A National Research Data Management Strategy for Canada: The Work of the National Data Archive Consultation Working Group

*by Yvette Hackett**

Introduction
I will be speaking today about an initiative currently underway in Canada related to the preservation of data, specifically data generated by and useful for research in the social sciences and humanities.

Background
I'd like to begin with some background information about my institution and its "relationship" with data over the years.

The National Archives of Canada has existed since 1872 and has, compared to some other countries, an extremely wide collecting mandate which we have dubbed "total archives". It means that we collect unpublished records from both public and private sector-sources in all media. It is the issue of publication which distinguishes our primary area of responsibility from that of the National Library of Canada, which collects published material.

In the early 1970's, an interest in "computer data" and its preservation developed within the institution and by 1973, a Machine Readable Archives Division had been created with a mandate to acquire research data from both public and private sector sources, in support of both the social and physical sciences. It became the de facto "national data archives" in Canada.

By the mid-1980's, automation had begun to affect the creation of what archives considered their "traditional" records - correspondence, memos, reports, case files, etc. In reacting to this situation, the National Archives decided to""integrate" the data archivists with the traditional archivists who would be most affected by the changes being wrought by automation, specifically those responsible for government paper, private paper, and cartographic records

The intention was to cross-train everyone to do both paper and electronic records. For many reasons, including timing, the lack of human and financial resources, and other government priorities, the NA's role in, and commitment to data acquisition, preservation and access in Canada slowly narrowed to focus on a very small number of government-generated databases, such as the census.

Immediate Triggers
Two events have occurred in recent years to move the issue of data preservation and access back onto the government's agenda. The first was a review of the National Archives and National Library of Canada's mandates, requested by Heritage Canada. Dr. John English was appointed to investigate. Among the interested groups who presented a submission during the hearings was the Canadian Association of Public Data Users (CAPDU) represented by Chuck Humphrey, Ernie Boyko and Wendy Watkins, who are well-known to many members of IASSIST.

Dr. English's most important recommendation, in this context at least, was his endorsement of CAPDU's position that Canada needed a National Data Management Strategy:

> We endorse the Canadian Association of Public Data Users proposal for a National Data Management Strategy in which the National Archives and the National Library play a facilitative role. The two institutions should play a partnership role in such a data archive and coordinate the federal government's relationship with such an archive.

The second triggering event was a workshop co-sponsored by the Social Sciences and Humanities Research Council (SSHRC) and the Organization for Economic Cooperation and Development (OECD). In exploring Canada's social sciences infrastructure, participants highlighted the problems of access to Canadian research data. This issue was of particular concern to SSHRC, which funds a large proportion of academic research in Canada. Their grants enable a great deal of data collection and analysis and participants at the workshop emphasized the fact that the lack of a national data strategy has made the resulting data sets difficult to access, and has hindered Canada's ability to coordinate national developments and participate in international initiatives.

Joint Investigation
The impact of these two events led SSHRC to contact the National Archives of Canada and propose a co-sponsored investigation. At this stage, it would be restricted to the

social sciences and humanities, with the hope that the work would attract the attention, and possibly, the participation of other "data" groups at a later date, including the natural sciences, health sciences, environmental sciences, etc. The National Data Archive Consultation Working Group was formed over the summer of 2000.

The 9-member Group includes representation from a number of disciplines with experience in the creation and use of research data, such as political science, history and English. Its membership also includes a representative from the archival community, Luciana Duranti, who some of you may know as the Project Director of InterPARES, an international research project investigating the authenticity of electronic records. The Working Group also includes one member from the data library community, Chuck Humphrey, and Sue (Gavrel) Bryant, a former President of IASSIST, who had worked in the Machine Readable Archives Division at the National Archives before moving on to a career in information management at Treasury Board, a central agency of the federal government.

Working Group Members

John ApSimon, Chair
Special Advisor to the President,
Carleton University

Gérard Boismenu
Political Science,
Université de Montréal

Sue Bryant
Treasury Board, PKI Secretariat

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

Chuck Humphrey
Data Library, University of Alberta

José Igartua
Département d'histoire,
Université du Québec à Montréal

Ian Lancashire
Department of English, University of Toronto
Michael Murphy

Rogers Communications Centre,
Ryerson University

Matthew Mendelsohn
Department of Political Studies,
Queen's University

In addition to the Working Group, a 14-member Resource Group was also invited to participate. The composition of the Resource Group is similar to that of the Working Group, adding subject expertise in sociology, geographic information systems, modern languages and new media. Ernie Boyko represents Canada's main statistical agency, Statistics Canada. Wendy Watkins provides additional representation from the data library community. Chuck, Ernie and Wendy also bring extensive experience from both sides of Canada's Data Liberation Initiative, an earlier and very successful project to improve research access to Statistics Canada's data. My role, and that of my colleague from the National Library of Canada, focussed primarily on providing information relating to the mandates and current activities of our respective institutions. David Moorman, a policy analyst at SSHRC, coordinated the two groups' activities.

Resource Group Members

Paul Bernard
Dept. of Sociology, Université de Montréal

Ernie Boyko
Library and Information Centre, Statistics Canada

Martin Brooks
Institute of Information Technology, National
Research Council

Joseph Desloges
Department of Geography, University of Toronto

Yvette Hackett
National Archives of Canada

Douglas Hodges
National Library of Canada

Terry Kuny
XIST Inc.

Timothy Jackson
Ryerson University

Wanda Noel
Barrister and Solicitor

Frits Pannekoek
Information Resources, University of Calgary

Michael Ridley
Chief Librarian, University of Guelph

Geoffrey Rockwell
Dept of Modern Languages, McMaster University

Fraser Taylor
Department of Geography, Carleton University

Wendy Watkins
Data Centre, Carleton University

Term of Reference

The Terms of Reference for the Working Group proposed a two-phase structure, with the second phase contingent on the results of the first phase. The focus of the investigation was distilled into a series of questions to be answered. The four Phase One questions included:

1. To what extent is there a **need** for a unified and coordinated data archiving function? Are modest changes to existing institutional policies and mechanisms adequate to meet current and future requirements?

2. What **gaps** exist in the mandates and structures of existing institutions in relation to management of research data?

3. Who will **benefit** from the improved management of research data and to what degree?

4. How will effective research data management, preservation and access contribute to Canadian **research capacity**?

Following a study of these"needs, gaps, benefits, and research capacity" questions, the Working Group would submit a preliminary report to SSHRC and the National Archives. The next steps would be determined both by the recommendation of the Working Group and the response of the sponsoring institutions.

A working methodology rapidly evolved that depended equally on members of the Working Group and the Resource Group. The process began with a Stakeholders' Meeting, held in Ottawa in October 2000. Fifty-five people attended, representing universities, federal government departments, research groups, academic associations, archives and libraries across the country. In the course of a day-long consultation, a wide range of problems with the current Canadian situation were identified. The key ones included:

- difficulty in locating Canadian data
- difficulty in gaining access to previously collected Canadian data, due to costs and the lack of any kind of central resource directory, or depository service
- a weak tradition, within the Canadian research community, of making data available for re-use or for replication studies
- a lack of "national data" leading to obstacles in Canadian participation in multi-national studies
- a lack of a recognized national institution to facilitate Canada's participation in international research projects, in international associations and in standards development

Participants in the stakeholders' meeting also heard of many initiatives currently underway, though most were addressing access issues only. They tended to be organized on an institutional, regional or disciplinary basis and were operating largely in isolation from each other. Over 20 participants followed up with written submissions to the Working Group.

The Stakeholders Meeting pointed out the need to clearly define the focus of the Group's investigations. The following "scope" statement was developed:

> A research data function would have the goal of preserving, managing and making publicly accessible digital information, structured through methodology and documentation, for the purpose of producing new knowledge. This function would address the gap that exists between the raw research materials and formally published results. Acquisition would include digital information both produced by researchers and of interest to researchers.

It emphasizes:

- the 3 facets of the required strategy - access, management and preservation;
- the digital nature of the material, and the importance of a structured methodology; and
- the focus on the gap between raw data and published results.

From a number of proposed research strategies, the Working Group focussed on three additional activities that they would undertake:

- the preparation of briefs outlining the "research data" situation in their particular areas of specialization;

- research to understand where "research data" fit into the mandates of existing institutions such as the National Archives and the National Library, as well as the role of the many university-based data libraries and archives.

- the organization of 4 surveys to elicit concrete data to support the opinions expressed at the Stakeholders Meeting; surveys were directed to SSHRC-funded researchers; university data

archivists; participating institutions in the Data Liberation Initiative; and finally a list of general stakeholders who had attended the October meeting, or otherwise expressed their interest in this issue.

The Working Group is scheduled to submit its Phase One report in early June. The report confirms that a serious gap exists in Canada's research infrastructure and argues that the preservation of research data, and the facilitation of on-going preservation, management and access to such data are important factors in building the research capacity so necessary to the growth of a "knowledge society". As a result of these findings, the report will recommend that Phase Two be undertaken to study possible mechanisms to accomplish these goals.

Plans - Phase Two

The Terms of Reference for Phase Two have already been established, in anticipation of Phase Two proceeding. This time, the questions include:

1. Is some form of **national data archiving agency** the right way to meet the needs of the research community?

2. Are there **alternative ways** of meeting the needs of researchers?

3. If a new national facility is recommended, what functions should it perform and what institutional form should it take?

4. What is the **most appropriate working relationship** between a new facility and existing agencies such as the **National Archives of Canada and the National Library of Canada**? How should duplication of responsibilities and services be avoided?

5. How can a data preservation and access facility best take advantage of emerging information and communication technologies to increase its efficiency and effectiveness?

A specific methodology for Phase Two has not yet been developed, but three obvious areas to pursue have emerged.

The first would involve research into the various organizational models already in place in Canada and abroad, and in specific disciplines. This process should include the identification of the strengths and weaknesses of each and an analysis of their applicability in the current Canadian landscape. A second issue could address jurisdictional issues, as the needs of the private and public sectors are considered, including federal, provincial and municipal levels of government and the full spectrum of Canada's academic community, all of whom are potentially both creators and users of research data. A third issue would address the availability and appropriateness of various funding mechanisms.

Assuming a prompt and positive response from its sponsoring agencies, the Working Group hopes to complete work by December 2001.

* Paper presented at the IASSIST/IFDO conference 2001 in Amsterdam. Yvette Hackett, National Archives of Canada, Yhackett@archives.ca