Management and Preservation of Geospatial Data

Ad-Hoc Committee on Archiving and Preserving Geospatial Data
Policy Advisory Network Node
GeoConnections

David L. Brown: Library and Archives Canada
Grace Welch: University of Ottawa
Christine Cullingworth: University of Ottawa

April 6, 2004

Document Identification

Title	Management and Preservation of Geospatial Data
Creator	David Brown, Library and Archives of Canada Grace Welch, University of Ottawa Christine Cullingworth, University of Ottawa
Subject	Geospatial Data, Geomatic Information
Description	
Publisher	
Contributor	
Date	April 6, 2004
Type	Text
Format	Microsoft Word 2000
Identifier	
Source	
Language	English
Relation	
Coverage	
Rights	

Standard Document Identification – Dublin Core Metadata Element Set Version 1.1 1999-07-02

Table of Contents

1 INTRODUCTION

"The preservation and re-use of digital data and information forms both the cornerstone of future economic growth and development, and the foundation for the future of memory".[1]

Geospatial data, information that references the geographic location of natural and man-made phenomena on the surface of the Earth, have become an indispensable information asset in today's society. Geospatial data are the fuel that will drive an estimated $45 to $67 billion (US) world market for geomatics based products and services by the year 2004.[2] According to Statistics Canada's 2000 survey of the mapping and surveying services industry, there are over 2,000 companies in Canada that generate $1.5 billion worth of annual revenues from geomatics based activities.[3] From the design of the roads on which we travel, to the location of our place of work, nearly every facet of everyday life is touched, in some way, by geospatial data.

Geospatial data are being produced by all levels of government and in the private sector at an unprecedented rate. However, long-term access to the wealth of these data will be compromised unless policies and procedures are created and implemented by geospatial data custodians to ensure their preservation and continued availability to policy makers, industry and researchers. Decisions about our economy, environment and society cannot be based simply on current data; temporal analysis is required to identify trends, evaluate impacts and make informed decisions. Data preservation policies currently in place in all levels of government are inconsistent, or even non-extant, and do not address the wide-range of information management issues created by the digital environment.

Alarm bells are beginning to sound about the potential loss of this valuable information. Terms such as "catastrophic" and "imperiled" are being used to describe what could happen if steps are not taken to ensure the long-term preservation of these data. Numerous examples already exist where data have been lost. One includes the loss of data that were compiled by the State of New York for the completion of a land use and natural resource inventory. The data can no longer be queried and investigated because the software required to read them no longer exists.[4] In Canada, there is the example of the Canada Land Data System, where valuable land use information collected for the Canada Land Inventory (CLI) was nearly lost until four federal government departments jointly undertook a massive restoration project. Agriculture and Agri-Food Canada, National Archives of

---

1 Ross, Seamus. (2000). *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship, London*: National Preservation Office, URL: http://www.bl.uk/services/preservation/occpaper.pdf, last accessed: Feb. 11, 2003, p. 45.

2 Hickling Arthurs Low (HAL) Corporation. (2001). *Geomatics Sector Human Resources Study*, report initiated by the Canadian Council of Land Surveyors, the Canadian Institute of Geomatics, and the Geomatics Industry Association of Canada.

3 Statistics Canada. (2000). *2000 Survey of Service Industries: Surveying and Mapping,* North American Industrial Classification Systems (NAICS) codes 54136 and 54137.

4 Tristam, Claire. (2002). *Data Extinction*, Technology Review, October 2002, pp. 37-42.

Canada (NA), Natural Resources Canada (NRCan), and Statistics Canada completed restoration activities along with a private sector company known as Spatialanalysis.[5]

While considerable study and research is now being undertaken to address the subject of preserving electronic information, very little study focuses on the unique challenges to successfully preserve geospatial data.[6] This study will address the requirement for geospatial data creators and custodians to develop archival policies and implement preservation strategies that ensure long-term access to data that can be used to study the interrelationships of man and his physical environment. "The vision of a rich information record just waiting to be harvested and processed by the technology enabled researcher of the future depends upon the survival of digital data".[7]

Part of the solution to the preservation problem lies in the development of an enhanced infrastructure for geospatial data that is based upon sound information management and technology standards, and the application of suitable techniques and tools. A common infrastructure provides the framework to ensure there is uniformity to meet the needs of the geomatics community. It also makes certain that data collection and maintenance are undertaken in the national interests of Canadian society.

An effective information management infrastructure must be comprised of three essential components to ensure its success:

1. Information management legislation and policy to ensure the information required by society is collected, managed, preserved, and made accessible.

2. A governance and accountability framework to ensure participation in the setting, change and adoption of legislation and policy.

---

5 Brown, David L. & Comeau, Mike. (1999). Restoration of the Canada Land Data System, *Association of Canadian Map Libraries and Archives Bulletin,* Number 106, Autumn, pp. 42-52.

6 Numerous studies have been completed about the need to preserve digital information outside the GIS technology domain. To obtain information from the USA, readers should consult: 1. Hedstrom, Margaret & Montgomery, Sheon; (1999), *Digital Preservation Needs and Requirements in RLG Member Institutions,* Research Libraries Group, Mountain View, California, 94041, USA. 2. Council on Library and Information Resources, and Library of Congress. (2002). *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, Washington, DC, 99 pp. The National Library of Australia has a staff presentation and papers series about preservation issues that can be found on the Internet at http://www.nla.gov.au/nla/staffpaper/preserve.html. In addition, it supports a website entitled *Preserving Access to Digital Information (PADI)* that can be used as a portal site to obtain information about digital preservation issues, http://www.nla.gov.au/padi. In the United Kingdom, the Digital Preservation Coalition website has been established to foster the dissemination of information about the preservation of digital materials, http://www.dpconline.org/text/digpres/presissues.html. In Canada, research about the long-term preservation of authentic electronic records is currently being completed under the auspices of the InterPARES Project, http://www.interpares.org/. With respect to preservations issues specifically relating to GIS, one should consult: 1. Bleakly, Denise R. (2002). *Long-Term Spatial Data Preservation and Archiving: What are the Issues?*, Sandia National Laboratories, Albuquerque, New Mexico and Livermore, California, 42 pp. 2. Zaslavsky, Ilya. (2001). *Archiving Spatial Data: Research Issues*, San Diego Supercomputer Center Technical Report, TR-2001-6.

7 Ross, Seamus. (2000). Op. Cit., p. 45.

3. A physical infrastructure to support the collection, management, preservation, and accessibility of the information or data.

This document addresses elements of all three components.

2 AIMS AND OBJECTIVES

The aim of this study is to:

1. Identify the legislative and policy framework required for the management of geospatial data created by federal, provincial and territorial governments.

2. Identify and define the information management issues that must be addressed for the integral management of geospatial data.

3. Review existing strategies and practices concerned with the preservation of geospatial data.

4. Identify available standards that are required and practiced for the logical and physical preservation of geospatial data.

5. Introduce archival criteria used for geospatial data selection and retention.

6. Develop a set of recommendations for the long-term archiving and preservation of geospatial data.

3 RESEARCH METHODOLOGY

An Ad-Hoc committee was created in 2001 by the Policy Advisory Network Node of GeoConnections to examine best practices for the preservation and archiving of geospatial data. Membership in the group included representatives from the Association of Canadian Map Libraries and Archives, Canadian Association of Public Data Users, National Archives of Canada, National Library of Canada, and Natural Resources Canada. As a first step in preparing the study, representatives from the Ordnance Survey, the United States Geological Survey (USGS), the Australia and New Zealand Land Information Council (ANZLIC) and the European Map Librarians Group were contacted to determine the status of policies in their jurisdictions. Almost universally, an ad-hoc approach has been taken, or issues related to the preservation of spatial data in digital form are still under study and review. There was general recognition that the lack of policies and standards was a serious problem, but one that could not be easily resolved.

In view of the lack of published standards and guidelines for the preservation of geomatic information, GeoConnections provided funding to contract a graduate student to work with members of the Ad-Hoc Committee on Archiving and Preserving Geospatial Data, to study current research about preserving geospatial data and propose solutions that could be used by Policy Node members to manage their data collections. During the summer of 2002, extensive research

was undertaken to examine the national and international literature about archiving, preserving and accessing digital information and identifying relevant practices.

In addition to an examination of secondary research sources, primary data were gathered by interviewing data creators and custodians.[8] These interviews were conducted to obtain empirical information about the issues and concerns that are experienced by these people in the conduct of their archiving and preservation activities. The survey was not designed to be completely exhaustive; it was done to identify key issues and questions that face Canadian geospatial data creators. Both personal and telephone interviews were held with seven government departments or program areas, all of which were major geospatial data users or producers.

This report summarizes the results of the study. It is supplemented with information from the National Archives of Canada. The focus of the report is on the preservation of government created information, since governments are the primary creators of geospatial data. Most of the country's government jurisdictions are not only mandated to collect data for the delivery of survey and mapping based programs and services, they must also perform specific roles and responsibilities vis-à-vis information management. These requirements are also identified in the report.

4 LEGISLATIVE AND INFORMATION POLICY FRAMEWORK

4.1 FEDERAL GOVERNMENT

4.1.1 DRAFT MANAGEMENT OF GOVERNMENT INFORMATION POLICY

The Government of Canada (GOC) is increasingly using information technologies to serve Canadians and to record its business, which requires it to ensure that information is accessible and useable over time and through technological change. Furthermore, the government must mange information to ensure that Canadians receive consistent service regardless of how they choose to obtain it, whether in-person, by telephone, through mail, or via the Internet.[9] The objective of the proposed 'Management of Government Information Policy' is to ensure that government information is managed effectively and efficiently throughout its life cycle.

The Policy should be ratified sometime during the spring of 2003. It provides direction on how government institutions, departments and agencies should create, use, manage and preserve information in a comprehensive and strategic manner. The policy applies to all institutions listed in

---

[8] The following individuals were interviewed: Peter Shute, Agriculture and Agri-Food Canada; Marc Lemaire, Mapping Service Branch, Natural Resources Canada (NRCan). Peter Paul, Atlas Framework Section, NRCan; Joseph Anawati, Jeff Karau, and Dr. Jacques Trencia, Canadian Forestry Service, NRCan; Timothy V. Evangelatos, Rick Mehlman, Doug Brynt, and Colin Bromfield, Canadian Hydrographic Service, Department of Fisheries and Oceans; Sylvain LeMay, Elections Canada; Brian Maloney, Greg Bay, and Rob Parry, Ontario Ministry of Natural Resources.

[9] This is the primary objective of the Government of Canada's 'Government On-Line' initiative which proposes to build the technical infrastructure to deliver all government services on the internet by the year 2005.

Schedules I, I.1 and II of the Financial Administration Act (FAA). The key premise of the policy is that the preferred future record of government will be digital.

The policy advocates that institutions:

- Ensure that governance and accountability structures are implemented for the cost effective and coordinated management of information under their control to support effective decision-making, services and program delivery.

- Provide the infrastructure for the effective and efficient management of information, regardless of its medium or format, to ensure its authenticity and integrity for as long as it is required by legislation, departmental statutes, and other laws and policies.

- Manage information to facilitate its universal access by anyone and in a manner that optimizes its sharing and re-use in accordance with legal and policy obligations.

- Document the decision-making processes throughout the evolution of policies, programs, and service delivery.

- Preserve information of enduring value to the Government of Canada or to Canadians.

- Establish a coordinated and comprehensive approach to describing the institution's information.

- Maintain a current and comprehensive classification structure(s), including metadata.

The leadership required to achieve the objectives of the policy will be provided through the Treasury Board Secretariat and the Library and Archives of Canada. These agencies are responsible for maintaining an overall understanding of the state of information management practices and providing the appropriate control mechanisms across government. These agencies will work with government institutions to help solve information management concerns and issues, and lead government-wide information management improvement initiatives.

Both the *National Archives of Canada Act* and the *National Library Act* will be harmonized into one piece of legislation to form the Library and Archives of Canada sometime during the summer of 2003. However, the key objects and functions of the current legislation will not substantively change.

4.1.2 NATIONAL ARCHIVES OF CANADA ACT

Under the auspices of the *National Archives of Canada Act*, the National Archives has responsibility for preserving the collective memory of the nation and the Government of Canada. Under Section 4 of the NA Act, the Archives can acquire 'records' (sec. 4.2.a) from the 'private and public' (sec. 4.1) sectors that it considers to be of national significance (sec. 4.1). Under the definition of a record in the *Act* this includes 'machine readable record[s]' which includes geospatial data.

The National Archives and federal government departments have specific roles and responsibilities related to the management of government information. Under Section 5, 'no record under the

control of a government institution … shall be destroyed or disposed of without the consent of the Archivist'. Under this scenario, the National Archives has a responsibility to work with federal government departments to develop record disposition authorities (RDAs) that identifies archival information that should be transferred to the legal control of the National Archives after its operational value to an organization has ceased. During this process, information is assessed (be it analogue or digital) to determine its enduring value to Canadians. For information that is identified as being of historical value, an agreement is developed that outlines the dual roles and responsibilities of the organization and the Archives to ensure the integrity of that information while it resides in the custody of a department. The National Archives also has a responsibility to work with departments to help them manage the body of information that is not considered to be of historical value, but contains considerable business value to an organization.

4.1.3 NATIONAL LIBRARY ACT

The National Library of Canada's primary function is to acquire, organize, preserve and make accessible the published heritage of Canada to all Canadians. Under the *National Library Act*, all newly published information in various formats (including digital) from federal government departments and agencies must be deposited with the National Library of Canada. Similar to the Archives, the Library has an information management role to play, but the emphasis is with respect to published information.

In 1995, the National Archives and the National Library entered into an agreement whereby the National Archives acquires and preserves both published and unpublished cartographic materials consisting of separately published single sheet and series maps, atlases, globes and geomatic information. Both institutions collect atlases and other published compilations of maps in all formats, including CD-ROM and other electronic formats.

Although separately published maps are not subject to legal deposit, the National Archivist may request the National Librarian to invoke the exception clause in the National Library Legal Deposit Regulations to acquire separately published maps of national significance. The *National Library Act* is currently under review, and there is a proposal to modify the Regulations to include maps (in all forms) as part of the legal deposit program.

4.1.4 COPYRIGHT ACT

The Canadian constitution provides that activities that are completed under the Government of Canada are to be carried out in the Queen's name. Today, the Crown includes units of the federal, provincial and territorial governments. Each of these units governs as part of the Crown and therefore exercises Crown copyright. Copyright protects 'intellectual' as opposed to 'physical' property.

Section 12 of the *Copyright Act* provides that the Crown owns copyright in any work that has been prepared or published by the Crown, or under its direction. The *Act* also stipulates that the copyright in the work shall continue for the remainder of the calendar year of first publication and for a period of 50 additional years. At the end of the 50 years term, a work falls into the 'public domain'.

Canadian Government Publishing of Communication Canada is responsible for giving permission

and/or issuing license agreements giving an individual, a publisher or the private sector written authorization to reproduce Crown works still under copyright. This is done with an author department's authorization.

Copyright in unpublished Crown works is perpetual. The majority of the information that is created in the federal government is of an unpublished nature. As a result, this information comes under scrutiny of the National Archives of Canada. Under these circumstances, Section 6 (1) of the *National Archives of Canada Act* implies that the transfer of archival records (in this case, unpublished geospatial data) includes not only the physical transfer of the records (care), but the transfer of 'crown copyright' and the intellectual property rights associated with them (control).

4.1.5 OTHER RELEVANT LEGISLATION

- Access to Information Act
- Canada Evidence Act
- Emergency Preparedness Act
- Personal Information Protection and Electronic Documents Act
- Privacy Act
- Statistics Act

4.2 PROVINCIAL AND TERRITORIAL GOVERNMENTS

There is a legislative basis for provincial and territorial governments to preserve and archive electronic information. This section identifies the legislative and policy basis for these governments to manage their electronic data holdings. The Association of Commonwealth Archivists and Records Managers (ACARM) have identified the archival legislation for most Commonwealth countries. The results of this review have been consulted and updated with information from provincial and territorial archival websites. [10]

4.2.1 ALBERTA

In Alberta, the *Historical Resources Act of* 1980 ensures that Alberta Community Development oversees the operations of the Provincial Archives of Alberta. It is also responsible for the management of records under the auspices of the *Government Organization Act* (Schedule II) and the *Records Management Regulation*. Under the regulation, departments and offices of the provincial government must manage their records according to corporate standards, guidelines and policies to support the delivery of their programs and services.

---

[10] Catherine Bailey of the National Archives of Canada prepared an extract of Canadian legislative information for the authors from information that was attributed to the following report: Parer, Dagmar, ed. (2001). *Archival Legislation for Commonwealth Countries, Area 4: Europe and North America.* Cairns, Australia : Association of Commonwealth Archivists and Records Managers (ACARM), pp. 79-110.

4.2.2 BRITISH COLUMBIA

The British Columbia Archives provides a government-wide framework for the management of all recorded information, including electronic, as defined in the *Interpretation Act*. The B.C. Archives administers the provisions of the *Document Disposal Act*, and establishes corporate standards for the creation, security, accessibility, retention, disposition and preservation of records under *Chapter 8* of the *General Management Operation Policy* and the *Recorded Information Management Manual.*

4.2.3 MANITOBA

Under the *Archives and Recordkeeping Act of 2001*, the Provincial Archives of Manitoba is responsible for the preservation and protection of archival records, including electronic, from provincial and local governments, courts of law, school authorities, businesses, organizations and individuals, and the provision of access to that information.

4.2.4 NEW BRUNSWICK

Established in 1967, the Provincial Archives of New Brunswick provides advice and assistance in records and information management to all government departments, agencies, corporations and commissions. Under the *Archives Act of 1977*, records management policies and procedures apply to all records in any form, including machine readable, which are created or received in the administration of public business.

4.2.5 NEWFOUNDLAND AND LABRADOR

The Provincial Archives of Newfoundland and Labrador (PANL) is mandated through the *1983 Archives Act* to preserve records of the government that are deemed to have enduring legal, fiscal, and evidential or research value. PANL is also responsible for the government's records and information management program and develops and implements information management policies, standards and services. No specific mention is made about electronic records.

4.2.6 NORTHWEST TERRITORIES AND NUNAVUT

Under the *Northwest Territories Archives Act of 1988,* the primary task of the NWT Archives is to ensure that a fair, accurate and complete record of government activity is identified, acquired and preserved. Records include any correspondence, memorandum, book, plan, map, drawing, diagram, pictorial or graphic work, photograph, film, microfilm, sound recording, video tape, machine-readable record, manuscript, inventory, pamphlet, periodical, photographic slide, micrographic, electronic data print-out, and any other documentary material regardless of its physical form or characteristics, held by or under control of a government body. The NWT Archives Act is currently being used as the basis of archival legislation for Nunavut.

4.2.7 NOVA SCOTIA

The mandate of Nova Scotia Archives and Records Management, as embodied in the *Public Archives*

Act of 1998 and the *Government Records Act of 1995-96*, is to acquire and preserve government records of provincial significance and facilitate access to them; develop policies, standards, procedures and services for records management activities in government; and, be the permanent repository of the records of public bodies. Records are defined as information or data that is recorded or stored by graphic, photographic, electronic, mechanical or other means.

4.2.8 ONTARIO

The Archives of Ontario was created in 1903 and is governed by the *1923 Archives Act*. Under a 1990 revision of the *Act,* the Archives of Ontario acquires original documents, parchments, manuscripts, papers, records and other matters in the executive and administrative departments of the Government or of the Assembly, or of any commission, office or branch of the public service. *The Act* does not mention anything about electronic records. However, the Archives intends to integrate the management of electronic records into its current policies and practices.

4.2.9 PRINCE EDWARD ISLAND

The Public Archives and Records Office of Prince Edward Island (PARO) acquires, preserves, and makes available for public research the records of the government of the province. The P.E.I. *Archives and Records Act of 2001* includes a provision for the acquisition and management of 'electronic data' and 'machine readable record[s]'.

4.2.10 QUEBEC

Under the auspices of the *1983 Archives Act*, the Archives nationales du Québec acquires documents from the government of Quebec. The Archives acquires documents created from government business activities. Documents are defined as any medium of information, including the data on it, legible directly or indirectly by machine.

4.2.11 SASKATCHEWAN

Under the auspices of *The Archives Act of 1945*, the Saskatchewan Archives' primary responsibility is to acquire the official records of the Government of Saskatchewan and provide records management services for the provincial government by developing and disseminating records/information management standards and providing advice on classification systems and electronic records management requirements.

4.2.12 YUKON TERRITORY

Under the auspices of the *Archives Act of 1979*, the Yukon Archives acquires, preserves and makes available documentary sources related to the Yukon and the circumpolar North, including government, private and corporate records, photographs, films, sound recordings, maps, newspapers, and other published sources.

Although all of the provincial and territorial governments have established archives legislation, some of which is supported with information management policy for digital information, it

would appear as though none of them is currently acquiring digital information. This is surprising considering that the vast majority of the information produced by governments is digital in nature. This trend is very concerning and represents either a lack of knowledge about the role of an archive in government operations. Or, it could also represent a lack of leadership to ensure that governance and accountability structures are implemented for the coordinated management of digital information under government control.

5 DATA AND INFORMATION MANAGEMENT ISSUES

Data collection, management, preservation and access activities in individual organizations and institutions are driven by the need for managers to make business decisions, and deliver products and services that are based on the use of reliable and accurate data. In many organizations today, data and information are maintained for only as long as they have immediate or short-term business value and they are not effectively managed after immediate interest in them has declined. Data and information go through phases of operational value to an organization and their value is often augmented and diminished over time. To maintain the accuracy and long-term value of data, there is a necessity for organizations to develop information management plans and adopt preservation strategies that are based upon an information life cycle management model (Figure 1). There is also a necessity for leadership in the management of geospatial data.

The goal of information management is to provide access to *information* that has been created and managed within an information management (IM) framework that assures its trustworthiness, integrity and authenticity over time. Information management and knowledge management have become key enablers for governments to deliver programs and services. Effective information management is critical to building an information infrastructure with the capacity to support and implement the Government of Canada's vision of a knowledge-based economy.[11] Fundamental to government's new business activity and governance models is an information management infrastructure of technology, functions, processes and standards that provide for public service accountability, sound strategic planning, and successful corporate business enterprise. It is this broad IM infrastructure of people, tools and protocols which support effective institutional decision-making, delivers programs and services to Canadians, and affords justification and protection in cases of audit, investigation and litigation.[12]

---

11 McDonald, John; (2000), Information Management in the Government of Canada - A Situation Analysis, For the Chief Information Office, Treasury Board Secretariat and the National Archivist, Ottawa, 41 pp.

12 Brown, Richard & Brown, David L. (2001), *Information Management in the Government of Canada: Improving Service Today and Planning for Tomorrow*, National Archives of Canada, unpublished report, August.

Figure 1: Record and Information Life Cycle Management Model

In terms of data management, most organizations find themselves situated some place along the following data management continuum, where:

1. Overall data architecture does not exist for the organization. There is no consistent definition of the data objects in use within the organization.
2. Overall data architecture is being developed, but is not yet complete and up to date. Data management standards such as the definition of data objects vary between organizational units and data transfer across organizational boundaries is costly and difficult.
3. Organization-wide data management standards have been defined and new applications adhere to those standards. Older applications need to be converted to conform to the standards.
4. Consistent standards for data formats and data transfer, and data object definitions are in place across the organization. There are widely available tools that support access to data objects across multiple platforms. There is widespread expertise in the use of these tools.
5. Consistent data management standards are in place across the organization and are compatible with those used externally, and thereby facilitate communication with clients, suppliers and business partners. Standards are rapidly adopted and adhered to across organizational boundaries, and provide a solid foundation for knowledge management and information sharing.[13]

Government program managers are entrusted with public resources and as data custodians they have a responsibility to protect the integrity of the data for which they are managerially accountable. Custodianship is the hub of an efficient and effective geospatial information management system

13 National Archives of Canada. (2003). *Information Management Capacity Check: Methodology and Guide*, 58 pp.

because it provides accountability for data management practices and indemnifies an authoritative source within the organization. Custodianship provides a means to facilitate data management on the behalf of others and it provides continuity in the delivery of a geospatial data infrastructure. Custodianship roles and responsibilities should ensure the ability of an organization to provide quality and accurate data for reporting, decision-making and research purposes.

Unfortunately, many managers have no idea where their organization lies along the data management continuum. In fact, most managers have no idea what types of data are collected for operational requirements. Nor, do they know anything about the manner in which those data have been collected. Often, when they do know the answers to these questions, they have no substantive information that describes the structural and semantic characteristics of those data to facilitate their future utility. We do know however, that the management of corporate data is not occurring in an orderly fashion, and many problems exist regardless of requirements stipulated in information management policy and legislation.[14]

In terms of volume, federal, provincial and territorial governments are storing terabytes of digital information, most of which is stored in a variety of logical formats and some of which is at risk of being lost because of the manner in which it is physically stored. Digital information is by its nature fragile and impermanent and will quickly become obsolete if it is not first, properly managed within the context under which it was created and used, and then moved to an environment that ensures its preservation over time. Preservation activities for geospatial data are but one element that organizations must address in the information life cycle management process. However, it is an issue of enormous importance, especially in the management of computer-readable digital assets. The goal of preservation is to ensure the maintenance and protection of a body of information for access by present and future generations.

6 TECHNOLOGY OBSOLESCENCE

One component of this study included interviewing data creators to obtain a better understanding about the practices that are currently being employed by organizations to ensure the long-term viability of their data. The challenge of managing and preserving digital information includes the development of a cost effective preservation strategy that will liberate the data from proprietary file formats that are dependant upon specific software and hardware. The creation of database backups that rely on the use of an operating system's restoration software cannot be considered to be a reliable long-term preservation strategy even though this approach may fulfill short-term operational needs. In addition, a preservation strategy must account for the volatility of the physical medium upon which the data are placed for short, medium and long-term storage requirements.

6.1 DATA REPRESENTATION

Some organizations and agencies have created data management policies for their geospatial data, but the application of these policies is often inconsistent. Interviews revealed that geospatial data are

[14] Treasury Board of Canada Secretariat. (2001). *IM Ready Report*, unpublished, Chief Information Officer Branch, Version 2, May, 23 pp.

being created in both vector and raster based data structures. Of these, the most frequently used file formats for storing and interchanging geospatial data include Environmental Systems Research Institute (ESRI) thematic coverage files, export files (E00), shape-files (shp), and the Spatial Data Engine (ArcSDE) format. Other supported formats included AutoDesk DXF; CARIS NTX and ASC files; PIX (PCI) files; GeoTIFF; TIFF; JPEG; COGIF; GIF; and, XML/GML encoding.

It is interesting to note that the majority of these data and interchange formats are based upon the application of industry standards rather than the adoption and implementation of national or ISO based standards. In fact, the geomatics area of the National Archives of Canada has adopted a number of industry based defacto standards such as the ESRI export file format for vector data and the uncompressed TIFF file format for raster data because they represent formats that are widely used and supported by geospatial data creators and users. The Archives also uses ISO standards such as IHO-S57 for the preservation and interchange of digital charts, but the CARIS ASC file format is also being supported. Unfortunately, from an operational perspective national and international standards are not being widely adopted and used within the geomatics community. As a result, many organizations support a number of functional geospatial data import and export formats.

Various derivatives of the Extensible Markup Language (XML) look promising for the future. XML is a meta-language that allows one to create descriptive tags for varying types of digital objects. Geography Markup Language (GML) is a dialect of XML and has been developed for handling geographic information. GML was designed with a number of objectives in mind, some of which overlap with the aims of XML. GML provides geospatial-encoding rules for both data transport and storage, and these rules are especially adaptable to an Internet GIS environment. The format is extensible enough to support a wide variety of geospatial functions and tasks, permits the efficient encoding of geospatial geometry, allows one to separate the spatial content from the non-spatial, and defines a common set of parameters for geographic objects to enable the interoperability of data between independently developed applications.

The use of varying hardware platforms from different manufactures also has an impact on the ability to interchange geospatial data between organizations. For instance, a CARIS NTX file that is created on a SUN Microsystems workstation that uses the Solaris operating system cannot be imported as a CARIS NTX file that uses the same CARIS software and version on a PC using the Windows operating system. The problem lies in the ability of each computer to interpret the character set values that were used to encode the data in each computer's operating environment. The problem is similar to that of taking 8-bit EBCDIC (Extended Binary Coded Decimal Interchange Code) data that have been created on an IBM mainframe computer and copying the data to a PC that uses the 8-bit ASCII (American Standard Code for Information Interchange) coded character set for information representation. The data from each platform will not be interoperable and interpretable unless they are migrated to a suitable interchange format. This task is not always as simple as using an ASCII to EBCDIC conversion software package, or vice versa, because some of the data may be represented by binaries that are stored and packed at the bit level. As a result, the data elements will be unintelligible without some form of custom programming.[15] To achieve this goal, one will require

[15] Brown and Comeau present an empirical account about problems associated with geospatial data conversion and migration activities in: Brown, David L. & Comeau, Mike. (2001). *Restoration of the Canada Land Inventory*, paper presented at the Social Science History Association Conference, Historical Geography Network, Archives in the Digital Age, Chicago, Illinois, November 15-18, 17 pp.

a detailed knowledge of the logical data model that is employed in the software environment. This knowledge can only be obtained if the structural and semantic elements of the geographic data model have been well documented.

Currently, Canadian data collection and management activities are driven by the individual needs of organizations and institutions. As a result, there is a lack of consistency in the use of homogeneous data structures both within and between organizations, especially at a national level. As indicated above, many data structures are concurrently used to manage geospatial data within corporate entities. From a managerial perspective, this is not only undesirable, but also problematic because it requires multiple investments in software, hardware and the development of people skills and knowledge management activities.

6.2 STORAGE TECHNOLOGIES

The physical media upon which data are stored must also be carefully considered. The market offers a large number of disc and tape storage devices and data management solutions that use a variety of optical, metal and polyurethane based storage mediums. Unfortunately, many of these Input/Output solutions are proprietary in nature and do not easily facilitate the interchange of data. One can have the best intentions by saving geospatial data in a standard based logical format, but unless similar standards based practices are extended to the physical storage environment, data obsolescence is sure to result.

In most offline storage environments, the life of the physical medium will usually outlast that of the device that was used to copy the data. Although storage media can theoretically last for hundreds of years (e.g., optical tape and disc) the life of the physical reading and writing device used to copy and restore data is in the order of three to five years. Over the last thirty years, the geomatic and archival professions have seen many examples of this type of obsolescence. One of the keys to preventing obsolescence is to control the handling and storage of the physical carrier upon which the data are placed and implementing proper data refreshing and migration procedures. As many organizations find out the difficult way, the expense of maintaining proper handling and storage conditions is insignificant when compared to the cost of replacing or attempting to recreate lost data.

The National Archives of Canada provides the following information related to the environmental parameters that should be implemented for the offline storage of data carriers. For instance, the NA is using a temperature guideline of 18 degrees Celsius and a relative humidity range of 40-45% for the storage of computer discs and tape in its archival storage facilities. The NA makes two copies of every archival collection on two different physical supports. Over the medium term, one copy is being stored on 8mm particulate metal tape and the other on DLT4000 cartridge tape. This methodology attempts to diversify the risks associated with relying on only one type of physical carrier for archival storage. Each copy is stored in a separate facility to ensure the information is protected from localized disasters. From an Input/Output perspective, the 'tar' tape management utility, which is compatible with most hardware platforms, is used during the tape reading and data copying processes.

7 PRESERVATION STRATEGIES

The international archival community proposes a range of approaches to overcome technological obsolescence. However, it would appear as though no one solution is available that addresses all

the archival challenges associated with the preservation of digital information. In fact, most archival institutions and government departments use a mixture of evolving approaches, and this same trend is expected to continue into the foreseeable future.

Kyong-Ho Lee has indicated that the most appropriate preservation strategy for any institution must be determined by considering the cost effectiveness of the preservation solution, legal issues, and user access requirements.[16] All preservations strategies must weight the costs and benefits of preserving everything and defining rules of acceptable loss if these strategies prove to be cost restrictive. Some of the following strategies represent the most popular methods that are available to neutralize the interdependencies of technology and data:

- Technology preservation, which is an approach that maintains old hardware and software technologies to facilitate access to old logical and physical storage formats.

- Technology emulation, which is a technique that enables contemporary software and hardware to mimic the functionality and look and feel of historic digital objects and applications.

- Information migration, which allows one to transfer digital information from one logical storage format to another so it can be represented in a current generation of computer hardware and software.

- Information encapsulation, which is a technique that is used to group together digital objects with metadata that provides the necessary information to access those objects at some point in the future.

Table 1 identifies the advantages and disadvantages associated with each preservation approach. From an operational perspective, at some point in time all organizations will migrate their data holdings to diminish the risk of information loss. Most archives and government departments support information migration as the most plausible preservation strategy. In order to reduce obsolescence risks, information migration activities should be incorporated into the daily operations of an organization's business processes, and be based upon corporately accepted logical and information storage standards and practices to ensure success.

16 Lee, Kyong-Ho, et. al. (2002). The State of the Art and Practice in Digital Preservation, *Journal of Research of the National Institute of Standards and Technology*, Volume 107, Number 1, January – February, p. 94.

TECHNIQUE	ADVANTAGES	DISADVANAGES	PRESERVATION PERIOD
Technology Preservation	Information is saved in its native format.	Maintaining original technology is both difficult and costly. Information is only accessible through original hardware and software.	Short-term solution and is impossible to implement over the long-term
Technology Emulation	Preserves an application's original functionality, and same look and feel.	Requires access to original source code to access application and replicate functionality in a new technology environment. Recreating emulated functionality is costly and difficult to achieve.	Short to medium-term in duration.
Information Migration	Original application software does not need to be maintained. There is a small risk of information loss.	There is a risk of altering records during conversion, and a probability of loosing record integrity. Methodology is a continuous process, and therefore costly over time. Administrative metadata must be developed to document the migration process.	Short to medium-term in duration
Information Encapsulation	Preserves both the records and the information about the records. Does not change the original file format.	Considerable time must be invested into the encapsulation format and system development.	Long-term solution.

Table 1: Preservation strategies, their advantages and disadvantages, and preservation periods.

7.1 TEMPORAL MANAGEMENT OF GEOGRAPHIC DATA

The temporal management of geospatial data introduces many challenges. A number of methodologies can be applied including the creation of data base versions, editions, or snapshots. The methodology chosen will depend upon an organization's legislated and operational business requirement to produce and maintain specific geospatial data products, or objects.

Departments in the Canadian federal government gather, analyze, interpret, use and distribute geographic information to support the government's surveying, mapping, charting, resource and infrastructure management activities and programs, and to deliver reliable services. In these activities, there is a requirement for organizations to define their core geospatial data products, develop the business case for their creation, and identify the risks and liabilities associated with their use. The development of the business case will enable organizations to define the requirements to manage specific geographic data objects at a variety of scales, levels of detail and degrees of accuracy. The database management strategy deployed should be based upon the business requirement to create and manage geographic data at the point, line, polygon or the composite geographic object level (i. e., map). The deployment strategy should also enable data management practices at the bit, byte, record, file or other levels of hierarchical detail.

Geospatial data creators and users usually want the most current and accurate information they can obtain. In these situations, data creators regularly supersede what they have on hand and in doing so they lose track of how geographic objects evolve over time. In other situations, creators and users are interested in analyzing temporal changes in geographic phenomena to discern how policy decisions have impacted the contemporary geographic distribution of these phenomena. These different research perspectives require different data management approaches.

One management approach is to use a versioned database model that can be used to symbolize characteristics associated with a piece of geographic data at a historical point in time. Over time, a versioned database represents a collection of manageable states that are discrete snapshots of every geographic phenomenon that is recorded in the database. The database design allows an organization to manage multiple point-in-time spatial representations for different types of geographic data. The design also allows one to investigate complicated time series spatial relationships. The advantage of the versioned database solution is that it enables one to recreate the spatial characteristics of a database's constituent geographic data entities at any one point in time. In a historical sense, it allows one to reproduce edition states on demand.

Another approach is to use the snapshot database model. This model is used to represent multiple geographic phenomena at specifically defined points in time. Snapshots of a database are created at regular intervals, and may be considered as an edition or picture of the database's state at static points in time. Each snapshot is generally used as a public or official view of the database until it is superceded with another. Over time multiple static snapshots may be utilized to analyze temporal changes in the spatial patterns of geographic phenomena.

A number of GIS software companies have created a variety of solutions that aid organizations with the management of geospatial data and promote archiving, preservation and access activities. The archival community is involved in an international effort to develop the Open Archival Information System (OAIS) Reference Model that defines the functional and technical components of systems that are used to manage and preserve digital information over time.[17] The OAIS model originated from an effort to address data management requirements in the space community, and it can be applied to the temporal management of geospatial data.

The United States National Archives and Records Administration (NARA) and the San Diego Supercomputer Center (SDSC) are currently adapting the OAIS reference model in the practical application of an information management architecture that is built around a persistent object preservation approach.[18] The project aims to develop a persistent archive to support the ingestion, archival storage, information discovery, and preservation of digital collections. The ultimate goal of this project is to preserve not only the original data, but the context that permits the data to be interpreted. The Persistent Archive Research Group of SDSC is defining the functional and technical requirements needed to develop the architecture and construct a system that uses data grid and clustering technologies.[19] If the research approach is successful, it will deliver a single technological solution that supports preservation mechanisms to access heterogeneous types of storage systems and information repositories that allow the re-creation of derived data products and objects. The outputs from this research, in combination with research that is being conducted by the International Research on Permanent Authentic Records Electronic Systems (InterPARES) case study team for

---

17 Consultative Committee for Space Data Systems (CCSDS). (2002). *Reference Model for an Open Archival Information System (OAIS)*, Document CCSDS 650.0-B-1, Blue Book, Issue 1, January, 148 pp.

18 Thibodeau, Kenneth. (2001). Building the Archives of the Future: Advances in Preserving Electronic Records at the National Archives and Records Administration, *D-Lib Magazine*, Volume 7, No. 2, February.

19 Moore, R.W., & Mersky, A. (2002). *Persistent Archive Basic Components – Draft 1.0,* Persistent Archive Research Group, Global Grid Forum, July, 30 pp.

the 'CyberCartographic Atlas of Antarctica', may be a suitable solution for the management and preservation of geospatial information objects.[20]

7.2 DOCUMENTATION AND METADATA

No organization can successfully preserve geospatial data with out the proper documentation. Metadata, data about data, provides the information necessary to discover and successfully interchange digital information between geospatial data producers and users. It is a central component of any information management and preservation strategy for geospatial data, especially for creators.

The creation and use of spatially oriented metadata is not a new concept. Map and nautical chart producers have created metadata for centuries to convey important information to users of their maps and charts. Historically, map legends have been created to provide a summarized description of map unit components and symbology, together with other supporting information about a map's title, date, scale, datum, location, data sources, reference numbers, credits and citation information. Librarians and archivists are key producers of metadata descriptions because they enable access and collection management activities.

In a digital world, metadata must be sufficient to support changes made to records through various generations of hardware and software, to support the reconstruction of the decision making process, and to provide audit trails throughout the life cycle of a record. Within the geomatics community, there is a need to develop metadata infrastructure to compliment preservations activities that support an increased requirement to share and distribute reliable geospatial data products between creators and users, at a reduced cost.[21]

The Australia New Zealand Land Information Council (ANZLIC) has indicated that metadata for geospatial information is required for a number of reasons.[22] The most important are:

---

[20] InterPARES is a multinational collaborative research initiative that is attempting to determine the archival requirements to maintain the authenticity of different types of electronic records; identify principles and practices that can be applied to the successful preservation of electronic records; and develop frameworks for preservation policies and standards. The Geomatics and Cartographic Research Centre, Carleton University Department of Geography is currently investigating the theoretical construct of CyberCartography which sees the map as a mechanism for managing digital information. The research team is using the 'CyberCartographic Atlas of Antarctica' as a case study under InterPARES 2, Domain 1, Task Force 2, Working Group 2.1, Records Creation and Maintenance of Scientific Records.

[21] Over the past decade, efforts have been made by Australia, Canada, Great Britain and the United States to develop national standards for geographic information. Today, many of these national efforts are focusing on the development of the International Organization for Standardization (ISO) 19100 suite of standards for geographic information (International Standard for Geographic Information) that are being developed by Technical Committee 211, Geographic Information/Geomatics. With respect to metadata, the Committee has released a draft version of ISO 19115, Geographic Information – Metadata that is intended for the description of geospatial databases and data files. The standard can be adapted to an organization's requirement to describe data holdings at various levels of descriptive detail.

[22] ANZLIC the Spatial Information Council. (2001). *ANZLIC Metadata Guidelines: Core metadata elements for geographic data in Australia and New Zealand, Version 2*, prepared for ANZLIC by the ANZLIC Metadata Working Group, February, 95 pp.

1. To provide detailed information about data content, provenance, collection methods and analysis techniques.

2. To provide information about geographic coverage and data extent, accuracy, quality, scale, resolution and projection.

3. To document information about copyright, intellectual property rights, restrictions on use and sharing, licensing, disclaimers, and liabilities.

4. To provide information about the processing history of a data collection and to document the processes that may result in the change of geographic information as a result of data refreshing, reformatting and migration activities.

5. To provide technical information about the hardware and software requirements that are needed to access and query a set of geospatial data.

Among the organizations that were interviewed for this study, the most common metadata standard being utilized is the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata.[23] A few organizations were interested in using an international descriptive standard for geospatial data. This is encouraging because there are plans to harmonize FGDC with the ISO 19100 suite of standards.

Only one department interviewed did not utilize any type of metadata standard. In those organizations that maintain metadata, descriptions are being created at various levels of detail. For instance, the Mapping Service Branch of Natural Resources Canada maintains separate levels of descriptive detail based on whether the data are available for external use. In all cases, it is the core geographic product that is described whether that product is held in a database, dataset, data warehouse, or other storage environment.

There is also anecdotal evidence to suggest that at the municipal level of government, and in the private sector, there is no adherence to a metadata standard and there is considerable variation in the level and completeness of documentation that accompany geospatial data. The availability of a cataloguing tool or module within GIS software (e.g., such as ESRI's ArcCatalog) should assist in the creation of metadata, but its universal use requires a commitment on the part of data producers to ensure that metadata activities become an integral part of the data creation process.

From the interviews with data producers, it is apparent that the recording of changes to data over time is often not consistently recorded. Agriculture and Agri-Food Canada and NRCan keep a record of the changes that occur to their data. The Canadian Hydrographic Service does as well because it has a legal requirement for the accuracy of its navigational charts. For the other organizations, the primary reason cited for not recording this type of information was the lack of a governance and

[23] The Federal Geographic Data Committee is a 19 member interagency committee composed of representatives from the Executive Office of the President, Cabinet-level and independent agencies. The FGDC is developing the United States National Spatial Data Infrastructure (NSDI) in cooperation with organizations from State, local and tribal governments, the academic community, and the private sector. The NSDI encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data.

accountability structure for data management activities. As a result, almost anyone can make changes to the data.

The following summarizes documentation practices:

1. Four of the eight organizations surveyed record the date of changes or modifications to data.

2. Four of the eight also record the reasons for these changes or modifications.

3. Only a few agencies keep a record of the number of records affected by the modifications.

4. One agency keeps a record of the software and version that was used on a given dataset.

5. No one keeps a record about the type of computer or operating system that was used to manage a data collection.

6. Four organizations keep a record of the origin of their data.

7. Several organizations keep a record about who worked on a series of data at the organizational level.

8. Four of the eight departments interviewed record the purpose, use, and sometimes the usage rules for the data.

8 NATIONAL PROGRAMS AND INITIATIVES

8.1 GEOCONNECTIONS

The GeoConnections program is a Government of Canada funded program to develop the Canadian Geospatial Data Infrastructure (CGDI), with the objective of harmonizing Canada's geospatial databases and making them accessible on the Internet. The impetus for the CGDI came from the Inter-Agency Committee on Geomatics (IACG), a committee created to coordinate geomatic activities in the Canadian federal government.

Through partnerships with federal, provincial and local governments, the private sector and academia, the GeoConnections program is promoting the use of standards and protocols to facilitate access to Canadian geospatial data. Extensive consultation with the Canadian Council on Geomatics (CCOG), which is a federal-provincial consultative committee for geomatics, and the Geomatics Industry Association of Canada (GIAC) is guiding GeoConnections' activities.

Five policy thrusts or themes drive the delivery of GeoConnections' programs and activities: access, framework data, standards, partnerships and policy. The Policy Advisory Node focuses on creating a supportive policy framework to promote the sharing and distribution of data by reviewing and harmonizing existing policies. It has made several recommendations for the government to change

data pricing policies and it is now addressing the need to harmonize distribution and user licensing policy. The Policy Node, through this study, is attempting to provide guidance to data producers on geospatial data archiving and preservation policies.

8.2 NATIONAL DATA ARCHIVES INITIATIVE

Canada currently lacks the necessary archival infrastructure to ensure that all geospatial data created by Canadian society are preserved and made publicly available. The National Archives of Canada has been a leader in addressing archival issues related to preserving geomatic information, but it lacks the capacity to serve the entire geomatics community. In 1996, the Canadian Global Change Program stated that the reasons for this are clear by stating that the National Archives of Canada has neither the funds nor the human resources to fulfill its legislative role to archive all the data that are produced by the Canadian federal government. Nor, does the National Archives of Canada have the infrastructure to provide archival services to other levels of government in Canada.[24] As a result, it has to be very selective in terms of what it does acquire, preserve and make available.

In part, to address this problem, the Social Sciences and Humanities Research Council (SSHRC) and the National Archives of Canada established the National Data Archive Consultation Working Group (NDACWG) to address the need for the Government of Canada to build the archival infrastructure that would be required to archive research data that are created in all sectors of Canadian society.[25] The working group's needs assessment report indicated that many of the building blocks necessary to fulfill this objective are already in place. The missing element is a funded coordination, management and preservation service. The working group has recommended that further study is required to define the characteristics of a National data archive network or service and establish the funding requirements to create such a service.

The establishment of this archival infrastructure would be complementary to the establishment of a Canadian Geospatial Data Infrastructure that has a goal of making Canada's numerous databases of geographic information accessible through the GeoGratis and GeoBase Internet portals. The archival infrastructure will also support GeoConnections' vision to increase access to the use of geospatial data and improve sharing amongst holders of data at all three levels of government. The GeoGratis website offers a variety of national datasets and geodetic products at no cost to researchers, but the data are static and their long-term preservation is questionable.

9 RELATED INTERNATIONAL PROGRAMS AND INITIATIVES

The management of electronic records has become one of the major challenges in the library and archival community. While considerable research has been undertaken, no definitive solution has yet

---

24 The Canadian Global Change Program. (1996). *Data Policy and Barriers to Data Access Across Canada: Issues for Global Change Research*, A discussion paper by the Data and Information Systems Panel, p. 52.

25 National Data Archive Consultation. (2002). *Building Infrastructure for Access to and Preservation of Research Data in Canada*, Report submitted by the NDAC Working Group to the Social Science and Humanities Research Council of Canada and the National Archives of Canada, 67 pp.

been found. Developments in these communities will benefit the geomatics community, as many of the issues are fundamentally the same. The following programs and initiatives are worth highlighting:

- *The Cedars Digital Preservation Project* is an initiative of the Consortium of University Research Libraries in the United Kingdom and Ireland. It has created a series of guides that concentrate on technical approaches for the preservation and access to digital data.

- *Digital Preservation Coalition* was established in 2001 to secure the preservation of digital resources in the United Kingdom and to work internationally to secure the world's global digital memory and knowledge base. Prominent members include the British Library, Public Record Office, Consortium of University Research Libraries, Joint Information Systems Committee of the Higher and Further Education Funding Councils (JISC).

- *The Electronic Resource Preservation and Access NETwork Project (ERPANET)* is an European Union funded initiative that has produced best practice guides, workshop materials and reports on the digital preservation of cultural heritage and scientific objects.

10 SELECTION OF DATA FOR LONG-TERM RETENTION

"In a digital environment, decisions taken at creation and selection have significant implications for preservation. The link between access and preservation is far more explicit than for paper and other traditional materials, as access to a digital resource can be lost within a relatively brief period of time if active steps are not taken to maintain (i.e. preserve) it from the beginning…The brief period during which digital resources will inevitably become inaccessible means that it makes sense to make decisions about selection and preservation simultaneously."[26]

Organizational policies and practices must be implemented for the creation, use, retention, dissemination, preservation, and disposition of geospatial data. All organizations should build a business case for the creation of core geospatial data products. This will help define the enduring value of the information to the business activities of the organization and determine the life cycle management rules that must be applied to its retention and eventual disposition. Jones and Beagrie have developed a decision tree model that can be used to guide collections management activities in organizations.[27]

Accepting organizational responsibility for preservation implies significant costs. An important step in any preservation strategy is to identify mission critical or core datasets that may require

[26] Beagrie, Neil and Maggie Jones; (2001). *The Preservation Management of Digital Material - A Handbook.* London: The British Library, 2001. 139 pp.

[27] Beagrie, Neil and Jones, Maggie. (2001). Ibid. pp. 85.

preservation management activities. Organizations should develop criteria to help them manage their core collections by addressing the following issues:[28]

- Is there a legal or policy requirement for geospatial data retention?
- Are the data critical to the delivery of the institutions' core business functions?
- What were the operational costs to collect and create the data? What would replication cost if the data were lost?
- Are the data unique?
- What are the qualities of the data?
- What is the potential value of the data and what is the probability of them being reused within the institution, or by others?

A positive answer to one or more of the above questions should automatically mean selection for long-term preservation.

11 CONCLUSIONS AND RECOMMENDATIONS

Despite the considerable research that is now underway, it is apparent that there is no single solution available that addresses all the archival challenges associated with the preservation of digital information. There is a need for a mixture of strategies that must be implemented within an effective information management structure. Implementation will be dependent upon an organization's legislative, policy and information management requirements and its ability to invest in long-term preservation activities.

11.1 INFORMATION MANAGEMENT AND GOVERNANCE ISSUES

Geospatial data preservation issues fall within the realm of a national information policy, and a national data management strategy. Working in partnership with the library and archives communities, data producers need to standardize and adopt organizational policies and practices to govern the creation, use, retention, dissemination, preservation, and disposition of geospatial data to ensure its authenticity and integrity for as long as it is required for legislation, departmental statues and other laws and policies. Canada is producing high-quality geospatial data, but the policies and procedures governing its long-term retention remain to be developed.

As the first step in ensuring the long-term preservation and retention of valuable resources, data producers must adopt an information life cycle management approach, which will ensure that their data will be managed proactively from creation to disposition. An efficient and effective geospatial information management system is based on a governance framework that provides accountability

---

[28] Both *The preservation management of digital material – a handbook* and FGDC Historical Data Working Group fact sheet: *Managing historical geospatial data records: a guide for federal agencies* (URL: http://www.fgdc.gov/nara/hdwgfsht.html , Last accessed Feb. 11, 2003) present guidelines/criteria for assessing the value of data.

for data management practices. In government, data producers must work with their archival agencies to identify information that is of historical value and implement procedures and practices to ensure the integrity of the information while it is under their control.

Metadata plays a pivotal role in any preservation strategy because it sets the context for creation and defines the circumstances surrounding an object's use. It is critical that metadata be defined and implemented at creation. It should also be as detailed as possible and be based and applied upon commonly accepted descriptive standards.

All organizations should build a business case for the creation of their core geospatial data products. It is essential that organizations assess the rationale for collecting and managing their data holdings and define the enduring value of these holdings to the delivery of their business activities. The rationale should be used to establish a set of information life cycle management criteria that can be used to establish data value and define retention and disposition rules that meet the operational requirements for the development of geospatial data products and services

Organizations must establish authoritative responsibility centers that empower individuals with the ability to define and apply the information management principles required to ensure the integrity of an organization's geospatial data holdings. A custodianship model is one option that can be used to provide a means to facilitate data management on the behalf of creators and users, and provide continuity in the delivery of a geospatial data infrastructure.

Recommendations:

- The Inter Agency Committee on Geomatics (IACG) should endorse the concept of custodianship. This will require data producing organizations to define their custodianship roles and responsibilities for the management of corporately held geospatial data products and objects. Custodianship should cover responsibilities relating to accountability, reliability and authenticity, metadata and documentation standards, record formats and associated access issues.

- Data producers should create a corporately approved data management strategy and plan that identifies data holdings, and provides standards for data collection, management and preservation practices.

- Data producers should appoint individuals, or establish a committee, to act as the organization's data custodian(s) to work within the organization and ensure that policies and procedures for geospatial data management are defined, understood and implemented.

- Decisions about the value of a data object should be made at the time of creation and should be evaluated and assessed in terms of its operational and preservation requirements.

- Data producers should create a data assessment fact sheet to assist in the identification of valuable data that requires long-term retention.

- A Canadian government metadata policy should be clearly articulated by the GeoConnections Policy Node so it can be used to provide guidance to data producers.

- The metadata policy should identify and promote the use of a metadata standard that can be used to describe geospatial data. The policy should be actively promoted throughout the geomatics community with the provision of fact sheets and information sessions.

11.2 THE PRESERVATION CHALLENGE

The challenge of managing and preserving digital information includes the development of a cost effective preservation strategy that will liberate the data from proprietary file formats. Information migration activities should be incorporated into the daily operations of an organization's business procedures, and be based upon corporately accepted logical formats and information storage standards and practices.

The creation of geospatial data products should be tied to industry standards, including ISO or open source solutions that liberate the data from proprietary formats. The Library and Archives of Canada is developing guidelines that identify computer file types, interchange formats and information standards that it is recommending to facilitate the interoperability of digital information in the Government of Canada.[29]

Considerable interest and experimentation is now underway with the use of XML and GML for the encoding of geographic information, but it is premature to recommend these formats as preservation standards.

Many data structures are concurrently used to manage geospatial data within corporate entities. From a managerial perspective, this is not only undesirable, but also problematic because it requires multiple investments in software, hardware and the development of people skills and knowledge management activities.

Recommendations:

- A working group comprised of representatives from the Library and Archives of Canada, Natural Resources Canada, Statistics Canada, and other interest organizations, should be established under the auspices of the Inter-Agency Committee on Geomatics to conduct test trials using XML and GML as possible formats that can be used for the preservation of geospatial data.

- In consultation with the library and archival communities, data producers should establish and use a minimum set of archival criteria that can help direct and promote preservation practices for geospatial data in the Canadian federal government.

- Data producers should be encouraged to incorporate the Library and Archives of Canada guidelines and best practices for the logical and physical storage of digital information into their data management strategies and plans.

[29] Library and Archives of Canada, (2004). *Guidelines for Computer File Types, Interchange Formats and Information Standards*. Version 1.0, February. Ottawa.: Library and Archives of Canada.

- There should be consultation with preservation specialists before the national profile for the ISO TC/211 19100 suite of standards is finalized, to ensure that preservation elements are included in the profile.

- Data producers should be encouraged to invest in activities that permit the temporal management of geographic objects at their finest level of geographic detail. The Open Archival Information System reference model, and research that is being conducted by the Unites States National Archives and Records Administration and the San Diego Supercomputer Center should be investigated to determine its applicability to the management of geospatial data.

Bibliography

1. ANZLIC the Spatial Information Council. (2001). *ANZLIC Metadata Guidelines: Core metadata elements for geographic data in Australia and New Zealand, Version 2*. February, 95 pp.

2. Beagrie, Neil and Jones, Maggie. (2001). *The Preservation Management of Digital Material - A Handbook*. London: The British Library.

3. Bleakly, Denise R. (2002). *Long-Term Spatial Data Preservation and Archiving: What are the Issues?*, Sand Report, SAND 2002-0107. Albuquerque, New Mexico : Sandia National Laboratories.

4. Brown, David L. & Comeau, Mike. (1999). "Restoration of the Canada Land Data System". *Association of Canadian Map Libraries and Archives Bulletin,* Number 106, Autumn, pp. 42-52.

5. Brown, David L. & Comeau, Mike. (2001). *Restoration of the Canada Land Inventory.* Paper presented at the Social Science History Association Conference, Historical Geography Network, Archives in the Digital Age, Chicago, Illinois, November 15-18.

6. Brown, Richard & Brown, David L. (2001). *Information Management in the Government of Canada: Improving Service Today and Planning for Tomorrow*. Unpublished report. Ottawa: National Archives of Canada.

7. Chen, Su-Shing. (2003). *Digital Library Preservation: Organizational Commitment, Archival Stability, and Technology Continuity*, Department of CISE, Gainesville, FL : University of Florida.

8. Consultative Committee for Space Data Systems (CCSDS). (2002). *Reference Model for an Open Archival Information System (OAIS)*. Document CCSDS 650.0-B-1, Blue Book, Issue 1.

9. Council on Library and Information Resources and Library of Congress, (2002). *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, Washington, DC : CLIR. Retrieved May 2002 from: http://www.clir.org/pubs/abstract/pub106abst.html.

10. Data and Information Systems Panel, Canadian Global Change Program. (1996). *Data Policy and Barriers to Data Access Across Canada: Issues for Global Change Research*. Retrieved February 2002 from: http://www.globalcentres.org/cgcp/english/html_documents/publications/data/toc.htm.

11. FGDC Historical Data Working Group (1998). *Managing historical geospatial data records: a guide for federal agencies* Retrieved February 2002 from website: http://www.fgdc.gov/nara/hdwgfsht.html

12. Hedstrom, Margaret & Montgomery, Sheon. (1998). *Digital Preservation Needs and Requirements in RLG Member Institutions.* Mountain View, CA. *:* Research Libraries Group. Retrieved April 2002 from: http://www.rlg.org/preserv/digpres.html

13. Hedstrom, Margaret. (1995). "Digital Preservation: A Time Bomb for Digital Data". In *U.S. Geological Survey: EROS Data Center Collections Policy Manual.* Index revised 2002 by

Denny Thurman and Diane Krell. Retrieved August 2002 from: http://www.uky.edu/~kiernan/DL/hedstrom.html.

14. Hickling Arthurs Low (HAL) Corporation. (2001). *Geomatics Sector Human Resources Study*. CD-ROM. Otawa : Canadian Council of Land Surveyors Canadian Institute of Geomatics and Geomatics Industry Association of Canada.

15. Lee, Kyong-Ho, et. al. (2002). The State of the Art and Practice in Digital Preservation, *Journal of Research of the National Institute of Standards and Technology*, Volume 107, Number 1.

16. Library and Archives of Canada. (2004). *Guidelines for Computer File Types, Interchange Formats and Information Standards*. Version 1.0, February. Ottawa.: Library and Archives of Canada.

17. McDonald, John. (2000). *Information Management in the Government of Canada - A Situation Analysis*. Unpublished report for the Chief Information Office, Treasury Board Secretariat and the National Archivist. Ottawa : National Archives of Canada.

18. Moore, R.W., & Mersky, A. (2002). *Persistent Archive Basic Components – Draft 1.0,* Persistent Archive Research Group, Global Grid Forum. Retrieved August 2002 from: http://www.sdsc.edu/NARA/Publications/Web/PA-cap.doc.

19. National Archives of Canada. (2003). *Information Management Capacity Check: Methodology and Guide*. Ottawa : National Archives of Canada.

20. National Data Archive Consultation. (2002). *Building Infrastructure for Access to and Preservation of Research Data in Canada*. Unpublished report submitted by the NDAC Working Group to the Social Science and Humanities Research Council of Canada and the National Archives of Canada. Ottawa : NDAC Working Group.

21. Open GIS Consortium. (2001). *Geography Markup Language*. Open GIS Implementation specification. OGC Document Number: 01-029. Retrieved July 2002 from: http://www.opengis.net/gml/01-029/GML2.pdf

22. Parer, Dagmar, ed. (2001). *Archival Legislation for Commonwealth Countries*. Cairns, Australia : Association of Commonwealth Archivists and Records Managers (ACARM).

23. Ross, Seamus. (2000). *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship*. London: National Preservation Office. Retrieved August 2002 from: http://www.bl.uk/services/preservation/occpaper.pdf.

24. Statistics Canada. (2000). *2000 Survey of Service Industries: Surveying and Mapping,* North American Industrial Classification Systems (NAICS) codes 54136 and 54137. Ottawa : Statistics Canada.

25. Thibodeau, Kenneth. (2001). "Building the Archives of the Future: Advances in Preserving Electronic Records at the National Archives and Records Administration". *D-Lib Magazine*, Volume 7, No. 2, February.

26. Treasury Board of Canada Secretariat. (2001). *IM Ready Report*. Version 2, Unpublished report. Ottawa : Chief Information Officer Branch.

27. Tristam, Claire. (2002). "Data Extinction". *Technology Review*, October 2002, 37-42.

28. Zaslavsky, Ilya. (2001). *Archiving Spatial Data: Research Issues*. San Diego Supercomputer Center Technical Report, TR-2001-6. Retrieved June 2002 from: http://www.sdsc.edu/TR/TR-2001-06.doc.pdf