things that interest them. As a result, setting priorities and sustaining strategic directions is more difficult with volunteers. Furthermore, volunteer organizations run the risk of becoming insular and isolationist.

1.5 Permanence. A model dependent on volunteers is only as strong as the community donating its labour. If the volunteer community is built around the personality of one individual, the community runs the risk of expiring upon the retirement, death, or career change of this person.

1.6 Security and Confidentiality. A volunteer organization will have difficulty addressing security and confidentiality concerns because it depends on a large number of relatively unsupervised volunteers who may not have the professional obligations or credentials to entrust security and confidentiality.

2. What are the advantages of the Volunteer Organization Model?

2.1 Community Involvement. A volunteer community can involve the larger community so they understand the need for data preservation and the value of professional preservation. A volunteer organization has the advantage of involving the very people whose political support are needed.

2.2 Using Underutilized Resources. A volunteer organization can leverage underutilized resources, especially processing and storage resources available over the network. For example, there is a significant number of PCs on the net that are idle all day. An appropriately designed screen saver or peer-to-peer tool could be distributed to volunteers that would exploit the excess processing and storage capacity on most networked personal computers.

2.3 Saving Money. The volunteer organization has the virtue of being cheap. Anything done by volunteers is something that does not have to be paid for if done properly. That said, it should be recognized that there are costs associated with setting up a volunteer organization that would not be otherwise incurred.

2.4 Cultivating Donors. A major issue in the success of a research data archive is getting the community to contribute to the archive. A volunteer organization is one way to cultivate educated donors. Those who are participating in a project are more likely

to donate appropriately to it and spread the word that donations are appreciated. The point is to find ways that members of the community can contribute. For example, there will need to be the development of migration or emulation tools in order to maintain access to research data. These tools could be developed following the Open Source model, which has proven successful elsewhere. By definition, research migration tools need to be open to ensure that users of research data can evaluate the processes that affect their data.

2.5 Extending Service. A volunteer organization could deal with donations that are outside the mandate or capability of a central archive. Should grey areas be encountered with data that for some reason fall outside the definition of research data, a volunteer community may be able to assist with such problematic content. For example, donations from industry that are not research data or donations that are arguably publications could be handled gracefully without alienating the donor. As such, the volunteer community would provide a companion organization to the main archive. By taking on well-defined tasks outside the core mandate of the main archive, the volunteer community frees the main archive to concentrate on its core mandate. A volunteer organization could be also provide coordinated services with other organizations that are beyond the research data mandate.

There was a sense that certain aspects of a VO could be incorporated with other organizational models to complement and enrich a national data archive. This would entail a balance between volunteers working in a supportive community with the staff of a national data archive.

Appendix D

Outline of Institutional Options for a National Research Data Archive

This appendix outlines various options for institutional options for a new National Research Data Network.

The models incorporate the thinking of the Working Group members, the input from the UK and Norwegian Data Archive directors, information gathered from a variety of stakeholders throughout the consultation process, and draws on advice received from the Privy Council Office and the Treasury Board.

In general, existing data archives around the world do not follow any specific model. All were shaped by the particular national context in which they were created. What they do have in common, however, is an underlying objective of ensuring that information gathered through the research process, or gathered for research purposes, should be effectively preserved and managed so that it is available to researchers for the purposes of creating new knowledge.

The models outlined here take into consideration this underlying objective as well as the particular context of the Canadian research community, the nature of Canadian research support mechanisms, and institutional structures and public governance systems.

1) New Division of the National Archives of Canada

Description:

- New division created within the existing National Archives operational, management and accountability structures;
- **Requirements:** re-formulation of the NA mandate; a-base increase to the NA budget; new personnel with expertise to operate a research-oriented data archive; facilities, administrative support and technology infrastructure;

Strengths:

- Takes advantage of existing and well-established operational, management and accountability structures;
- Takes advantage of existing archival culture and first-rate archival infrastructure;
- Does not require legislation or a change in the structures of the federal government;
- New division would be able to build on existing support services (e.g. human resources, accounting and payroll, IT support, etc.).

Weaknesses:

- Requires an expansion and re-formulation of the existing NA mandate;
- Requires a new type of service orientation, tailored to the needs of the research community rather than the general public;
- Single, centralised location under the direct control of the federal government, and, therefore, regarded as part of the federal bureaucracy;

- Does not have an existing, well-established relationship with the broad research community;
- Does not have a direct, on-going link with university data services;
- Currently, does not have the necessary space for a new division in its Ottawa facilities;
- Funding subject to changes in National Archives and overall federal budget priorities.

2) Special Operating Agency, attached to the National Archives of Canada (e.g. Passport Office)

Description:

- From *Framework for Alternative Program Delivery*: "This is an operational unit of a department that the Treasury Board has designated as an SOA. It operates within the departmental legislative framework and Treasury Board policies, and is accountable to the deputy head. A framework document and business plan establish accountability. These agencies promote a more businesslike approach within the departmental context. They have tailored authorities and flexibility delegated from the department and the Treasury Board".
- **Requirements:** authorisation by the Treasury Board and support from the National Archivist; re-formulation of the NA mandate; A-base increase to the NA budget; new personnel with expertise to operate a research-oriented data archive; facilities, administrative support and technology infrastructure.

Strengths:

- Has operational flexibility and a degree of independence within the National Archives departmental structure (accountable directly to the National Archivist);
- Takes advantage of existing and well-established operational structures;
- Takes advantage of existing archival culture and first-rate archival infrastructure;
- Does not require legislation or a change in the structures of the federal government;
- New division would be able to use existing support services (e.g. human resources, accounting and payroll, IT support, etc.).

Weaknesses:

- Requires an expansion and re-formulation of the existing NA mandate;
- Single, centralised location under the direct control of the federal government, and, therefore, regarded as part of the federal bureaucracy;
- Would have to establish a relationship with the broad research community;
- Would have to develop direct, on-going links with universities;
- Currently, does not have the necessary space for a new division in its Ottawa facilities;
- Funding subject to changes in National Archives and overall federal budget priorities.

3) Separate Statutory Agency (e.g. Canadian Space Agency, Statistics Canada)

Description:

- From *Framework for Alternative Program Delivery*: "Separate agencies are components of the Public Service…They operate under constituent legislation and are responsible to a minister and deputy head. These organisations operate within the policy framework set out by the minister."
- **Requirements:** Act of Parliament; annual vote in Parliament, within the budget process, for operating funds; facilities, staff, administrative support and technology infrastructure;
- The agency also requires a home department, probably Industry Canada, and a minister willing to promote its establishment.

Strengths:
- Would become a permanent institution, secured by legislation;
- Could be considered as an element in the current Innovation Agenda, focused on research, capacity building, stewardship and international competitiveness;
- Clear lines of authority and accountability within the federal government structures;
- Would have a ministerial champion, and if under Industry Canada, a key minister within Cabinet;
- Given its legislated status, the agency would most likely be able to secure a substantial budget, right from start-up;
- Both infrastructure and operational funding would come from a single source;
- Like StatsCan, the agency would have the potential, and the means, to develop a reputation as a "trusted system";
- Would have official national representation status in the international arena.

Weaknesses:
- Would require legislation, and therefore, a ministerial champion in Parliament;
- Could be a complex and lengthy process, one that is subject to the legislative agenda, reactions by the opposition in Parliament, government priorities, etc.
- May not be seen as a project large enough, or significant enough, for legislation;
- Would be seen by academics as a federal agency, and would have to work to establish a relationship with the research community;
- Creates a somewhat centralized facility, but one that could establish regional offices at universities.

4) Departmental Corporation (e.g. National Research Council)

Description:
- From *Framework for Alternative Program Delivery*: "This is a corporation established by an Act of Parliament…It performs administrative, research, supervisory or regulatory functions. It is subject to the same administrative requirements and controls as departments, but it has increased decision-making independence from government."
- **Requirements:** Act of Parliament; annual vote in Parliament, within the budget process, for operating funds; facilities, staff, administrative support and technology infrastructure;

- The agency also requires a home department, probably Industry Canada, and a minister willing to promote its establishment.

Strengths:
- Essentially the same as for a Separate Statutory Agency;
- In addition, would have the same "arm's length" relationship as the research councils, allowing for considerable independence of action;
- Arm's length status would allow for a closer relationship with research community than otherwise;
- Would have official national representation status in the international arena.

Weaknesses:
- Essentially the same as for a Statutory Agency, except the scale of the agency may be more appropriate for a Departmental Corporation;
- Slightly less secure long-term status than a Statutory Agency, but would still be considered a permanent body of the federal government (as permanent as, say, SSHRC);
- Creates a centralized facility, but one that could establish regional offices at universities.

5) Public Partnership (e.g. Canadian Centre for Justice Statistics)

Description:
- From *Framework for Alternative Program Delivery*: "This is a relationship formed when the federal government and other levels of government or governmental organizations agree to work co-operatively toward shared or compatible objectives. The partnership is based on a formal agreement specifying its purpose and nature, and the terms and conditions governing it, such as financing, staffing and reporting. The specific instrument may be established pursuant to Governor-in-Council approval. This category includes joint enterprises, which are corporations jointly owned by the federal government with other levels of government."
- **Requirements:** Order-in-Council to establish the agency; champion within Cabinet to promote the initiative; agreement with the provinces on a wide range of issues; offices, facilities, staff, administrative support and technology infrastructure.

Strengths:
- Follows precedents for the creation of bodies similar to a data archiving agency;
- Order-in-Council establishes permanency almost as strong as legislation;
- Potentially broadens the scope of the agency to include a direct connection with provincial statistical agencies;
- Relatively secure funding, but from multiple sources.

Weaknesses:
- Exceptionally complicated processes to follow;
- Requires negotiating agreements between numerous players, may prove lengthy;

- Would require a strong and committed champion within Cabinet;
- Subject to the policy and budget priorities of up to 14 different governments;
- Does not have the arm's length status of a Departmental Corporation or independence of a Statutory Agency.

6) University-based Centre (e.g. TRIUMF, Univ. of British Columbia)

Description:

- A research support centre, based on a university campus, operated as a joint venture by associated members (universities), with a financial contribution from SSHRC;
- Application would be made to the Canada Foundation for Innovation for infrastructure start-up costs;
- Operational funding would flow through SSHRC to the centre. The centre would be accountable to the SSHRC President and Board of Directors;
- Each associated member would constitute a node in the centre's network, and would have a seat on the Board of Management;
- Every Canadian university with a research data service would be eligible to become an associated member, and would both benefit from, and be responsible for, the operations of the centre;
- The Board of Management would be responsible for overseeing the operations of the centre, and setting general policy directions;
- A Director would be responsible for management of the centre, day-to-day operations, human resources, financial transactions, and reporting;
- The Director would be assisted by Advisory Committees composed of both data users and archival experts;
- The location of the centre could be selected through a SSHRC-managed peer-review competitive process amongst the associated members.
- **Requirements**: a supplement to the SSHRC A-base budget dedicated to the centre (flow through funds); application to CFI for infrastructure funding; a site selection process; creation of governance and organizational structures; offices, facilities, staff, administrative support and technology infrastructure.

Strengths:

- Follows an established model for locating a federally funded research facility on a university campus;
- Direct, immediate, on-site connection with the research community;
- Ownership, operations and policies would be in the hands of the university associated members;
- Builds on the existing data services, expertise and technology infrastructure within universities;
- Could use a hybrid centralized/de-centralized system, where the centre takes care of preservation and data set processing and the associated members act as local facilities for access to data, deposit of data, on-site advice, and educational activities;
- Associated members could include non-university bodies such as provincial archives;

- Scope of the agency is scalable, could include NSERC and CIHR areas of science, with associated funding agreements with these agencies;
- Digital archival research activities would take advantage of proximity to university based information science researchers.

Weaknesses:

- Lacks the assured long-term stability that comes with legislation;
- No directly equivalent institutional models in the social sciences;
- Funding would come from two, and possible more, sources – CFI and SSHRC
- CFI funding would require 60% matching funds (although it may be possible to draw on provincial matching funds programs);
- Governance structure would involve many partners and be relatively complex;
- Not all associated members would have the same level of resources to provide local services;
- May not have official national representation status in the international arena.

Appendix E
Topical Bibliography

Functions

"Review: Data Historica: A Guide to Electronic Resources for Historians Compiled by the Netherlands Historical Data Archive." *Historical Social Research* [Germany] 23(4), 1998.

Anderson, B. "Secondary Analysis and Gateway Libraries." *Behavioral and Social Sciences Librarian*. 17:2. 109-110 (1999).

Austin, Erik W. and Dunn Christopher S. "Protecting Confidentiality in Archival Data Resources." *IASSIST Newsletter*. 22:2. 16-24 (1998).

Bennett, J. C. "Archiving and Data Management." *Annual Conference – 1996 Sep: York:* National Preservation Office; 1997.

Bonatti, P., et al. "An Access Control Model for Data Archives." *International Conference; 16th – 2001 June: Paris:* International Federation for Information Processing. Boston; London; Kluwer Academic; c2001.

Buck, Michael. "Problems in Cataloguing Computer Files." *International Cataloguing and Bibliographic Control* v18 (1989).

Carley, Michael, and Josefina J Card. "The Social Science Electronic Data Library: Serving the Needs of Data Librarians and Users." *IASSIST-(International-Association-for-Social-Science-Information; 24: No 3 Fall 2000* (2000).

Church, J. "Getting Our Data Used: A National Statistics Office Perspective." *NATO Advanced Research Workshop – 1997 Aug : Colchester:* IOS Press; Ohmsha; 1998.

Consultative Committee for Space Data Systems. "Reference Model for an Open Archival Information System (OAIS): Recommendation Concerning Space Data Systems Standards." White Book CCSDS 650.0-W-4.0, September 17, 1998. http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

Corti, Louise, Janet Foster, and Paul Thompson. "The Need for a Qualitative Data Archival Policy." *Qualitative Health Research* 6, 1, Feb (1996).

Crothers, A. L. "Institute for Research in Social Science Louis Harris Data Archive (Social and Behavioral Sciences) (Brief Article)." *CHOICE: Current Reviews for Academic Libraries* v38 Special Issue p. 131(2). (2001).

Dorobek, Christopher J. "Project Aims to Modernize Data Preservation Process." *Government Computer News* v18 (August 30, 1999).

Duffy, C. "An Introduction to the Performing Arts Data Service." *Literary & Linguistic Computing : Journal of the Association for Literary and Linguistic Computing:* 1997.

Economic and Social Research Council. “ESRC Green Paper on Data Policy and Data Archiving” October 2000. http://www.ilrt.bris.ac.uk/ubris/esrc/p3.html

Egret, D., and F. Genova. "The Role of Data Centres in the Era of Electronic Publishing." *Astrophysics and Space Science Library. 224, (1997)* D. Reidel Pub. Co.

Finke, Roger; et al. "Democratizing Access to Data: the American Religion Data Archive" *IASSIST-(International-Association-for-Social-Science-Information); 23: No 1 Spring 1999*.

Fraser, Michael. "From Concordances to Subject Portals: Supporting the Text-Centred Humanities Community." *Computers and the Humanities v 34 No3 Aug 2000* p. (265-1978).

Gabrynowicz, Joanne Irene. "The Work of the US National Satellite Land Remote Sensing Data Archive Advisory Committee." *Space Policy* (Feb2001).

Green, Ann. Dionne JoAnn. Dennis Martin. "Preserving the Whole: A Two-Track Approach to Rescuing Social Science Data and Metadata." Washington, DC: Digital Library Federation, 1999.

Jesske Muller, Birgit, and Christoph Reichert. "Social Research on Developing Countries: Reasons for and Aims of a Development Studies Data Collection." *Historical Social Research / Historische Sozialforschung* 15-2 (1990).

Jones, Margaret. "A Workbook for the Preservation Management of Digital Materials." *New Review of Academic Librarianship* v 6 (2000).

Jones, Roger. "Australia's Electronic Heritage: Comments on the ACLIS Task Force Recommendations: Augmented Title: Perspectives on the Towards Federation 2001 Conference, Canberra, March 1992." *Australian Library Review v10 Nov 1993* p. (447-1956).

Kinzy S. "Data Warehousing Moves GIS Beyond Maps." *GIS-World* 1998; pp. 1958-1961.

Kuula, Arja. "Making Qualitative Research Material Reusable: Case in Finland." *IASSIST-(International-Association-for-Social-Science-Information); 24: No 2 Summer 2000*.

Leadbeter, Deana, and Lorna Guinness. "Development of a Health Data Archive for Bangladesh: an Example of a Cost Effective and Sustainable Approach to Information Sharing in a Developing Country." *IASSIST-(International-Association-for-Social-Science-Information); 23: No 2 Summer 1999*.

Lesaoana, Maseka A. "Data Archiving in Africa: The South African Experience." *IASSIST Newsletter. 21:1. 4-7.* (1997).

Levada, Y. "The Russian Data Infrastructure." *NATO Advanced Research Workshop – 1997 Aug : Colchester:* IOS Press; Ohmsha; 1998.

Lievesley, D. "Access to European Data." *Conference – 1995 Dec: Dublin:* Dublin; Institute of Public Administration, 1996.

—. "Making Data Available - the Experience of the ESRC Data Archive." *Seminar on the Future History of Our Landscape – 1992 Oct: London:* Royal Society of London; 1994.

—. "The Role and Future Development of the Data Archive." *Conference – 1995 Dec: Dublin:* Dublin; Institute of Public Administration; 1996.

—. "The Value of Data Dissemination to Both Data Users and Producers." *NATO Advanced Research Workshop – 1997 Aug: Colchester:* IOS Press; Ohmsha; 1998.

Lievesley, D, and I Masser. "Socio-Economic Data: Making the Pieces Fit." *GIS-Europe* 1994.

McCook, Kathleen-de-la-Pena. "Social Scientific Information Needs for Numeric Data: the Evolution of the International Data Archive Infrastructure." *Collection Management v 9 Spring 1987. P. 1-53*.

Mochmann, E. "European Co-Operation in Social Science Data Dissemination." *NATO Advanced Research Workshop – 1997 Aug: Colchester:* IOS Press; Ohmsha; 1998.

National Research Council. "Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources". Washington, DC: National Academy Press, 1995.

Nielsen, Per. "Merging Cultures: Danish Integration of Academic Data Service into Traditional Archive System." *IASSIST Newsletter. 19:2. 27-35* (1995).

Owen, Catherine, et al. "Meeting the Challenge of Film Research in the Electronic Age" *D-Lib-Magazine; 6: No 3 Mar 2000*.

Rasmussen, Karsten Boye. "Documentation - What We Have and What We Want: Report of an Enquete of Data Archives and Their Staff." *IASSIST Newsletter. 19:1. 22-35* (1995).

Richards, Julian D. "Preservation and Re-Use of Digital Data: the Role of the Archaeology Data Service." *Antiquity v 71 Dec 1997* pp. 1057-1997.

Ryssevik, Jostein, and Simon Musgrave. "The Social Science Dream Machine: Resource Discovery, Analysis, and Delivery on the Web." *Social Science Computer Review* Vol. 19 Issue 2 (May2001).

Scheuch, Erwin K. "From a Data Archive to an Infrastructure for the Social Sciences." International Social Science Journal [Great Britain] 1990 42(1): 93-111.

Schurer, K. "Twenty-Five Years of Preserving and Disseminating Data: the Experience of the ESRC Data Archive. An Abstract." *7th Anglo-Nordic Seminar – 1994 Oct : Windsor:* Nordic Council for Scientific Information and Research Libraries. The British Library; Research and Development Department. NORDINFO; 1995.

Singh, Jagtar, et al. "Accessing Indian Numeric and Statistical Data: a Critical Study of the Suprastructure and Infrastructure in India." *IASSIST-(International-Association-for-Social-Science-Information); 24: No 3 Fall 2000*.

Tanenbaum Eric, and Mochmann Ekkehard. "Towards a European Database for Comparative Social Research." *Historical-Social-Research / Historische-Sozialforschung; 1996, 21, 2(78), 118-122*.

Treadwell, W. "DDI, the Data Documentation Initiative: An Introduction to the Standard and Its Role in Social Science Data Access." *2000 Jul: Chicago, IL:* American Library Association. Scarecrow Press; 2002.

Vries, Repke-Eduard-de. "ILSES: How Library and Data Archive Meet in Active Support of Research in the Social Sciences." *INSPEL v 31 No4 1997. P. 234-41*.

Whelan, B. J. "An Irish Social Science Data Archive: Prospect and Problems." *Conference – 1995 Dec: Dublin:* Dublin; Institute of Public Administration; 1996.

Mandate / Principles / Ethical Issues

"Access to Data From Medical Research." *New England Journal of Medicine* (12/25/86).

"Case for Free Access to Fundamental Data." *Lancet (North American Edition) v 353 No9166 May 22 1999. p. 1721*

"Conservatives Challenge Science Community on Data Access." *Issues in Science and Technology v 15 No4 Summer 1999*.

"Data Access and Privacy Concerns." *Monthly Labor Review v 117 Jan 1994*.

"Enhancing Access to Primary Sources Internationally: A Joint RLG-CURL Symposium Held at University College London" June 1997. http://www.rlg.org/psprog.html

"Health Privacy Overview: Background and Current Law." *Congressional Digest v 79 No8/9 Aug/Sept 2000. p. 194-196* .

History Data Service (HDS). "Collections Development Policy." http://hds.essex.ac.uk/collpol.asp

ICPSR Collections Policy - http://www.icpsr.umich.edu/ORG/Policies/colldev.html

"Law on Access to Research Data Pleases Business, Alarms Science." *Pediatrics* (Oct99 Part 1 of 2).

"Neuroscientists Debate Data Sharing . Virtual Structure for Canada's Biomedical Research." *Nature:* 2000.

"News - Scientists Await Impact of Public Data Sharing" *The Scientist:* 2000.

"Open Exchange of Scientific Data Should Be Protected." *Public Health Reports v 112 July/Aug 1997. p. 268* .

"Preserving and Sharing Statistical Material." Working Group on the Preservation and Sharing of Statistical Material: Information for Data Producer/UK Data Archive (UKDA). http://www.data-archive.ac.uk/home/PreservingSharing.pdf

"Protecting Privacy in an Information Sharing Environment." *Fredericton: The Task Force, 1994*.

"Reinforcing Access to Research Data." *Nature* (1/18/96).

Adam, John. "Research in Electronic Age Should Be Open." *IEEE Spectrum* (Jun97).

Alberts, Bruce, and Kenneth Shine. "Scientists and the Integrity of Research." *Science v 266 Dec 9 1994*.

Aldhous, Peter. "Prospect of Data Sharing Gives Brain Mappers a Headache." *Nature v 406 No6795 Aug 3 2000. p. 445-446*.

Allen, R S. "An Overview of the Federal Geographic Data Committee, National Spatial Data Infrastructure, National Geospatial Data Clearinghouse, and the Digital Geospatial Metadata Standard: What Will It Mean for Tomorrow's Libraries?" *Bulletin, Special Libraries Association, Geography and Map Division* 180 (1995).

Altman, Micha, et al. "A Digital Library for the Dissemination and Replication of Quantitative Social Science Research: The Virtual Data Center." *Social Science Computer Review 19, No. 4 (2001): 458-470*.

Anderson, R. "Undermining Data Privacy in Health Information." *BMJ. 322(7284):442-3, 2001 Feb 24.* .

Annas, George J. "Rules for Research on Human Genetic Variation- Lessons From Iceland." *The New England Journal of Medicine v 342 No24 June 15 2000* p. 1830-2000.

Applegate, David. "AGI Workshop Examines Geoscience Data Preservation." *Geotimes* v 41 (1996).

Archaeology Data Service (ADS). "ADS Collections Policy: 3rd edition."
http://ads.ahds.ac.uk/project/collpol3.html

Austin, Erik W. and Dunn Christopher S. "Protecting Confidentiality in Archival Data Resources." *IASSIST Newsletter. 22:2. 16-24* (1998).

Barber, B., P. Skerman, and R. Saldanha. "The Definition of Data Privacy for Europe." *Conference – 1997 Mar: Harrogate:* BJHC Limited; 1997.

Baron, James N. "Data Sharing as a Public Good." *American Sociological Review1988* 53, 1, Feb, vi-viii.

Beedham, Hilary. "The Royal Statistical Society Working Group on Archiving Data." *IASSIST-(International-Association-for-Social-Science-Information); 23: No 3 Fall 1999* (1999).

Blume, Peter ed. "Special Feature: The Citizen in the Information Society; Legal Issues From the Perspective of the Citizen." *Journal of Information, Law and Technology (JILT); Issue 1 F 27 1998*.

Campbell, Paulette Walker. "Controversial Proposal on Public Access to Research Data Draws 10,000 Comments." *Chronicle of Higher Education* (4/16/99).

Carew, R. "Institutional Arrangements and Public Agricultural Research in Canada." *Rev-Agric-Econ* 23 (2001).

Carter, J, and -R. "Perspectives on Sharing Data in Geographic Information Systems." *Photogrammetric-Engineering-and-Remote-Sensing. 1992. 58(11), Pp 1557-1560.*

Ceci, Stephen J. "Scientists' Attitudes Toward Data Sharing." *Science Technology & Human Values. V13 N1-2 P45-52 Win-Spr 1988* .

Cooley, William W. "Confidentiality of Education Data and Data Access." Pennsylvania Educational Policy Studies Number 7.

Corti, Louise, Janet Foster, and Paul Thompson. "The Need for a Qualitative Data Archival Policy." *Qualitative Health Research* 6, 1, Feb (1996).

Craig, W. J. "Why We Can't Share Data: Institutional Inertia." *Meeting Entitled "Institutions Sharing Geographic Information" – 1992 Feb:* National Center for Geographic Information and Analysis. New Brunswick; Center for Urban Policy Research; 1995.

Crissman, James W, James S Bus, and Ron R Miller. "Toxicology: Judge Data or Dollars?" *Environmental Health Perspectives v 107 No10 Oct 1999*.

Curry, M. R. "Data Protection and Intellectual Property: Information Systems and the Americanization of the New Europe." *Environment-and-Planning-A. 1996. 28/5, 891-908.*

da Silva, Wilson. "Europe Gets Tough on Data Privacy." *New Scientist v 154 May 10 1997. P. 10.*

Dalton, Rex. "Young, Worldly and Unhelpful All Miss Out on Data Sharing." *Nature v 404 No6773 Mar 2 2000*.

Dangel, E. T. 3rd, and B Mishkin. "Collaboration and Data Sharing: Continued." *Science. 1996 Jan 12;271(5246):129-30*.

Data and Information Systems Panel. "Data Policy and Barriers to Data Access in Canada: Issues for Global Change Research: A Discussion Paper" Canadian Global Change Program http://www.globalcentres.org/cgcp/english/html_documents/publications/data/toc.htm

Dove, Alan. "Research Data Access Bill Issued." *Nature Medicine v 6 No1 Jan 2000*.

—. "US Public May Gain Access to Research Data." *Nature Medicine v 5 No1 Jan 1999. P. 8*.

Downing, John D H. "Computers for Political Change: PeaceNet and Public Data Access." Journal of Communication. v39 n3 p154-62 Sum 1989.

Dreier, T. "Development and Perspectives of Copyright: From Gutenberg to Data Highways'." *International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies – 1995 Aug : Vienna:* Munich; Vienna; Oldenbourg; Osterreichische Computer Gesellschaft; 1995.

Dresser, Rebecca. "Accountability in Science and Government: Is Access the Answer?" *The Hastings Center Report v 30 No3 May/June 2000*.

Dueker K, et al. "A Geographic Information System Framework for Transportation Data Sharing." *Transportation-Research-Part-C:-Emerging-Technologies* 2000; (1936).

Duncan, Joseph W. "Confidentiality and the Future of the U.S. Statistical System." *American Statistician* Vol. 30 Issue 1 (Feb76).

Dunn, Christopher S. and Erik W. Austin. "Protecting Confidentiality in Archival Data Resources.". ICPSR Bulletin, Fall 1998. http://www.icpsr.umich.edu/ORG/Publications/Bulletin/Fall98/article.html

Eberle, Carl Eugen. "Legislation to Protect the Confidentiality of Personal Data: Implications for Social Science Research; Implikationen Des Datenschutzes Fur Die Empirische Sozialforschung." *Zeitschrift-Fur-Soziologie; 1981, 10, 2, Apr, 196-211*.

Eliot Marshall. "Genome Researchers Take the Pledge: Data Sharing." *Science* 1996 v272 n5261 p477(2).

Elliott, P, et al. "Intersalt Data - Science Demands Data Sharing." *British Medical Journal* 315 (1997).

Els, C., T. Verschoor, and H. Oosthuizen. "The Protection of Personal and Medical Data–a Call for Confidentiality." *South African Medical Journal. 85(8):773-5, 1995 Aug* .

Estabrooks, C. A. Romyn DM. "Data Sharing in Nursing Research: Advantages and Challenges." *Can J Nurs Res. 1995 Spring;27(1):77-88.*

Farrell, Rita. "Who Owns the Knowledge?" *UNESCO Sources* (May99 Issue 112).

Fellegi, I. P. "On the Question of Statistical Confidentiality." *Journal of the American Statistical Association* (Mar72).

Fienberg, S. E. "Statistical Perspectives on Confidentiality and Data Access in Public Health." *Statistics in Medicine. 20(9-10):1347-56, 2001 May 15-30* .

Frankel, M. S. "Public Access to Data." *Science. 1999 Feb 19;283(5405):1114*.

Gibson, J. L. "Cautious Reflections on a Data-Archiving Policy for Political Science." *Political Science & Politics* 28 (1995).

Goodwin, Candice. "Personal Privacy v Public Health." *New Scientist v 140 Dec 11 1993*.

Gordon, W. "Commentary on Economical and Ethical Reasons for Protecting Data." *Fordham Annual Conference; 8th – 2000:* Juris Publishing; 2001.

Greenstein, Daniel I, and Sarah Porter. "Scholars' Information Needs in a Digital Age: Executive Summary." *New Review of Academic Librarianship v 4 1998*.

Hammersley, Martyn. "Qualitative Data Archiving: Some Reflections on Its Prospects and Problems." *Sociology* Feb 1997 v31 n1 p131(12).

Haseltine, Florence P. "Unexpected Trouble With the Freedom of Information (FOIA): An Open Letter to the Public." *Journal of Women's Health & Gender-Based Medicine* (Apr99).

Hilgartner, Stephen. Brandt-Rauf Sherry I. "Data Access, Ownership, and Control: Toward Empirical Studies of Access Practices." *Knowledge: Creation, Diffusion, Utilization. V15 N4 P355-72 Jun 1994* .

Howell Robert G. "Database Protection and Canadian Laws, State of Law As of June 15, 1998.". Ottawa: Industry Canada ; Hull, PQ : Canadian Heritage, 1998.

Imwalle, Christine. "Data Warehousing Enables Sharing." *American City and County v 113 No8 July 1998*.

Isaacs, Lindsay. "Shared Responsibility, Shared Reward." *American City and County v 115 No15 Nov 2000. P. 44-52* .

Jeffords, James. "Confidentiality of Medical Information: Protecting Privacy in an Electronic Age." *Professional Psychology, Research and Practice v 30 No2 Apr 1999. P. 115-116*.

Johnson, David J. and Michel E. Sabourin. "Universally Accessible Databases in the Advancement of Knowledge From Psychological Research ." *International Journal of Psychology* (Jun2001).

Jones, Roger. "Comment on 'The Need for the Preservation of Australian Created Electronic Information: a Position Paper'." *Australian Academic and Research Libraries v 24 Mar 1993* pp. 1947-1993.

Kaiser, Jocelyn. "Researchers and Lawmakers Clash Over Access to Data." *Science v 277 July 25 1997* p. 467.

Karr, Alan F., and Jaeyong Lee. "Disseminating Information but Protecting Confidentiality." *Computer* (Feb2001).

Kastner, T. "On the Need for Policy Requiring Data-Sharing Among Researchers Publishing in AAMR Journals: Critique of Conroy and Adler (1998)." *Ment Retard. 2000 Dec;38(6):519-29.*

Key, Sandra W., et al. "Innovations in Research Threatened by Proposals to Restrict Exchange of Data." *AIDS Weekly Plus* (01/05/98).

Kiernan, Vincent. "Debate Continues on Cost of Global Access to Scientific Data." *Laser Focus World* (Jul97).

Lawler, Andrew. "Treaty Draft Raises Scientific Hackles." *Science v 274 Oct 25 1996* p. (494).

—. "U.S., Europe Clash Over Plan to Set Policy on Data Access." *Science v 268 Apr 28 1995* p. (493).

Litman, J. "Information Privacy/Information Property." *Stanford Law Review* .

Longley, P. "Intellectual Property Rights and Digital Data." *Environment and Planning*.

Love, Denise E., Luis M.C. Paita, and William S. Custer. "Data Sharing and Dissemination Strategies for Fostering Competition in Health Care." *Health Services Research* 36 issue 1 p. 277. (2001).

Macilwain, Colin. "Business Lobby Set to Take EPA to Court Over Data Access." *Nature v 403 No6767 Jan 20 2000* p. 236.

—. "Researchers Resist Copyright Laws That Could Endanger Data Access." *Nature* (10/24/96).

—. "Scientists Fight for Right to Withhold Data." *Nature* Vol 397 No 6719 (1999).

—. "US University Fears Allayed Over Access to Research Data." *Nature* (10/21/99).

Marshall, Eliot. "Agencies, Journals Set Some Rules." *Science v 248 May 25 1990. P. 954*.

—. "Data Sharing: a Declining Ethic?" *Science v 248 May 25 1990. P. 952+*.

—. "Epidemiologists Wary of Opening Up Their Data." *Science* v 290 no 5489 (2000).

Maurer, S. M. Hugenholtz PB, Onsrud HJ. "Intellectual Property. Europe's Database Experiment." *Science. 2001 Oct 26;294(5543):789-90.*.

Maurer, SM. "Coping With Change: Intellectual Property Rights, New Legislation, and the Human Mutation Database Initiative." *Human Mutation*. 15:22-29, 2000

Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. "The Data Are Out There, or Are They?: Implications for Archiving and Revisiting Qualitative Data." *Sociology* Nov 1998 v32 i4 p733(1).

Mellors, Colin, and David Pollitt. "Policing the Communications Revolution: a Case-Study of Data Protection Legislation." *West-European-Politics; 9:195-211 O 1986*.

Miller, Heather G, and Wendy H Baldwin. "A Terse Amendment Produces Broad Change in Data Access." *American Journal of Public Health v 91 No5 May 2001* p. (824-2001).

Mills, M. E. "Data Privacy and Confidentiality in the Public Arena." *Proceedings / AMIA Annual Fall Symposium. :42-5, 1997* .

Mishkin, B. "Urgently Needed: Policies on Access to Data by Erstwhile Collaborators." *Science. 1995 Nov 10;270(5238):927-8.*.

Nobel, J. "Data Confidentiality and Data Access - Practical and Legal Issues in the Netherlands." *International Seminar – 1994 Nov : Luxembourg:* Eurostat.; 1995.

Normile, Dennis. "Asian Network Seeks Data Sharing." *Science v 267 Mar 31 1995* p. 1902.

Ochert, Ayala. "US Fury Over Data Access Legislation." *Times Higher Education Supplement* Issue 1387 (06/04/99).

Olsen, Florence. "Publishers and Users of Data Bases Disagree Over Intellectual-Property Legislation." *Chronicle of Higher Education* (7/9/99).

Peterson, M.J.. "Community and Individual Stakes in the Collection, Analysis, and Availability of Data. (Response to Articles by Gary King and Paul S. Herrnson in This Issue, P.444 and P.452, Respectively)." *PS: Political Science & Politics* Sept 1995 v28 n3 p462(3).

Pickett, R., and D. Quackenbush. "Lock the Door and Throw Away the Key? . If We Cannot Provide Access to Data, Why Collect It?" *Annual Conference – 1995 Nov : New Orleans; LA:* CAUSE. CAUSE; 1996.

Plater, S. Seeley E, Dixon LA. "Two Routes to Privacy Protection: a Comparison of Health Information Legislation in Canada and the United States." *J Womens Health. 1998 Aug;7(6):665-72*.

Porter, J, et al. "Circumventing a Dilemma: Historical Approaches to Data Sharing in Ecological Research." *in: Environmental Information Management and Analysis. (Taylor & Francis), 1994, Pp 193-202*.

Porter, Sarah. "Reports From the Front: Six Perspectives on Scholar's Information Requirements in the Digital Age." *New Review of Academic Librarianship v 4* (1998).

Posey, D A, G Dutfield, and K Plenderleith. "Collaborative Research and Intellectual Property-Rights." *Biodiversity and Conservation*.

Ready, Tinker. "Will Research Be Restricted by New Data Privacy Rules." *Nature Medicine v 7 No2 Feb 2001*.

Reidpath, Daniel D, and Pascale A Allotey. "Data Sharing in Medical Research: an Empirical Investigation." *Bioethics* v 15 no2 (2001).

Rennolls, K. "Intersalt Data - Science Demands Data Sharing." *British Medical Journal* 315 (1997).

Robbin, Alice, and Heather Koball. "Seeking Explanation in Theory: Reflections on the Social Practices of Organizations That Distribute Public Use Microdata Files for Research Purposes." *Journal of the American Society for Information Science and Technology New York* 52 (2001).

Rockwell, R. C., and R. P. Abeles. "Sharing and Archiving Data Is Fundamental to Scientific Progress." *Journals of Gerontology. Series B, Psychological Sciences & Social Sciences. 53(1):S5-8, 1998 Jan*.

—. "Guest Editorial: Sharing and Archiving Data Is Fundamental to Scientific Progress." *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences v 53B No1 Jan 1998*.

Rowen, L, G K S Wong, and R P and others Lane. "INTELLECTUAL PROPERTY: Publication Rights in the Era of Open Data Release Policies." *Science:* 2000.

Rudder, Catherine-E. "APSA Responds to OMB's Draft Regulations." *PS, Political Science and Politics v 32 No2 June 1999. P. 188-189*.

Samuelson, P. "Copyright, Digital Data, and Fair Use in Digital Networked Environments." *Conference – 1994 May : Montreal; Canada:* Kluwer Law International; 1995.

Santoro, M. D., and S Gopalakrishnan. "Relationship Dynamics Between University Research Centers and Industrial Firms: Their Impact on Technology Transfer Activities." Chicago, Illinois: *Technology Transfer Society* v. 26 (1-2) (Jan 2001.).

Schafer, U. "Public Archives Between Data Access and Data Protection." *DLM-Forum; 2nd – 1999 Oct : Brussels:* Commission of the European Communities. Luxembourg; Office for Official Publications of the European Communities; 2000.

Schwartz, Ephraim, Heather Harreld, and Tom Sullivan. "Data Sharing's Call to Arms." *InfoWorld Framingham* 23 (2001).

Shaman, Susan M. Shapiro Daniel. "Data-Sharing Models." *New Directions for Institutional Research. N89 P29-40 Spr 1996* .

Shelby, R. "Accountability and Transparency: Public Access to Federally Funded Research Data." *Harvard Journal On Legislation*.

Sheppard, M. "Data Ownership and Copyright." *8th Annual Symposium on Geographic Information Systems in Forestry, Environmental and Natural Resources Management – 1993 Feb : Vancouver; Canada:* Polaris Conferences; 1994.

Sieber, Joan E ed. "Sharing Social Science Data: Advantages and Challenges." Sage Publications, 1991.

Siegfried, Susan L. "The Policy Landscape." *The Art Bulletin v 79 June 1997* p. (209-1997).

Simpson, Roy L. "How Can We Keep Private Data Private?" *Nursing Management* (May 2001).

Singer, E., and S. Presser. "Public Attitudes Toward Data Sharing: Findings and Implications." *Annual Meeting – 1997 Aug : Anaheim; CA:* American Statistical Association; Section on Survey Research Methods. ASA; 1997.

Singer, Eleanor, Schaeffer Nora Cate, and Raghunathan Trivellore. "Public Attitudes Toward Data Sharing by Federal Agencies." *International Journal of Public Opinion Research* Fall 1997 v9 n3 p277(9).

Skinner, T. "Improving Access to Unidentified Data." *International Seminar – 1994 Nov : Luxembourg:* Eurostat; 1995.

Smaglik, Paul. "Publication Deal for Celera Sparks Row Over Data Access." *Nature v 408 No6814 Dec 14 2000* p. (759).

Stevens, A. R. "Collaboration and Data Sharing." *Science* 270(5243) (1995).

Stevens, Ann R. "Ownership and Retention of Data." NACUA Publication Series. Washington, DC: National Council of Univ. Research Administrators.

Sullum, Jacob. "Data on Demand." *Reason* (Feb99).

Townsend, Sean, Cressida Chappell and Oscar Struijve "Digitising History: A Guide to Creating Digital Resources from Historical Documents." HDS/AHDS. http://hds.essex.ac.uk/g2gp/digitising_history/index.asp

UK Data Archive "Collections Development Policy" http://www.data-archive.ac.uk/depositingData/collectionsPolicy.asp

Varian, Hal R. "The Information Economy." *Scientific American v 273 Sept 1995* p. (1995).

Wheeler, David L. "A Bitter Feud Over Authorship." *Chronicle of Higher Education. V41 N38 PA8,12 Jun 2 1995* .

Whitman, E. D., M. E. Frisse, and M. G. Kahn. "The Impact of Data Sharing on Data Quality." *19th Annual Symposium – 1995 Oct : New Orleans; LA:* American Medical Informatics Association. Hanley & Belfus; 1995.

Williamson, Alan R. "Access to Data and Intellectual Property." *ASM News* 8 (1998).

Young, Douglas L. "A Comment on the Role of Professional Journals in Facilitating Data Access." *Journal-of-Agricultural-Economics-Research; 43:45-7 Spring 1991*.

Organizational Models

"Access to Cultural Heritage Through an on-Line Multimedia Data Service: Application to the Archive Folders of France's General Inventory of Monuments and Art Treasures." *Conference – 1997 May : Barcelona; Spain:* Graphic Communications Association. Graphic Communications Association; [1997].

Humanities and Arts on the Information Highways: A Profile" Coalition for Networked Information. November 1997. http://www.cni.org/projects/humartiway/humartiway-rpt.part1.html

"Preserving Digital Information: Report of the Task Force on Archiving of Digital Information". Commision of Preserving Digital Information and the Research Libraries Group, Inc. May 1996. http://www.rlg.org/ArchTF/tfadi.index.htm

"The Royal Statistical Society Working Group on Archiving Data. AU: Beedham,-Hilary." *IASSIST-(International-Association-for-Social-Science-Information); 23: No 3 Fall 1999*.

"Trusted Digital Repositories: Attributed and Responsibilities" RLG / OCLC. May 2002 http://www.rlg.org/longterm/repositories.pdf

Bainbridge, W. S. "International Network for Integrated Social Science." *Social Science Computer Review* .

Brophy, Peter, et al. "Towards a National Agency for Resource Discovery Scoping Study" British Library Research and Innovation Report 58. 1997. http://www.ukoln.ac.uk/services/papers/bl/blri058/nard.html

Corti, Louise, Janet Foster, and Paul Thompson. "The Need for a Qualitative Data Archival Policy." *Qualitative Health Research* 6, 1, Feb (1996).

Eiteljorg, H. "The Archaeological Data Archive Project." *Conference – 1994 Mar : Glasgow:* Tempus Reparatum; 1995.

Fonseca F, et al. "Ontologies and Knowledge Sharing in Urban GIS." *Computers-Environment-and-Urban-Systems* 2000; (251-271).

Gatenby, Pam. "Digital Archiving – Developing Policy and Best Practice Guidelines at the National Library of Australia" International Council For Scientific and Technical Information. Jan 2000. http://www.icsti.org/2000workshop/gatenby.html

Garcia, A. "A Conceptual Framework for Clinical Data Warehouses." *A Medical Informatics Odyssey: Visions of the Future and Lessons From the Past:* Washington, DC: Hanley and Belfus; 2001.

Haberkorn-Susan-B, and Traugott-Michael-W. "National Archive of Computer-Readable Data on Aging." *Special-Collections. 1982; Vol. 1 (No. 3-4): P. 73-83*.

Harris, R. Olby N. "Archives for Earth Observation Data." *Space Policy* .

—. "Earth Observation Data Archiving in the USA and Europe." *Space Policy*.

Haynes, David, and David Steatfield. “A National Co-Ordinating Body for Digital Archiving?” *Ariadne.* 15. May 1998. http://www.ariadne.ac.uk/issue15/digital/

Hodge, Gail, and Bonnie C Carroll. “Digital Electonic Archiving: The State of the Art and the State of the Practice.” International Council for Scientific and Technical Information / CENDI. April 1999. http://www.icsti.org/99ga/digarch99_TOCP.pdf ; http://www.icsti.org/99ga/digarch99_ExecP.pdf ; http://www.icsti.org/99ga/digarch99_MainP.pdf.

Leadbeter, Deana, and Lorna Guinness. "Development of a Health Data Archive for Bangladesh: an Example of a Cost Effective and Sustainable Approach to Information Sharing in a Developing Country." *IASSIST-(International-Association-for-Social-Science-Information); 23: No 2 Summer 1999*.

Nielsen, Per. "Merging Cultures: Danish Integration of Academic Data Service into Traditional Archive System." *IASSIST Newsletter. 19:2. 27-35* (1995).

Pack, Thomas, and Jeff Pemberton. “A Harbinger of Change: The Cutting Edge Library at the Los Alamos National Laboratory.” *Online Magazine*, March 1999. http://www.onlineinc.com/articles/onlinemag/pack993.html

Porter, J, et al. "Circumventing a Dilemma: Historical Approaches to Data Sharing in Ecological Research." *in: Environmental Information Management and Analysis. (Taylor & Francis), 1994, Pp 193-202* .

Reed, Ken. Blunsdon Betsy. Rimme Malcolm. "The Social and Organizational Life Data Archive (SOLDA)." *Journal of Computing in Higher Education. V11 N2 P63-74 Spr 2000* .

Roberts, A. M. Cluckie IDGray LGriffith RJLane AMoore RJPedder MA. "Data Management and Data Archive for the HYREX Programme." *Hydrology and Earth System Sciences*.

Roghmann, K. J., and R. J. Haggerty. "Mini Data Archives: the Rochester Child Health Studies Masterfiles." *Inquiry. 9(1):66-71, 1972 Mar.*

Scheuch, Erwin K. "From a Data Archive to an Infrastructure for the Social Sciences." *International-Social-Science-Journal* 42, 1(123), Feb, (1990).

Schürer, Kevin. “The Irish Data Archive Feasibility Project - A Report” The Data Archive, University of Essex.

Tanenbaum Eric, and Mochmann Ekkehard. "Towards a European Database for Comparative Social Research." *Historical-Social-Research / Historische-Sozialforschung; 1996, 21, 2(78), 118-122*.

Thomas, S. P. "Issues in Data Management and Storage." *Journal of Neuroscience Nursing. 25(4):243-245, 1993 Aug* .

Waters. Donald J. “Toward a System of Digital Archives: Some Technological, Political and Economic Considerations: The Rearranging Effect Presented at the Association of Research Libraries Meeting, October 1998. http://arl.cni.org/arl/proceedings/131/waters.html

Zelenock, Tom, and Kaye Marz "Archiving Social Science Data: A Collaborative Process.". ICPSR Bulletin, May
1997. Inter-University Consortium for Political and Social Research (ICPSR)
http://www.icpsr.umich.edu/ORG/Publications/Bulletin/May97/article.html

Technology Usage / Adaptations

"Defending Secrets, Sharing Data : New Locks and Keys for Electronic Information." *Washington, DC : Congress of the U.S., Office of Technology Assessment, U.S. G.P.O., [1987] Vi, 187*.

"Selected Papers From TEI 10: Celebrating the Tenth Anniversary of the Text Encoding Initiative." *Computers and the Humanities v 33 No1/2 Apr 1999* p. (206).

Badard T, and Richard D. "Using XML for the Exchange of Updating Information Between Geographical Information Systems." *Computers-Environment-and-Urban-Systems* 2001; (2001).

Balstad Miller, R. "On-Line Access and Analysis of Data: Developing A Social Science Information Management Strategy." *1996 Mar : Arlington; VA:* United States; Department of Commerce; Bureau of the Census. Department of Commerce; 1996.

Bentley, J. E. "A Data Access Challenge: Build a SAS Frame Application for Access to an Informix Data Warehouse." *Annual International Conference; 23rd – 1998 Mar : Nashville; TN:* SAS Institute User Group. SAS; 1998.

Bonatti, P., et al. "An Access Control Model for Data Archives." *International Conference; 16th – 2001 Jun : Paris:* International Federation for Information Processing. Boston; London; Kluwer Academic; c2001.

Brantgarde, Lennart. "Documentation of Academic Research Data and the Hypertext Concept- a Presentation of the Swedish System ELSA." *Tidskrift for Dokumentation* v 48 No2 1993 p. 1956-1961.

Card, Josefina J. and McKean Elizabeth. "The Development of Software to Facilitate Use of Archived Data Sets." *IASSIST Newsletter.* 19:3. 4-11 (1995).

Chappell, C. Struijve O. and Anderson S. "The History Data Service – Using Technology to Enhance Access." *IASSIST Newsletter.* 22:4. 17-20. (1998).

Chen, Su-Shing. "PERSPECTIVES - The Paradox of Digital Preservation - Preserving Digital Information Requires Proven Methods for Maintaining, Accessing, and Authenticating Technology-Generated Data." *Computer:* 2001.

Chiang, Katherine-S, Jan Olsen, and William-V Garrison. "Beyond the Data Archive: the Creation of an Interactive Numeric File Retrieval System." *Library Hi Tech* v 11 No3 1993. P. 57-72 .

Cleden, D, Proud R , and Hills N . "New Interface Approach to Complex Data Archive Systems for Remote Sensing and Mineral Exploration Applications." *in: Remote Sensing: an Operational Technology for the Mining and Petroleum Industries. Conference, IMM, London, 1990. (IMM), 1990, Pp 15-25* .

Curran, Patrick. "INCORE Metadatabase Server - Technical Aspects and Constraints." *IASSIST Newsletter. 19:3. 12-19* (1995).

Ellis, R Darin. And Others. "Gero-Informatics and the Internet: Loading Gerontology Information on the World Wide Web (WWW)." *Gerontologist. V36 N1 P100-05 Feb 1996* .

Elson, Lee, Mark Allen, and Jeff Goldsmith. "An Example of a Network-Based Approach to Data Access, Visualization, Interactive Analysis, and Distribution." *Bulletin of the American Meteorological Society v 81 No3 Mar 2000*.

Emery, W, et al. "On-Line Access to Weather Satellite Imagery and Image Manipulation Software." *Bulletin,-American-Meteorological-Society* 1995.

Fergerson, James C. "Appendix A: A Computer and Network Resource Guide to Support Data Sharing." *New Directions for Institutional Research. N89 P79-90 Spr 1996* .

Franzen, H. "Facilitated Access to Information Sources: New Approaches in Organisation, Management and Delivery of Data and." *International Workshop – 1999 Oct : Alghero, Italy:* European Commission. European Commission, Directorate-General for Research; 2001.

Gardner T, Ruggles S, and Sobek M. "IPUMS Data Extraction System." *Historical-Methods* (1999).

Gartner H, et al. "OPALIS - An Open Paleoecological Information System." *Geo-Informations-Systeme* 2000.

Goffe, William L, and Robert P Parks. "The Future Information Infrastructure in Economics." *Journal of Economic Perspectives v 11 Summer 1997* p. (1975-1994).

Grout, Catherine, Phill Purdy, Janine Rymer "Creating Digital Resources for the Visual Arts: Standards and GoodPractice.". Visual Arts Data Service (VADS)/ AHDS. http://vads.ahds.ac.uk/guides/creating_guide/contents.html

Harger, C., et al. "The Genome Sequence DataBase (GSDB): Improving Data Quality and Data Access." *Nucleic Acids Research. 26(1):21-6, 1998 Jan 1* .

Hawick, K. A. Coddington PD. "Interfacing to Distributed Active Data Archives." *Future Generation Computer Systems* .

Heikkila, Eric J. "GIS Is Dead; Long Live GIS." *Journal of the American Planning Association v 64 No3 Summer 1998* p. (350-1960).

Hinke, T. H., et al. "For Scientific Data Discovery: Why Can't the Archive Be More Like the Web?" *International Conference; 9th – 1997 Aug : Olympia; WA:* IEEE. IEEE; 1997.

Ide, Nancy M., Sperberg M., and C McQueen. "The TEI: History, Goals, and Future." *Computers and the Humanities v 29 No1 1995* (1995).

Jacob, J. "Statistics and Data Protection: A Global View." *International Seminar – 1994 Nov : Luxembourg:* Eurostat; 1995.

Kerschberg, L. "Knowledge Management in Heterogeneous Data Warehouse Environments." *International Conference; 3rd – 2001 Sep : Munich, Germany:* New York; Springer-Verlag; 2001.

Kilbride, W. "Piloting an Archive for a New Millennium: Researching into, and Researching With Digital Data." *Conference – 1999 Nov : Stoke on Trent:* Society of Museum Archaeologists. SMA; 2001.

Kim T. J. "Metadata for Geo-Spatial Data Sharing: A Comparative Analysis." *Annals-of-Regional-Science* 33(2) (1999).

Kruse, Robin L., Bernard G. Ewigman, and George C. Tremblay. "The Zipper: a Method for Using Personal Identifiers to Link Data While Preserving Confidentiality." *Child Abuse & Neglect* (

Sep2001).

Kumar, R. "Ensuring Data Security in Interrelated Tabular Data." *Symposium – 1994 May : Oakland; CA:* IEEE Computer Society Press; 1994.

Letourneau F, Bedard Y, and Moulin B. "Prospects for the Use of the Data Warehouse Concept for Geo-referenced Digital Libraries on the Internet." *Geomatica* 52 (2) (1998).

Lins, J. E. "On-Line Access and Analysis of Data: Developing A Social Science Information Management Strategy: Floor Discussion." *1996 Mar : Arlington; VA:* United States; Department of Commerce; Bureau of the Census. Department of Commerce; 1996.

Ludacher, Bertram, Richard Marciano, and Reagan Moore. "Special Section on Advanced XML Data Processing - Preservation of Digital Data With Self-Validating, Self-Instantiating Knowledge-Based Archives." *SIGMOD Record:* 2001.

McCain, KW. "Sharing Digitized Research-Related Information on the World Wide Web." *Journal of the American Society for Information Science*.

Miller, P. "The Importance of Metadata to Archaeology: One View From Within the Archaeology Data Service: BAR International Series." *BAR International Series (Supplementary)*.

Moore, Fred. "Long Term Data Preservation." *Computer Technology Review* Third Quarter (1999).

—. "Profiling the Storage Hierarchy." *Computer Technology Review Los Angeles* 20 (2000).

Morrison, A. "Delivering Electronic Texts Over the Web: the Current and Planned Practices of the Oxford Text Archive." *Computers and the Humanities* v 33 no 1/2 (1999).

Morrison, Alan, Michael Popham and Karen Wikander "Creating and Documenting Electronic Texts: A Guide to
Good Practice.". OTA/AHDS. http://ota.ahds.ac.uk/documents/creating/

Morrison, Christopher. "Coalition Invents Handshakes for Biomed R&D Data Sharing." *IEEE Spectrum New York* 38 (2001).

Olson, R. J., et al. "Managing Data From Multiple Disciplines, Scales, and Sites to Support Synthesis and Modeling." *Remote-Sensing-of-Environment. 1999; 70(1): 99-107* .

Oswald, R. "Shared Vision: Manitoba Partners Build Land Information Data Warehouse." *Geo-Info-Systems* 1996.

Pirenne, B., P. Benvenuti, and R. Albrecht. "Implementing a New Data Archive Paradigm to Face HST Increased Data Flow." *Annual Conference; 6th – 1996 Sep : Charlottesville; VA:* Astronomical Society of the Pacific; 1997.

Roddick, J. F. "Data Warehousing and Data Mining - Are We Working on the Right Things?" *Workshop – 1998 Nov : Singapore:* Berlin; Springer; 1999.

Seligman, Len, and Arnon Rosenthal. "RESEARCH - XML's Impact on Databases and Data Sharing." *Computer:* 2001.

Shaya, E., et al. "XML at the ADC: Steps to a Next Generation Data Archive." *Alfred Benzon Symposium – 1998 Aug : Copenhagen:* Alfred Benzon Foundation. Munksgaard; 1999.

Sprehe, J Timothy. "The U.S. Census Bureau's Data Access and Dissemination System (DADS)." Government Information Quarterly. v14 n1 p91-100 1997.

Vries, Repke Eduard de. "Oracle Eases Data Sharing." *Informationweek. No. 610,* (1996).

SOCIAL SCIENCES AND HUMANITIES RESEARCH COUNCIL OF CANADA

We build understanding

The Social Sciences and Humanities Research Council of Canada (SSHRC) is an arm's length federal agency that promotes and supports university-based research and training in the social sciences and humanities. Created by an act of Parliament in 1977, SSHRC is governed by a 22-member Council that reports to Parliament through the Minister of Industry.

SSHRC-funded research fuels innovative thinking about real life issues, including the economy, education, health care, the environment, immigration, globalization, language, ethics, peace, security, human rights, law, poverty, mass communication, politics, literature, addiction, pop culture, sexuality, religion, Aboriginal rights, the past, our future.

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

350 Albert Street
P.O. Box 1610
Ottawa, ON K1P 6G4
Canada

Phone: (613) 992-0691
Fax: (613) 992-1787
Internet: www.sshrc.ca