Archiving, Management and Preservation of Geospatial Data

Summary Report and Recommendations

GeoConnections Policy Advisory Node

Working Group on Archiving and Preserving Geospatial Data

David L. Brown: Cartography, Architecture & Geomatics Division, Library and Archives Canada
Grace Welch: University of Ottawa
Christine Cullingworth: Researcher

February 2005

TABLE OF CONTENTS

Executive Summary and Recommendations…… i
1. Introduction…….. 1
2. Background…….. 1
3. Contextual Environment for Preservation Activities…… 2
3.1 Government Policy Environment…….. 2
3.2 Research Environment……... 3
3.3 Institutional Environment…….. 4
4. Data Preservation Issues…….. 5
4.1 Technology Obsolescence…….. 5
4.1.1 Data Representation and Formats…… 5
4.1.2 Storage Technologies…… 6
4.1.3 Technological Obsolescence Solutions…… 6
4.2 Temporal Management of Government Data…….. 8
4.3 Documentation and Metadata……... 9
5. Conclusions and Recommendations……... 9
5.1 Institutional Based Action…….. 10
5.2 National Level Action……. 11
References……... 13

Executive Summary and Recommendations

A working group was created under the GeoConnections Policy Node to identify issues and solutions related to the long-term archiving and preservation of geospatial data. This summary report provides highlights and recommendations, based on a more detailed background study document entitled "The Management and Preservation of 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 archival preservation and continued availability to policy makers, industry and researchers. 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 archiving and preserve geospatial data.

This study concludes that there is no single solution available that addresses all the archival challenges associated with the preservation of digital information. The finding of this study is consistent with other efforts, as noted by Au Yeung (2004) "…the real solution for digital preservation may lie less in technology and more in policy". It is however clear that 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.

Recommendations for further action and research related to the long-term archiving and preservation of geospatial data can be divided between: 1) those activities that are institutionally based; and 2) those that require national policies, standards and guidelines that should be collaboratively developed.

Recommendations for Institutional Activities

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.

Recommendation #1 -- Organizations should define and implement policies and practices for the creation, use, retention, dissemination, preservation, and disposition of geospatial data.

> Suggestion on approach -- This will require building a business case for the creation of core geospatial data products. The business case helps define the enduring value of the information and determines the life cycle management rules that must be applied to its retention and eventual disposition. Beagrie and Jones (2002) have developed a decision tree model that can be used to guide collections management activities in organizations

(http://www.dpconline.org/graphics/handbook/figure4.html). The US Federal Geographic Data Committee (FGDC) Historical Data Working Group (1998) have also published a factsheet to assist in this activity which requires that an organization answer such questions as:

- 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 these questions should automatically mean selection for long-term preservation.

Recommendation # 2 -- 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.

Suggestion on approach -- 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. Under such an approach data producers need to:

- create a corporately approved data management strategy and plan that identifies data holdings, and provides standards for data collection, management and preservation practices;
- appoint an individual, 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;
- create a data assessment fact sheet to assist in the identification of valuable data that requires long-term retention;
- assign the value of a data object at the time of creation based on operational and preservation requirements;
- provide detailed metadata based on internationally and national approved metadata standards. Metadata and data models should be stored with the data;

- inventory existing digital holdings and assess their value for long-term retention and preservation and define a time-table for migration and media refreshing;
- standardize data formats within the organization and widely disseminate this information within the organization;
- adopt industry standards such as ISO or open source solutions that liberate the data from proprietary formats.

Recommendations for National Action

Recommendation # 1 -- 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, government 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.

Suggestion on approach -- Canada is producing high-quality geospatial data, but the policies and procedures governing its long-term retention remain to be developed.

National agencies and bodies such as the federal Inter-Agency Committee on Geomatics (IACG), the Canadian Council on Geomatics, federal agencies such as the Library and Archives Canada, along with national councils and Boards such as the Canadian General Standards Board should work together to develop the following policy areas:

- promote the use of the approved metadata standard in the geomatics community through publicity, workshops, and the creation of tools;
- promote the adoption of non-proprietary standards for the creation and exchange of geospatial data such as GML and Open Geospatial Consortium specifications
- promote the use of Library and Archives Canada guidelines and best practices for the logical and physical storage of digital information;
- develop checklists, guidelines and other practical tools for the selection of data for preservation;
- define and 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;

- investigate and recommend sustainable business models for preservation that promote cooperation to share the responsibility and costs associated with preservation activities
- work with the research communities such as InterPares and others involved in leading-edge research related to preservation issues especially as it relates to geospatial data, in order to incorporate new developments into information management policies and guidelines.

Recommendation # 2 – the Canadian Council on Geomatics should create a task force and invite interested stakeholders to develop priority policy areas identified above.

1. Introduction

A working group was created under the GeoConnections Policy Node to identify issues and solutions related to the long-term archiving and preservation of geospatial data. A detailed report and recommendations was prepared in July 2003. This summary report provides highlights and recommendations based on a more detailed background study document entitled "The Management and Preservation of Geospatial Data".

2. Background

"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". **(Ross, 2000)**

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. (Hickling 2001). 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. 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 archival 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-existent, 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. (Tristam, 2002). 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 along with a private sector company known as Spatialanalysis. (Brown, 1999).

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 archiving and preserve geospatial data

3. Contextual Environment for Preservation Activities

3. 1 Government Policy Environment

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. To ensure that government information is managed effectively and efficiently throughout its life cycle the GOC ratified the 'Management of Government Information Policy' in May 2003. The policy 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 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 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. The *National Archives of Canada Act* and the *National Library Act* were harmonized into one piece of legislation to form the Library and Archives Canada on May 21, 2004.

Other relevant federal legislation which impacts the preservation of geospatial data includes: Copyright Act, Access to Information Act, Canada Evidence Act, Personal Information and Electronic Documents Act, Privacy Act, Statistics Act.

With specific reference to geospatial data at the federal level, the Inter-Agency Committee on Geomatics (IACG), is a senior committee created to coordinate geomatics activities in the Canadian federal government. It has a key role in developing and supporting policies related to the preservation and management of geospatial information. Another important initiative is GeoConnections which 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. 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. The GeoConnections 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, harmonizes distribution and user licensing policy

At the provincial and territorial government level, there is a legislative basis to preserve and archive electronic information. 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 (as of 2003).

Without considerable investigation, it is difficult to determine what policies and procedures exist at the municipal level. Anecdotal evidence suggests however that there are very few activities related to the preservation of geospatial data.

3. 2 Research Environment

Most of the research related to the preservation of the digital information is taking place outside the GIS technology domain. The management of electronic records has become one of the major challenges in the library and archival community and there are a number of projects and programs underway. While considerable research has been undertaken, no definitive solution has yet 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.

- *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.

- *PADI (Preserving Access to Digital Information)* is a comprehensive portal to international digital preservation resources and activities, maintained at the National Library of Australia. PADI aims to facilitate the development of strategies and guidelines for the preservation of access to digital information.

There is a paucity of studies specifically related to the preservation of geospatial data with two exceptions: a study by Bleakley (2002) identified a number of issues related to the long-term preservation of spatial data and a report by Zaslavsky (2001) looked at research issues specific to archiving spatial data.

Despite the lack of definitive solutions, there are a number of themes which emerge in the research literature which are applicable to any preservation activity related to digital information. A recent paper on best practices for museums (Au Yeung, 2004) gives a succinct summary of key points related to digital preservation: preservation must take place at the point of creation; preservation is best accomplished through a distributed approach, through cooperation, adherence to standards and interoperability is essential; preservation metadata is critical; and lastly, many preservation efforts are modeled on the Open Archival Information Systems (OAIS) reference model.

It is unfortunate, that there are very few studies on the costs of preservation and how organizations can support preservation activities. One of the few studies on this topic is a recent report on economic issues associated with preserving digital resources over the long term, especially the question of incentives for preservation (Lavoie, 2003).

3.3 Institutional Environment

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, organizations need to develop information management plans and adopt preservation strategies that are based upon an information life cycle management model. 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. As noted by Au Yeung (2004), "there is also general consensus on digital archiving or preservation as a continuous activity in the form of a services of managed activities…or as a lifecycle management approach".

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.

4. Data Preservation Issues

4.1 Technology Obsolescence

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.

4.1.1 Data Representation and Formats

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. 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.

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.

4.1. 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.

4.1. 3 Technological Obsolescence Solutions

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.

All preservations strategies must weigh 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.

The table below identifies the advantages and disadvantages associated with each preservation approach.

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.

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.

4.2 Temporal Management of Government 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.

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 (Consultative Committee for Space Data Systems, 2002). 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 (Thibodeau, 2001). The project aims to develop a persistent archive to support the ingestion, archival storage, information discovery, and preservation of digital collections. 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 the 'CyberCartographic Atlas of Antarctica', may be a suitable solution for the management and preservation of geospatial information objects.

4.3 Documentation and Metadata

Metadata plays a pivotal role in any long-term preservation strategy. 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 recognized as one of the most important elements of any information management and preservation strategy for geospatial data.

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.

The most common metadata standard being utilized at the federal level is the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata. This standard is being replaced by a new international standard, ISO 19115. Harmonization activities and the development of a North American profile are now underway.

5. 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. As noted by Au Yeung (2004) "…the real solution for digital preservation may lie less in technology and more in policy". It is however clear that 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.

Recommendations for further action and research related to the long-term archiving and preservation of geospatial data can be divided between: 1)those activities that are institutionally based; and 2) those that require national policies, standards and guidelines that should be collaboratively developed between such organizations as the Library and Archives Canada, the federal Inter-Agency Committee on Geomatics, the Canadian Council

on Geomatics along with national councils and Boards such as the Canadian General Standards Board.

5. 1 Recommendations for Institutional-based Action

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.

Recommendation #1 -- Organizations should define and implement policies and practices for the creation, use, retention, dissemination, preservation, and disposition of geospatial data.

> Suggestion on approach -- This will require building a business case for the creation of core geospatial data products. The business case helps 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. Beagrie and Jones (2002) have developed a decision tree model that can be used to guide collections management activities in organizations. The US Federal Geographic Data Committee (FGDC) Historical Data Working Group (1998) have also published a factsheet to assist in this activity which requires that an organization answer such questions as:
>
> - 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 these questions should automatically mean selection for long-term preservation.

Recommendation # 2 --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.

> Suggestion on approach -- 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. Under such an approach data producers need to:
>
> - create a corporately approved data management strategy and plan that identifies data holdings, and provides standards for data collection, management and preservation practices;

- appoint an individual, 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;

- create a data assessment fact sheet to assist in the identification of valuable data that requires long-term retention;

- assign the value of a data object at the time of creation based on operational and preservation requirements;

- provide detailed metadata based on internationally and national approved metadata standards. Metadata and data models should be stored with the data;

- inventory existing digital holdings and assess their value for long-term retention and preservation and define a time-table for migration and media refreshing;

- standardize data formats within the organization and widely disseminate this information within the organization;

- adopt industry standards such as ISO or open source solutions that liberate the data from proprietary formats.

5.2 Recommendations for National Action

Recommendation # 1 -- 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, government 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.

Suggestion on approach -- Canada is producing high-quality geospatial data, but the policies and procedures governing its long-term retention remain to be developed.

National agencies and bodies such as the federal Inter-Agency Committee on Geomatics (IACG), the Canadian Council on Geomatics, federal agencies such as the Library and Archives Canada, along with national councils and Boards such as the Canadian General Standards Board, etc. should work together to develop the following policy areas:

- promote the use of the approved metadata standard in the geomatics community through publicity, workshops, and the creation of tools;

- promote the adoption of non-proprietary standards for the creation and exchange of geospatial data such as GML and Open Geospatial Consortium specifications;

- promote the use of Library and Archives Canada guidelines and best practices for the logical and physical storage of digital information;

- develop checklists, guidelines and other practical tools for the selection of data for preservation;

- define and 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;

- investigate and recommend sustainable business models for preservation that promote cooperation to share the responsibility and costs associated with preservation activities

- work with the research communities such as InterPares and others involved in leading-edge research related to preservation issues especially as it relates to geospatial data, in order to incorporate new developments into information management policies and guidelines.

Recommendation # 3 A final direction for recommendations is that the Canadian Council on Geomatics (the joint federal/provincial/territorial) forum create a task force and invite interested stakeholders from the academic community, private sector and other federal and provincial/territorial agencies that can collaborate to develop priority policy areas identified above.

References

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. Lavoie, Brian F. (2003). The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making. White paper published electronically by OCLC Research. Retrieved July 2004 from: http://www.oclc.org/research/projects/digipres/incentives-dp.pdf

17. 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.

18. 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.

19. 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.

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

21. 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.

22. 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

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

24. 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.

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

26. 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.

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

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

29. 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