TA 2: Final Report: Geospatial Data Archiving and Preservation

Science & Technology Policy Research and Analysis Resource Team

Prepared for:
Natural Resources Canada
GeoConnections Operational Framework Team

March 25, 2011

HAL Ref: 7985

Table of Contents

1. Introduction

This report is written to meet the following objective: conduct research and provide analysis and recommendations on issue of archiving and preserving CGDI's geospatial data assets, including solutions for perpetual access. The report is based primarily on research of available documents and literature, supplemented by consultations with a primary stakeholder, Library and Archives Canada (LAC).

The report is structured in nine chapters, including this introduction. The second chapter provides a synopsis and analysis of some of the more significant legislative requirements for the archiving and preservation of Canada's geospatial data. The third chapter reports on the consultation with Library and Archives Canada concerning their guidelines and best practices that might be used across government. Chapter 4 assesses geospatial considerations in Canada's current information management policy, highlighting key Treasury Board policies and directives. The fifth and sixth chapters report the results of an environmental scan of Canadian consultations and reports and Canadian research on digital data archiving and preservation, respectively. Chapter 7 identifies some best practices based on an environmental scan of geospatial data archiving and preservation in Canada and abroad. The eighth chapter identifies a few business models for archiving and preservation, especially those that promote cooperation and sharing responsibility and costs for preservation activities. The final chapter provides conclusions drawn from the research findings and analysis and recommendations for follow up action.

2. Canada's Legislative Requirements for Information Archiving and Preservation

2.1 Introduction

The initial sections of this chapter provide research findings and analysis of six (6) acts and two (2) regulations that direct how government institutions are to preserve, manage and disseminate Government of Canada information and knowledge, as shown in Table 1. The Library and Archives Act applies to a specific government organization, while the other five Acts discussed here direct how Government of Canada information is to be managed and disseminated irrespective of the creating institution. The two regulations discussed are associated with the Library and Archives of Canada Act and the Privacy Act, respectively. These sections assess how these acts and regulations govern the preservation, management and dissemination of geospatial data, databases and maps which are, irrespective of their form, also government information. These acts and regulations provide the context within which TBS and departmental guidelines, directives, and policies related to the management of government information are created.

Table 1: Acts and Regulations Specific to the Preservation and Management of Government Information

Acts	Regulations
Library and Archives of Canada Act (2004, c. 11)	**Legal Deposit of Publications Regulations** (SOR/2006-337)
Copyright Act (R.S., 1985, c. C-42)	
Access to Information Act (R.S., 1985, c. A-1)	
Privacy Act (R.S., 1985, c. P-21)	**Privacy Regulations** (SOR/83-508)
Personal Information Protection and Electronic Documents Act [2000, c. 5]	
Canada Evidence Act (R.S., 1985, c. C-5)	

Following these initial sections are sections on seventeen (17) acts and five (5) regulations that deal specifically with geospatial information, primarily those under the responsibility of the Minister of Natural Resources[1] (see the complete NRCan list in Appendix C). The *Remote*

[1] Natural Resources Canada, 2010, *List of acts for which the Minister of Natural Resources is responsible*, Accessed March 2011 (http://www.nrcan-rncan.gc.ca/com/resoress/actacte-eng.php)

Sensing Space Systems Act (2005, c. 45) and the *Remote Sensing Space Systems Regulations* (SOR/2007-66) are also included here since they have an impact on how the Canada Centre for Remote Sensing manages its geospatial data sets, and the *Charts and Nautical Publications Regulations (*SOR/95-149) are included because they deal with marine geospatial information. These acts and regulations are listed in Table 2.

Table 2: Acts and Regulations Specific to Geospatial Information

Acts	Regulations
Department of Natural Resources Act (1994, c. 41)	
Resources and Technical Surveys Act (R.S., 1985, c. R-7)	
Canada Lands Surveys Act (R.S., 1985, c. L-6)	
Forestry Act (R.S., 1985, c. F-30)	**Regulations Respecting the Report on the State of Canada's Forests** (SOR/95-479)
Canada Oil and Gas Operations Act (R.S., 1985, c. O-7)	**Canada Oil and Gas Geophysical Operations Regulations** (SOR/96-117)
Canada Petroleum Resources Act (1985, c. 36 (2nd Supp.))	
Canada-Newfoundland Atlantic Accord Implementation Act (1987, c. 3)	
Canada-Nova Scotia Offshore Petroleum Resources Accord Implementation Act (1988, c. 28)	
Energy Efficiency Act (1992, c. 36)	
Energy Monitoring Act (R.S., 1985, c. E-8)	
National Energy Board Act (R.S., 1985, c. N-7)	**National Energy Board Electricity Regulations** (SOR/97-130)
Northern Pipeline Act (R.S., 1985, c. N-26)	
Nuclear Energy Act (R.S., 1985, c. A-16)	
Nuclear Fuel Waste Act (2002, c. 23)	
Nuclear Safety and Control Act (1997, c. 9)	
Uranium Mines and Mills Regulations (SOR/2000-206)	
Remote Sensing Space Systems Act (2005, c. 45)	**Remote Sensing Space Systems Regulations** (SOR/2007-66)
	Charts and Nautical Publications Regulations (SOR/95-149)

These acts and regulations were examined to determine if they referred to archiving and preservation or geospatial data management (e.g., administration, licensing, privacy, cost

recovery), and to assess if there are specific clauses on the creation of geospatial data, databases and maps to fulfill the requirements of the acts and regulations. This is considered important, as these may be considered government records - that must be managed and maintained, which would probably result in their being appraised as having high historical and archival value since these are associated with the statutes of Canada. Appraisal, which is a key element in the process of archiving and should be considered as part of a record's life-cycle, is defined as: "the process of determining the value and thus the disposition of records based upon their current administrative, legal, and fiscal use; their evidential and information value; their arrangement and condition; their intrinsic value; and their relationship to other records"[2]. Geospatial data, databases and maps mentioned in or associated with these acts and regulations would also be considered critical to the operationalization of these acts and regulations.

The Government of Canada produces geospatial data, databases and maps in many other departments and agencies and each of these would have to adhere to acts and regulations for which they are responsible. It is beyond the scope of this E-Scan to explore them all; however, as will be seen in the analysis below, many of these acts and regulations can be very specific regarding geospatial data, dataset and map form, format, use, viewing, recording, distribution, licensing, etc. In some instances there is mention of raw data, how these are to be processed, the systems used to view and process the data and the use of cryptography. There are also mentions of registrars and catalogs. In some cases acts and regulations make specific reference to archiving, deposition, data protection plans, the number of years data are to be kept and data curation. Finally, because these geospatial datasets are critical to the management of Canada's boundaries, the health and safety of Canadians, Canada's economy and the attestation of lands, these government records can all be called into evidence to resolve international and national disputes, or any other judicial proceeding. Clearly, the geospatial datasets mentioned here must be well managed once they are created and should be preserved accordingly.

2.2 Library and Archives of Canada Act[3]

The *Library and Archives of Canada Act* (2004, c. 11), herein referred to as the *LAC Act*, stipulates that: "the documentary heritage of Canada be preserved for the benefit of present and future generations"[4]; there be an institution to service and to ensure that Canadian knowledge is accessible to all; that such institution facilitate "cooperation among the communities involved in the acquisition, preservation and diffusion of knowledge" and finally, that it serve as the memory of the Government of Canada. In addition, the *LAC Act* states that the Minister may "establish an Advisory Council to advise the Librarian and Archivist with regard to making the

[2] Appraisal: [Archives - A Glossary of Archival and Records Terminology (The Society of American Archivists)], InterPARES 2 Terminology Database (http://www.interpares.org/ip2/ip2_terminology_db.cfm)

[3] Justice Canada, Current to February 26, 2011, *Library and Archives of Canada Act (2004, c. 11)*, Accessed March 2011 from (http://laws.justice.gc.ca/PDF/Statute/L/L-7.7.pdf).

[4] LAC Act, Preamble p.1

documentary heritage known to Canadians and to anyone with an interest in Canada and facilitating access to it"[5]. The objects[6] of LAC, among others, are to "facilitate the management of information by government institutions"; "coordinate the library services of government institutions"; and "support the development of the library and archival communities". The Librarian and Archivist of Canada has a number of powers[7], the following of which are a subset that specifically relate to the objectives of this environmental scan (E-Scan):

- acquire publications and records or obtain the care, custody or control of them;
- take measures to catalogue, classify, identify, preserve and restore publications and records;
- enter into agreements with other libraries, archives or institutions in and outside Canada;
- advise government institutions concerning the management of information produced or used by them and provide services for that purpose;
- provide leadership and direction for library services of government institutions;
- provide professional, technical and financial support to those involved in the preservation and promotion of the documentary heritage and in providing access to it; and in addition to these powers LAC; and

In addition, LAC may take a representative sample of the documentary material of interest to Canada that is accessible to the public without restriction through the Internet or any similar medium.

The following definitions in the *Library and Archives of Canada Act* are relevant for geospatial information[8]:

- **Documentary heritage**: publications and records of interest to Canada.
- **Government record**: a record that is under the control of a government institution.
- **Ministerial record**: a record of a member of the Queen's Privy Council for Canada who holds the office of a minister and that pertains to that office, other than a record that is of a personal or political nature or that is a government record.
- **Publication**: any library matter that is made available in multiple copies or at multiple locations, whether without charge or otherwise, to the public generally or to qualifying members of the public by subscription or otherwise. Publications may be made available through any medium and may be in any form, including printed material, on-line items or recordings.

5 LAC Act, Establishment of Advisory Council, p.3
6 LAC Act, Objects, p.3
7 LAC Act, Powers of Librarian and Archivist, pp. 3-4
8 LAC Act, Definitions, pp. 1-2, and also p. 6 for recording.

- **Record**: any documentary material other than a publication, regardless of medium or form.
- **Recording**: anything that requires a machine in order to use its content, whether sounds, images or other information.

The *LAC Act* makes no specific reference to form, format or media, which means that geospatial data, databases and maps are considered to be government records, that may have documentary heritage qualities. Record creators are therefore responsible for managing this content in the event that these records are to be accessioned by LAC. LAC's responsibility is to advise, provide leadership, technical expertise, professional expertise and financial support to assist NRCan and other federal geospatial information producers with preserving their records.

With respect to the disposal of records, "no government or ministerial record, whether or not it is surplus property of a government institution, shall be disposed of, including by being destroyed, without the written consent of the Librarian and Archivist or of a person to whom the Librarian and Archivist has, in writing, delegated the power to give such consents[9]". Confidences of the Queen's Privy Council and other records as stipulated in 69 (1) Canada Access to Information Act[10] require that the Librarian and Archivist request special permissions to accession and disseminate these, otherwise he or she "has a right of access to any record to whose disposition he or she has been asked to consent" and any officer or employee of a government institution may grant that access. Also, if records are under any security requirements, the usual oaths of secrecy are required to be taken and only persons who would normally have access to these can access and view those records from the Archives. Such records could, in some cases, include geospatial information.

If records are at risk of serious damage or destruction, the Librarian and Archivist may require their transfer in the manner and at the time that the Librarian and Archivist specifies and he or she "shall have the care and control of all records of a government institution whose functions have ceased[11]" and "all publications that have become surplus to the requirements of any government institution shall be placed in the care or control of the Librarian and Archivist[12]". Furthermore, within six months after the completion of any data collection done for the purposes of public opinion research, that data should be sent to the Librarian and Archivist.[13]

Government publishers are to make legal deposit, at their own expense, of two copies of all publications with the Librarian and Archivist. This includes every version, edition or form of that publication. The *Legal Deposit of Publications Regulations*[14] provides direction on how to

9 LAC Act, Destruction or disposal, p.4
10 Justice Canada, Current to February 26, 2011, *Access to Information Act (R.S., 1985, c. A-1)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/A/A-1.pdf)
11 LAC Act, Government records at risk, p.7
12 LAC Act, Surplus publications, p.8
13 LAC Act, Public opinion research, pp.7-8
14 Justice Canada, Current to February 26, 2011, *Legal Deposit of Publications Regulations (SOR/2006-337)*, Accessed March 2011, (http://laws.justice.gc.ca/PDF/Regulation/S/SOR-2006-337.pdf).

submit digital publications which include, relevant to geospatial information: websites, including portals, personal websites, service sites, intranets and websites consisting primarily of links to other sites; multimedia publications consisting of two or more physical parts in different formats; and dynamic databases and raw data. In addition, a copy of a public recording in the form and quality that is considered of archival quality is to be deposited at LAC if it is considered to be of historical or archival value. These deposited publications and recordings would then belong to Her Majesty and form part of the collection of the Library and Archives of Canada.

A digital geospatial map, dataset and database can be a record, a government record, a recording, a publication, and can also be of documentary heritage as defined in the LAC Act. How these maps, datasets and databases are classified, categorized or labeled is contingent upon the record creating organization's legislative, policy and information management requirements (see for example NRCan Acts and Regulations findings and analysis). For instance, the Atlas of Canada may be a publication and, or a recording and both could be considered to be of documentary heritage. NRCan, the record creator, could appraise[15] the Atlas as being of documentary heritage, as could Atlas stakeholders (e.g., school teachers), or the Librarian and Archivist may decide that is the case and direct creators to deposit a copy at LAC. How and when to capture snapshots of the Atlas, whether or not that is the best method to adopt, and when to accession[16] it in the Archive would have to be collaboratively determined between LAC and NRCan.

A digital geospatial map, dataset or database can also be a ministerial record if it formed part of a particular decision in that office. These digital objects would then be part of that particular ministerial record set and could be subject to restricted access based on the *Access to Information Act*, particularly if part of a ministerial briefing or a memorandum to cabinet or if they affect national security (see analysis of *Access to Information Act* in Section 2.4).

A record-creating institution (e.g., Mapping Information Branch, CCRS, DFO, etc.) could for example, define what it constitutes a record based on how that record is acted upon, set aside for action or is referenced in the institution. Records are created during a particular institution's business process (e.g., creation of a database and an accompanying map of Canada's oil and gas exploration). A record is considered stable when its form is fixed and it is set aside. This is problematic in the context of many databases and maps which are continuously being updated.

[15] n., The process of assessing the value of records for the purpose of determining the length and conditions of their preservation. (InterPARES 2 Terminology Database - http://www.interpares.org/ip2/ip2_terminology_db.cfm)

[16] v., To take legal and physical custody of a body of records and to document it in a register. (InterPARES 2 Terminology Database - http://www.interpares.org/ip2/ip2_terminology_db.cfm)

These digital objects could potentially remain active[17] throughout their in-determinant life-cycle.

Also, some digital objects are created by many record-creating institutions, whose business practices may be different than those of the Government of Canada even though the assembly and dissemination of the dataset is within the purview of the Government of Canada. Framework datasets, for instance, are the products of provincial and territorial jurisdictions but are disseminated and assembled by the Federal Government under the Geomatics Accord. The same is true for distributed maps, where the maps are rendered by a host institution while geospatial data used to create that map are housed in a variety of institutions and made accessible via open standards and specifications. In that case, the software to render the map may need to be preserved, and the institutions that share their data may need to ensure that these are maintained in such a way that they can continuously be rendered. Should a distributed map be accessioned by an Archive, all of the data and the software may be required and agreements with multiple institutions may need to be negotiated.

The determination of what would be considered suitable for archival purposes, or how to appraise geospatial data, databases and maps should be part of a larger records management process, as discussed in Chapter 4. All government institutions are to manage their records in such a way that they could seamlessly be accessioned into the Archive. According to the *LAC Act*, LAC is mandated to work collaboratively with government institutions to help them identify digital maps, datasets and databases to be archived. Institutions must identify information that has ongoing business value and decide what parts of this information must be preserved. Archiving can be done by the record-creating institution (e.g., distributed archiving model) or at LAC, and how this is done must be collaboratively determined.

If a program or activity is terminated, the associated geospatial data, datasets and maps cannot be destroyed as Government institutions do not have the right to dispose of records even if a department is changing direction or made redundant. The TBS *Multi-Institutional Disposition Authority* (MIDA) provides guidelines to institutions in the event that a program ends. LAC would collaborate with the geospatial information producer on the appraisal and accessioning of those records and would determine how to maintain them for preservation purposes. Data producers or the Librarian and Archivist may have identified records that are at risk of deterioration, and if so, LAC is mandated to assist with their preservation (e.g., early digital editions of the Atlas of Canada could have been identified as being at risk). LAC already works with NRCan, in particular the National Air Photo Library (NAPL) and the GeoGratis

[17] Active Record: n., Records needed by the creator for the purpose of carrying out the actions for which they were created or for frequent reference. (InterPARES 2 Terminology Database - http://www.interpares.org/ip2/ip2_terminology_db.cfm)

Program, where LAC maintains and owns the original copies of the Canada Land Inventory[18] and associated data, which are disseminated by GeoGratis[19].

LAC may take a representative sample of documentary material of interest to Canada that is accessible to the public without restriction through the Internet or any similar medium. This has been done by LAC and some useful geospatial information and data can be accessed through the now defunct domain name web crawl conducted by LAC, also known as the Government of Canada Web Archive[20]. NRCan appears in the department list of this web archive[21]; the GeoConnections program does not appear in a list but is part of the NRCan web crawls[22]. In a recent test, the *Atlas of Canada* was crawled four times and the interactive part of the Atlas was not available, while it was possible to access some of the static maps. The GeoGratis site was available but the GeoGratis data collection was not searchable[23]. While helpful, these Internet web crawls clearly are not to be considered as suitable preservation solutions for dynamic or interactive geospatial data, databases and maps and cannot be relied upon as a sustainable means to preserve Government of Canada websites. GeoConnections reports and publications are accessible from this web archive; however, it is important to note that all GeoConnections documents published in paper, or PDF form on the Internet, must be deposited with LAC.

The 2005 GeoConnections study on *Archiving, Management and Preservation of Geospatial Data* indicated that "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)". Also, "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"[24]. It is beyond the scope of this study to investigate these

[18] CLI in LAC catalogue: http://collectionscanada.gc.ca/pam_archives/index.php?fuseaction=genitem.displayItem&lang=eng&rec_nbr=133094&rec_nbr_list=133094,3691606,823187,213916,1723948,1785071,825107,2310705,2310704,3683452

[19] GeoGratis CLI Search Results: http://www.geogratis.cgdi.gc.ca/geogratis/en/collection/search.do?search=keyword&keyword=canada+land+inventory

[20] LAC, *Government of Canada Web Archive*, Accessed February 2011 (http://www.collectionscanada.gc.ca/webarchives/index-e.html)

[21] LAC, *Government of Canada Web Archive Department List*, Accessed March 2011, http://www.collectionscanada.gc.ca/webarchives/department-list/index-e.html.

[22] LAC, *Government of Canada Web Archive, NRCan web crawls*, Accessed March 2011, (http://www.collectionscanada.gc.ca/webarchives/*/http://www.nrcan-rncan.gc.ca)

[23] LAC, *Government of Canada Web Archive GeoGratis web crawl*, Accessed March 2011 (http://www.collectionscanada.gc.ca/webarchives/20071116015613/http://geogratis.cgdi.gc.ca/geogratis/en/index.html).

[24] GeoConnections, 2005, *Archiving, Management and Preservation of Geospatial Data: Summary Report and Recommendations*, Produced by David L. Brown, Grace Welch and Christine Cullingworth as an outcome of the Policy Advisory Node.

statements; however, as the 2005 study recommended, the CCOG could take a proactive role in the development of a collaborative approach to preserving Canada's geospatial knowledge and heritage.

2.3 Copyright Act[25]

When archiving geospatial data, databases and maps, the *Copyright Act* (R.S., 1985, c. C-42) must also be taken into consideration, as would licenses and access rights. Geospatial data, databases and maps may fall within the following definitions in the Copyright Act:

- **Artistic work**: includes paintings, drawings, maps, charts, plans, photographs, engravings, sculptures, works of artistic craftsmanship, architectural works, and compilations of artistic works;
- **Book:** which is a volume or a part or division of a volume, in printed form, but does not include (a) a pamphlet, (b) a newspaper, review, magazine or other periodical, (c) a map, chart, plan or sheet music where the map, chart, plan or sheet music is separately published, and (d) an instruction or repair manual that accompanies a product or that is supplied as an accessory to a service;
- **Collective Work**: which is (a) an encyclopedia, dictionary, year book or similar work, (b) a newspaper, review, magazine or similar periodical, and (c) any work written in distinct parts by different authors, or in which works or parts of works of different authors are incorporated;
- **Compilation:** (a) a work resulting from the selection or arrangement of literary, dramatic, musical or artistic works or of parts thereof, or (b) a work resulting from the selection or arrangement of data;
- **Computer Program:** which is a set of instructions or statements, expressed, fixed, embodied or stored in any manner, that is to be used directly or indirectly in a computer in order to bring about a specific result;
- **Every original literary, dramatic, musical and artistic work:** includes every original production in the literary, scientific or artistic domain, whatever may be the mode or form of its expression, such as compilations, books, pamphlets and other writings, lectures, dramatic or dramatico-musical works, musical works, translations, illustrations, sketches and plastic works relative to geography, topography, architecture or science;
- **Literary work:** includes tables, computer programs, and compilations of literary works;

(http://www.geoconnections.org/publications/policyDocs/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf) p. 3

[25] Justice Canada, Current to February 26, 2011, *Copyright Act (R.S., 1985, c. C-42)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/C/C-42.pdf).

- **Definition of "publication":** 2.2 (1) For the purposes of this Act, "publication" means (a) in relation to works, (i) making copies of a work available to the public

If any data, databases or maps are produced by an employee of the Federal Government, then copyright belongs to Her Majesty. Works created by the Federal Government of Canada can be licensed and NRCan's and other federal producers' geospatial data, databases and maps are covered by many licenses. However, there are no universal licenses for geospatial or any other data being used within the Government of Canada. The *Dissemination of Government Geographic Data in Canada: Guide to Best Practices*[26] includes four types of government geographic data licensing models most commonly used in Canada: the unrestricted use model, the end-user model, the reseller model and the value-added reseller model.

The *Guide to Best Practices*

> provides a rationale for appropriate uses, explains how each model builds on common structures, demonstrates their inter-relationships and provides clear guidance to assist licensing practitioners in selecting the most appropriate model and license agreement. Recommended approaches to fundamental concepts such as ownership of intellectual property, liability, duration and termination are discussed in detail for the benefit of licensing practitioners, and are guided by data dissemination policy directives currently in force across federal departments and agencies.

A useful critique and analysis of these licenses can be found in a paper authored by Elizabeth Judge and Teresa Scassa entitled *Intellectual property and the licensing of Canadian government geospatial data: an examination of GeoConnections' recommendations for best practices and template licenses*[27] published in the Canadian Geographer. Other jurisdictions that also have Crown Copyright are adopting international interoperable licenses such Creative Commons and Open Data Commons licenses.[28]

The form of a geospatial database, database or map will define the nature of the work, albeit, as indicated by Scassa, intellectual property rights as these pertain to geospatial data, databases and maps can be uncertain.[29] For example, a distributed electronic atlas may be considered a compilation, a collected work and in some sense a computer program. The individual datasets used to assemble that atlas may also be considered compilations, with the means to assemble them computer programs. Also, a database may contain data produced by a number of scientists for illustrative purposes, or may be an assembly of data derived from multiple sensors. In that case, the database may be considered a compilation of facts, and the assembly

---

[26] GeoConnections, 2008, Dissemination of Government Geographic Data in Canada: Guide to Best Practices, Accessed March 2011, (http://www.geoconnections.org/publications/Best_practices_guide/Guide_to_Best_Practices_Summer_2008_Final_EN.pdf)

[27] Elizabeth Judge and Teresa Scassa, 2010. "Intellectual property and the licensing of Canadian government geospatial data: an examination of GeoConnections' recommendations for best practices and template licenses", *The Canadian Geographer* 54, no 3, pp.366–374

[28] Teresa Scassa, 2011. *Final Report: Review of IP Law and Instruments (Copyright, Licensing) in the Context of Geospatial Data*, prepared by Hickling Arthurs Low Corporation for Natural Resources Canada

[29] Scassa *supra* note 28

of those facts into a unique arrangement may be carried out within a context of multiple agreements and/or technological arrangements. If the geospatial data, irrespective of form, are produced by the Federal Government these would all fall under Crown Copyright and the TBS has procedures in place to administer and license that content[30].

As previously discussed, NRCan and other geospatial data producing agencies (e.g., DFO, Environment Canada, Statistics Canada, etc.) create their own data licenses which may or may not be legally interoperable and easy for licensees to manage or adhere to, particularly in a distributed mapping context. In addition, should the Government of Canada procure data, the provisions of the TBS *Policy on Title to Intellectual Property Arising under Crown Procurement Contracts*[31] will apply. It is beyond the scope of this E-Scan to be definitive on matters pertaining to copyright and IP. However, it is important to recognize that these still apply when geospatial data are archived, irrespective of their form, and the means to manage these rights and responsibilities need to be considered when data are accessioned and re-disseminated by the archive. In the case of many geospatial data, datasets and databases, metadata help with this process as authorship, rights, ownership, date stamps and use limitations are clearly indicated, particularly if the TBS *Standard on Geospatial Data*[32] has been used.

2.4 Access to Information Act[33]

The purpose of the *Access to Information Act* (R.S., 1985, c. A-1) is to

> extend the present laws of Canada to provide a right of access to information in records under the control of a government institution in accordance with the principles that government information should be available to the public, that necessary exceptions to the right of access should be limited and specific and that decisions on the disclosure of government information should be reviewed independently of government.

This act is intended to complement and not replace existing procedures for access to government information and is not intended to limit in any way access to the type of government information that is normally available to the general public.

It is the responsibility of the Government of Canada to make records available to Canadians, and access is subject to regulations. Also, the Minister should on a yearly basis provide "a

[30] TBS, 2008. *Procedures for the Administration and Licensing of Crown Copyright*, Accessed March 2011 (http://www.tbs-sct.gc.ca/pubs_pol/sipubs/comm/cpgcp-ppcgc04-eng.asp#_Toc137523383).

[31] TBS, 2000, *Policy on Title to Intellectual Property Arising Under Crown Procurement Contracts*, Accessed March 2011 (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=13697&section=text#cha3).

[32] TBS, 2009. *Standard on Geospatial Data*, Accessed March 2011 (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16553&section=text).

[33] Justice Canada, Current to February 26, 2011, Access to Information Act (R.S., 1985, c. A-1), Accessed March 2011, (http://laws.justice.gc.ca/PDF/Statute/A/A-1.pdf)

description of all classes of records under the control of each government institution in sufficient detail to facilitate the exercise of the right of access under this Act" and "the title and address of the appropriate officer for each government institution to whom requests for access to records under this Act should be sent". A subset of geospatial data can be found in data discovery portals or catalogues and can be discovered in metadata. These may be considered suitable dataset descriptions of this class of records. It is understood that most of Canada's geospatial data assets are not registered into catalogues and portals either in their departments or in a centralized location.

The *Act* includes a range of circumstances under which a Government institution may refuse to share information, such as where:

- the conduct by the Government of Canada of federal-provincial affairs could be potentially injured
- the conduct of international affairs, the defense of Canada or any state allied or associated with Canada or the detection, prevention or suppression of subversive or hostile activities could potentially be injured
- the information was obtained in confidence from the government of another jurisdiction
- the information was obtained in the course of lawful investigations (e.g., police investigations, and ongoing investigations by the Auditor General of Canada; the Commissioner of Official Languages for Canada; the Information Commissioner; and the Privacy Commissioner, Chief Electoral Officer, or Public Sector Integrity Commissioner)
- the disclosure of information could threaten the safety of an individual
- the information affects the strategic economic position of Canada
- the information contains trade secrets or financial, commercial, scientific or technical information that belongs to, and has consistently been treated as confidential by: (a) the Canada Post Corporation; (b) Export Development Canada; (c) the Public Sector Pension Investment Board; or (d) VIA Rail Canada Inc.
- a record contains personal information as defined in section 3 of the Privacy Act
- the information concerns Cabinet Confidences or advice to a minister

The *Act* does not apply to "confidences of the Queen's Privy Council for Canada". In addition, "where a certificate under section 38.13 of the *Canada Evidence Act* prohibiting the disclosure of information contained in a record is issued before a complaint is filed under this Act in respect of a request for access to that information", the *Act* does not apply to that information. Also, any restricted information cannot be disclosed. For example, as it pertains to information about the environment, the head of a government institution shall not refuse to disclose a part of a record if that part contains the results of product or environmental testing carried out by or on behalf of a government institution unless the testing was done as a service to a person, a

group of persons or an organization other than a government institution and for a fee. However, disclosure is authorized if

> (a) the disclosure would be in the public interest as it relates to public health, public safety or protection of the environment; and
>
> (b) the public interest in disclosure clearly outweighs in importance any financial loss or gain to a third party, any prejudice to the security of its structures, networks or systems, any prejudice to its competitive position or any interference with its contractual or other negotiations.

There may be cases where some geospatial data, databases and maps cannot be publicly disclosed as outlined in the *Access to Information Act*. For example, these may be geospatial data used as part of a cabinet submission, data produced by National Defence or data used in ongoing investigations. Irrespective of the reasons precluding their disclosure, when data are archived the same access rules apply with the same restrictions, rights, sensitivities and time limitations. In essence the same information management and access procedures are simply transferred to the archive and it is presumed that there are practices in place to ensure that the *Access to Information Act* is adhered to when geospatial data are created. The same care and guidelines used to manage records while they are active apply when data are accessioned into an archive at LAC, where technological measures to protect these data are in place, and will need to be created by geospatial data creators if they archive their own data in a distributed archive context.

2.5 Privacy Act[34]

The purpose of the *Privacy Act* (R.S., 1985, c. P-21) is to

> extend the present laws of Canada that protect the privacy of individuals with respect to personal information about themselves held by a government institution and that provide individuals with a right of access to that information.

Personal information under the *Privacy Act* is information that identifies an individual that is recorded in any form including the following categories that could relate to geospatial information:

> (c) any identifying number, symbol or other particular assigned to the individual,
>
> (d) the address, fingerprints or blood type of the individual,
>
> (i) the name of the individual where it appears with other personal information relating to the individual or where the disclosure of the name itself would reveal information about the individual.

According to the *Privacy Act*, personal information should not be collected by a government institution unless it relates directly to an operating program or activity of the institution; and wherever possible, a government institution should only be collecting personal information

[34] Justice Canada, Current to February 26, 2011, *Privacy Act (R.S., 1985, c. P-21)*, accessed March 2011.

intended to be used for an administrative purpose directly from the individual to whom it relates. Personal information is also to be collected with informed consent. The *Privacy Act* stipulates how data should be maintained, and the quality of the data along with deposition rules. In addition, personal information is only to be used for the purposes that it was initially collected and strict rules apply on how personal information can and cannot be disclosed.

Personal information may be disclosed to Library and Archives Canada for archival purposes and LAC would have to manage and disseminate those same private data in accordance with the *Privacy Act:*

> Subject to any other Act of Parliament, personal information under the custody or control of the Library and Archives of Canada that has been transferred there by a government institution for historical or archival purposes may be disclosed in accordance with the regulations to any person or body for research or statistical purposes.

The *Privacy Act* states when and where personal information can be disclosed. Ministers are mandated in the *Privacy Act* to provide an index of databases or data banks that contain personal information and classes of private information not contained in those banks or databases. The *Act* also stipulates the rights individuals have over their own information, and similar restrictions apply to the disclosure of data as outlined in the *Access to Information Act* above, as these pertain to security, safety, ongoing investigations, confidences to the Privy Council, etc. in addition to privacy regarding medical information and data. The *Privacy Regulations*[35] provide detailed information on record retention, access procedures, disclosure of information as it pertains to mental health data, and management of data when in the Archive:

> 6. Personal information that has been transferred to the control of the Library and Archives of Canada by a government institution for archival or historical purposes may be disclosed to any person or body for research or statistical purposes where
>
> (a) the information is of such a nature that disclosure would not constitute an unwarranted invasion of the privacy of the individual to whom the information relates;
>
> (b) the disclosure is in accordance with paragraph 8(2)(j) or (k) of the Act;
>
> (c) 110 years have elapsed following the birth of the individual to whom the information relates; or
>
> (d) in cases where the information was obtained through the taking of a census or survey, 92 years have elapsed following the census or survey containing the information.

Personal and private information is frequently found in research and administrative databases in non-anonymized form, and can also be found in geospatial databases in anonymized form and/or in geographically aggregated form. In an era where there is growing demand for government administrative data to be combined with demographic data such as in the example of public health, it becomes critical to ensure that the personal privacy of individuals is maintained. TBS records management guidelines and directives (see Chapter 4) already provide

[35] Justice Canada, Current to February 26, 2011, *Privacy Regulations (SOR/83-508)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Regulation/S/SOR-83-508.pdf).

a framework for the management of geospatial data that may include private information. What is not yet well understood is how the cross-matching and layering of myriad geospatial data derived from many institutions could lead to the identification of individuals. GeoConnections has done some research in this area, and has developed some guidelines on how to map and aggregate personal data and how to maintain privacy while doing so[36,37].

Also, like the *Access to Information Act*, it is clear that the *Privacy Act*, *Privacy Regulations* and *Personal Information Protection and Electronic Documents Act* (discussed below) apply when it comes to the preservation and archiving of geospatial data. In addition, it is presumed that mechanisms are already in place with regards to the management of private data irrespective of form, and that metadata to describe these privacy issues and access rights are in place.

2.6 Personal Information Protection and Electronic Documents Act[38]

In addition to the *Privacy Act* and the *Privacy Regulations*, there is the *Personal Information Protection and Electronic Documents Act* (2000, c. 5), which states that one purpose of the act is:

> to establish, in an era in which technology increasingly facilitates the circulation and exchange of information, rules to govern the collection, use and disclosure of personal information in a manner that recognizes the right of privacy of individuals with respect to their personal information and the need of organizations to collect, use or disclose personal information for purposes that a reasonable person would consider appropriate in the circumstances.

The *Personal Information Protection and Electronic Documents Act* has the following provisions for when information can be collected without consent:

> (a) the collection is clearly in the interests of the individual and consent cannot be obtained in a timely way;
>
> (b) it is reasonable to expect that the collection with the knowledge or consent of the individual would compromise the availability or the accuracy of the information and the collection is reasonable for purposes related to investigating a breach of an agreement or a contravention of the laws of Canada or a province;
>
> (c) the collection is solely for journalistic, artistic or literary purposes;
>
> (d) the information is publicly available and is specified by the regulations; or

---

36 GeoConnections, 2010. *Geospatial Privacy Awareness and Risk Management Guide for Federal Agencies*

37 GeoConnections, 2010. *A Manager's Guide to Public Health Geomatics*, AMEC Earth & Environmental a division of AMEC Americas Limited, Accessed March 2011, (http://www.geoconnections.org/publications/Key_documents/ManagerGuide_PubHealthGeomatics_EN.pdf).

38 Justice Canada, Current to February 26, 2011, *Personal Information Protection and Electronic Documents Act [2000, c. 5]*, Accessed March 2011, (http://laws.justice.gc.ca/PDF/Statute/P/P-8.6.pdf)

> (e) the collection is made for the purpose of making a disclosure

and when information can be used without consent such as in the case of an emergency, for statistical or scholarly study or research, etc.

The *Act* also states that a minister can use electronic means to create, collect, receive, store, transfer, distribute, publish or otherwise deal with documents or information whenever a federal law does not specify the manner of doing so. This covers electronic filing, electronic statutes, documents as evidence of proof, seals of authenticity, digital signatures, statements made under oath, and Section 5 of this Act provides an excellent set of clearly articulated principles set out in the National Standard of Canada Entitled Model Code for the Protection of Personal Information.

2.7 Canada Evidence Act[39]

The *Canada Evidence Act* (R.S., 1985, c. C-5) provides clear statements on the provision of documentary evidence irrespective of a record's form (e.g., geospatial data, database or map). Of note is how to assess the authenticity, integrity and the certification of the digital records being called into evidence. For instance,

> 31.1 Any person seeking to admit an electronic document as evidence has the burden of proving its authenticity by evidence capable of supporting a finding that the electronic document is that which it is purported to be.
>
> 31.2 (1) The best evidence rule in respect of an electronic document is satisfied
>
> (a) on proof of the integrity of the electronic documents system by or in which the electronic document was recorded or stored; or
>
> (b) if an evidentiary presumption established under section 31.4 applies.

Assessing the authenticity of digital records, irrespective of form, is not a straightforward process and this should not to be taken lightly, particularly if a geospatial dataset, database or map is called into evidence in a judicial proceeding. The University of British Columbia based International Research on Permanent Authentic Records in Electronic Systems (InterPARES)[40] research project investigated the issue of authenticity as it pertained to scientific and government records[41], and created guidelines and benchmarks. It is important to note that

[39] Justice Canada, Current to February 26, 2011, *Canada Evidence Act (R.S., 1985, c. C-5)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/C/C-5.pdf).

[40] International Research on Permanent Authentic Records in Electronic Systems (InterPARES), (http://www.interpares.org/)

[41] InterPARES 2, AUTHENTICITY, RELIABILITY AND ACCURACY OF DIGITAL RECORDS IN THE ARTISTIC, SCIENTIFIC AND GOVERNMENTAL SECTORS, Report authored by John Roeder, The University of British Columbia, Philip Eppard, University of Albany, State University of New York, William Underwood, Georgia Tech Research Institute and Tracey P. Lauriault, Carleton University, Accessed March 2011 (http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_part_3_domain2_task_force.pdf)

geomatics practitioners do not normally use the term authenticity, but instead use the terms lineage, data provenance or data integrity. The *Benchmark Requirements for Authenticity*, referred to in the InterPARES 2 report, are essentially the same elements that are found in good metadata (e.g., the ISO 19115 standard). While this may be the case, the practices of managing these data also form part of how they are kept authentic.

Authentication, which is often used in reference to digital information, is not the same as authenticity. Authentication simply refers to the fact that data are kept secure by a process of passwords or by another security system, and that the data have not been modified in or during transfer. Authenticity is closely tied to the concept of trustworthiness. An object that is believed to be and proven to be authentic is considered to be trustworthy. Trust is the quality that underpins social relations, and business and juridical transactions. Trust is often built or erected on the guarantees that data or records are authentic, reliable, and accurate. These qualities, among other features, suggest that a thing or person is trustworthy (for further elaboration on authenticity as it pertains to data see the paper, *Today's Data are Part of Tomorrow's Research: Archival Issues in the Sciences*[42]).

However, where there are no metadata, or in the case where a user wishes to assess if a data set has been tampered with during transfer, the concept of presumption of authenticity is introduced and defined as "an inference as to the fact of a record's authenticity that is drawn from known facts about the manner in which that record has been created and maintained"[43]. For example, in the US "the authenticity of records and documents is usually presumed, rather than requiring affirmation. Federal rules of evidence stipulate that to be presumed authentic, records and documents must be created in the 'regular practice' of business and that there be no overt reason to suspect the trustworthiness of the record (Uniform Rules of Evidence, as approved July 1999)."[44] This being the case, the context, practices, associated documentation, validation processes and authentication, and access measures would suggest that data produced by NRCan and other geospatial data producing departments and agencies would be presumed authentic from an archival perspective. If records are well managed the following should not be problematic:

> 31.3 For the purposes of subsection 31.2(1), in the absence of evidence to the contrary, the integrity of an electronic documents system by or in which an electronic document is recorded or stored is proven
>
> (a) by evidence capable of supporting a finding that at all material times the computer system or other similar device used by the electronic documents system was operating properly or, if it was not, the fact of its not operating properly did not affect the integrity of the electronic document and there are no other reasonable grounds to doubt the integrity of the electronic documents system;

---

[42] Lauriault, T. P., B. Craig, P. L. Pulsifer, and D. R. F. Taylor, 2008, Today's Data are Part of Tomorrow's Research: Archival Issues in the Sciences. *Archivaria #64*.

[43] InterPARES 2, *Terminology and Glossary*, (Vancouver, 2007), http://interpares.org/ip2/ip2_terminology_db.cfm (accessed January 17, 2007).

[44] Pearce-Moses, *A Glossary of Archival and Records Terminology*, http://www.archivists.org/glossary (accessed 23 August 2007).

> (b) if it is established that the electronic document was recorded or stored by a party who is adverse in interest to the party seeking to introduce it; or
>
> (c) if it is established that the electronic document was recorded or stored in the usual and ordinary course of business by a person who is not a party and who did not record or store it under the control of the party seeking to introduce it.
>
> 31.5 For the purpose of determining under any rule of law whether an electronic document is admissible, evidence may be presented in respect of any standard, procedure, usage or practice concerning the manner in which electronic documents are to be recorded or stored, having regard to the type of business, enterprise or endeavour that used, recorded or stored the electronic document and the nature and purpose of the electronic document.

A computer system, data that is contained within it, an electronic document and an electronic document system can all be called into evidence. The presumption that a valuable and authentic archive of data will emerge over time rests on a foundation of trust built between the original provider, his or her successors, including archivists and archival agencies, and the end user. That foundation is anchored in two pillars: 1) an explicit description of the sources of the data and of the changes and processes that the data have undergone over time, so that any user is able to come to a decision about whether the data fit their proposed use; and 2) the continuing authority that the portal or catalog, for instance, maintains as a viable community of practice and data. This would largely be the result of a robust mandate, a stable sponsorship that provides an assured source of funding, open and accessible policies governing access to the data, and a declared or understood commitment to the public good. It is presumed that geospatial data producing departments such as NRCan manage their data this way, and that they follow TBS records management guidelines and records (see Chapter 4).

2.8 Department of Natural Resources Act[45]

In the *Department of Natural Resources Act* (1994, c. 41), the powers, duties and functions of the Minister make explicit reference to the types of geospatial data he or she is responsible for, namely "technical surveys relating to any matter other than a matter to which the powers, duties and functions of the Minister of the Environment and the Minister of Fisheries and Oceans extend by law". The Minister's General Duties that are involved in the collection, maintenance and dissemination of geospatial data, databases and maps are as follows:

> (c) assist in the development and promotion of Canadian scientific and technological capabilities;
>
> (d) participate in the development and application of codes and standards for technical surveys and natural resources products and for the management and use of natural resources;
>
> (f) participate in the enhancement and promotion of market access for Canada's natural resources products and technical surveys industries, both domestically and internationally;

[45] Justice Canada, Current to February 26, 2011, *Department of Natural Resources Act (1994, c. 41)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/N/N-20.8.pdf).

> (g) promote the development and use of remote sensing technology;
>
> (i) gather, compile, analyze, coordinate and disseminate information respecting scientific, technological, economic, industrial, managerial, marketing and related activities and developments affecting Canada's natural resources.

Clearly geospatial data, databases and maps are required to create and formulate policies that pertain to Canada's natural resources, sustainable development and responsible use of natural resources which also form part of the Minister's duties. In addition, the Minister requires geospatial data to inform the production of reports in compliance with the *Report on the State of Canada's Forests Regulations*[46]. The geospatial data, databases and maps created also have to be accompanied with the methodologies and models associated with how these were collected since these form part of the codes and standards for technical surveys. This act makes clear that the geospatial data, databases and maps produced by NRCan for it to fulfill its powers, duties and functions are government records which should be managed accordingly, irrespective of who in the Department creates them (e.g., scientists, data procured from an outside agency). As previously discussed, these data all fall under Crown Copyright, and are governed by the Acts and Regulations listed in Table 1, along with all of the TBS records management guidelines and records discussed in Chapter 4.

2.9 Resources and Technical Surveys Act[47]

The duties of the Minister pertaining to geospatial data, databases and maps in the *Resources and Technical Surveys Act* (R.S., 1985, c. R-7) are to:

> (a) collect and publish statistics of the mineral exploration, development and production and of the mining and metallurgical industries of Canada, and such data regarding the economic minerals of Canada as relate to the processes and activities connected with their utilization, and collect and preserve records of mines and mining works in Canada;
>
> (b) make detailed investigations of mining camps and areas containing economic minerals or deposits of other economic substances, for the purpose of determining the mode of occurrence and the extent and character of the ore-bodies and deposits of the economic minerals or other economic substances;
>
> (c) make a full and scientific examination and survey of the geological structure and mineralogy of Canada;
>
> (d) make the chemical, mechanical, metallurgical and other researches and investigations that are necessary or desirable to carry out the purposes and provisions of this Act and particularly to aid the mining and metallurgical industry of Canada;

[46] Justice Canada, Current to February 26, 2011, *Regulations Respecting the Report on the State of Canada's Forests (SOR/95-479)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Regulation/S/SOR-95-479.pdf).

[47] Justice Canada, Current to February 26, 2011, *Resources and Technical Surveys Act (R.S., 1985, c. R-7)*, Accessed march 2011 (http://laws.justice.gc.ca/PDF/Statute/R/R-7.pdf)

> (e) collect and prepare for exhibition such specimens of the different ores and associated rocks and minerals of Canada and other materials as are necessary to afford a knowledge of the geology and mineralogy and the mining and metallurgical resources and industries of Canada; and
>
> (f) prepare and publish the maps, plans, sections, diagrams, drawings, documents and data that are necessary to illustrate and elucidate any reports of investigations and surveys made pursuant to this Act.

Additional relevant provisions of the *Act* include:

> 4. The Minister may, for the purpose of obtaining a basis for the representation of the mineral and mining resources and of the geographical and geological features of any part of Canada, cause the measurements, observations, investigations and physiographic, exploratory and reconnaissance surveys to be made that are necessary for or in connection with the preparation of maps, sketches, plans, sections or diagrams.
>
> 5. The Minister may authorize the distribution or sale of products, data, duplicate specimens, maps and other documents produced or issued by or on behalf of the Department.
>
> ***Powers:***
>
> (a) conduct or cooperate with persons conducting applied and basic research programs and investigations and economic studies in relation to those resources, and for that purpose maintain and operate research institutes, laboratories, observatories and other facilities for exploration and research related to the source, origin, properties, development or use of those resources.

2.10 Canada Lands Surveys Act[48]

The geospatial data, databases and maps created as part of the responsibilities of the Surveyor General of Canada under the *Canada Lands Surveys Act* (R.S., 1985, c. L-6) are to be administered as follows:

> The Surveyor General, subject to the direction of the Minister, has the management of surveys under this Act and the custody of all the original plans, journals, field notes and other papers connected with those surveys.

The *Canada Lands Surveys Act* does not make specific reference to digital geospatial data, databases or maps but does make reference to registering paper copies with the registrar, which is presumably an electronic catalogue that may or may not include official digital copies of original plans, journals, field notes and other papers associated with surveys. The records created by a land surveyor are sworn and authentic government records and:

> the Surveyor General shall require every surveyor to verify and affirm by oath or otherwise to the satisfaction of the Surveyor General on each return of his surveys under this Act that the surveyor has faithfully and correctly executed such surveys in accordance with this Act and with any instructions issued to the surveyor by the Surveyor General.

[48] Justice Canada, Current to February 26, 2011, *Canada Lands Surveys Act (R.S., 1985, c. L-6)*, Accessed March 2011, (http://laws.justice.gc.ca/PDF/Statute/L/L-6.pdf)

The surveyor's field notes are also considered government records since the *Canada Lands Surveys Act* states that:

> every surveyor shall keep exact and regular field notes of all his surveys under this Act and shall file them with the Surveyor General in the order of time in which the surveys have been performed.

This assertion also applies to Special Surveys as indicated in the *Act* and plans associated with these Special Surveys may be included in the Canada Gazette and copies must be sworn to attest to their authenticity. The results of special surveys once approved are sent to the Registrar for filing. The *Canada Land Surveys Act* is also directive in how data are to be gathered by specifying the unit of measure, the devices to be used for measuring, and the manner of survey, which are important metadata elements. The Surveyor General decides how plans on Canada Lands are to be plotted, the contents of those plans, how the Surveyor General confirms these plans to make them official, the correction of errors, the effects of new plans, and the filing of plans in the registrar. Certified copies of plans may also be called into evidence.

The geospatial data produced by or on behalf of the Surveyor General are also government records, particularly since his or her responsibilities include confirming the location of true boundary lines:

> 32. (1) All boundary lines of roads, streets, lanes, lots, parcels or other authorized subdivisions of Canada Lands that are defined by monuments in surveys made under this Part shall, after confirmation of the plans by the Surveyor General, be the true boundary lines of those roads, streets, lanes, lots, parcels or other authorized subdivisions, whether or not they are found to contain, on a measurement, the exact area or dimensions described or expressed in a plan, letters patent, grant or other instrument affecting those Canada Lands.

2.11 Canada Lands Surveyors Act[49]

Data and databases associated with the *Canada Lands Surveyors Act*, (1998, c. 14) are only referred to under the Complaints section, whereby an officer of the Complaints Committee shall have access to the data and database of the member of the Association, Canada Lands Surveyor or permit holder in order to investigate the complaint. While not a judicial proceeding, survey data, databases and maps could be called into evidence in order to resolve a survey dispute.

[49] Justice Canada, Current to February 26, 2011, *Canada Lands Surveyors Act (1998, c. 14)*, Accessed March 2011, (http://laws.justice.gc.ca/PDF/Statute/L/L-5.8.pdf)

2.12 Forestry Act[50]

In the carrying out of his or her powers under the *Forestry Act* (R.S., 1985, c. F-30), the Minister of Natural Resources Canada may require geospatial data, databases and maps in order to:

- conduct research relating to the protection, management and utilization of the forest resources of Canada and the better utilization of forest products and may establish and maintain laboratories and other necessary facilities for those purposes;
- undertake, promote or recommend measures for the encouragement of public cooperation in the protection and wise use of the forest resources of Canada;
- enter into agreements with the government of any province or with any person for forest protection and management or forest utilization, for the conduct of research related thereto or for forestry publicity or education;
- provide for the making of forestry surveys and provide advice relating to the protection and management of forests on lands administered by any department or agency of the Government of Canada or belonging to Her Majesty in right of Canada;
- conduct economic studies relating to the forest resources, forest industries and marketing of forest products, make investigations designed to aid the forest industries and woodlot owners of Canada and assist external aid programs relating to forestry; and
- manage Forest Experimental Areas.

Geospatial data, databases and maps may be required and may accompany permits allowing for the cutting and removal of timber as per the *Regulations Respecting the Cutting and Removal of Timber* (SOR/94-118)[51], and also for the report on the State of Canada's Forests as required in the *Regulations Respecting the Report on the State of Canada's Forests* (SOR/95-479).

2.13 Canada-Newfoundland Atlantic Accord Implementation Act[52]

Under the *Canada-Newfoundland Atlantic Accord Implementation Act* (1987, c. 3), the Federal Minister may "cause charts to be issued setting out the offshore area or any portion thereof as may be set out consistent with the nature and scale of the chart", and "in any legal or other proceedings, a chart purporting to be issued by or under the authority of the Federal Minister is conclusive proof of the limits of the offshore area or portion thereof set out in the chart without

[50] Justice Canada, Current to February 26, 2011, *Forestry Act (R.S., 1985, c. F-30)*, Accessed March 2011, (http://laws.justice.gc.ca/PDF/Statute/F/F-30.pdf)

[51] Justice Canada, Current to February 26, 2011, *Regulations Respecting the Cutting and Removal of Timber (SOR/94-118)*, Accessed March 2011 (http://laws.justice.gc.ca/eng/SOR-94-118/page-2.html#anchorbo-ga:s_3).

[52] Justice Canada, Current to February 26, 2011, *Canada-Newfoundland Atlantic Accord Implementation Act (1987, c. 3)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/C/C-7.5.pdf).

proof of the signature or official character of the person purporting to have issued the chart". These charts would be considered government records. These charts may also be called into evidence:

> In any legal or other proceedings, a chart purporting to be issued by or under the authority of the Federal Minister is conclusive proof of the limits of the offshore area or portion thereof set out in the chart without proof of the signature or official character of the person purporting to have issued the chart.

In addition, the Canada-Newfoundland Offshore Petroleum Board may request access to information that may include geospatial data, databases and maps as follows:

> 18. (1) The Federal Minister and the Provincial Minister are entitled to access to any information or documentation relating to petroleum resource activities in the offshore area that is provided for the purposes of this Act or any regulation made thereunder and such information or documentation shall, on the request of either Minister, be disclosed to that Minister without requiring the consent of the party who provided the information or documentation.
>
> Applicable provision
>
> (2) Section 119 applies, with such modifications as the circumstances require, in respect of any disclosure of information or documentation or the production or giving of evidence relating thereto by a Minister as if the references in that section to the administration or enforcement of a Part of this Act included references to the administration or enforcement of the Provincial Act or any Part thereof.

In addition, the Canada-Newfoundland Offshore Petroleum Board is to manage geospatial data, databases and maps as follows:

> The Board shall establish, maintain and operate a facility in the Province for the storage and curatorship of all geophysical records and geological and hydrocarbon samples relating to the offshore area.

A number of other official government records that fall under the *Act* may include geospatial data, databases and maps that are part of: development plans, feasibility studies, production orders, environmental impact statements, socio-economic impact statements, exploration licenses, declarations of areas deemed as discovery or exploration sites, spill management, data collected by safety and conservation officers, and the results of geological, geophysical or geotechnical work performed on or in relation to any portion of the offshore area.

2.14 Canada-Nova Scotia Offshore Petroleum Resources Accord Implementation Act[53]

The types of geospatial data, databases and maps that would be officially generated as part of *Canada-Nova Scotia Offshore Petroleum Resources Accord Implementation Act* (1099, c. 28) are similar to the records discussed above in the context of the *Canada-Newfoundland Atlantic Accord Implementation Act.* As is the case above, the Canada-Nova Scotia Offshore Petroleum Resources Board

> shall have responsibility for the storage and curatorship, in a facility in the Province, of all geophysical and geological records and reports, reports respecting wells and materials recovered from wells in the offshore area and, without limiting the generality of the foregoing, drill cuttings, fluid samples, hydrocarbon samples and cores recovered from wells in the offshore area.

2.15 Canada Oil and Gas Operations Act

There are no explicit references to geospatial data, databases and maps or charts in the *Canada Oil and Gas Operations Act* (R.S., 1985, c. O-7), as are seen in the *Canada-Newfoundland Atlantic Accord Implementation Act* and the *Canada-Nova Scotia Offshore Petroleum Resources Accord Implementation Act* previously discussed. However, it is presumed that these would be produced to accompany any approval and management plans. The associated *Canada Oil and Gas Geophysical Operations Regulations* (SOR/96-117) make clear reference to geospatial data.

2.16 Canada Oil and Gas Geophysical Operations Regulations[54]

The *Canada Oil and Gas Geophysical Operations Regulations* (SOR/96-117) explicitly state that the following geospatial data are to be retained for at least 15 years and deposition must be officially approved by the Chief Conservation Officer:

> 39. (1) Every operator shall, after completion of a geophysical operation, retain in Canada the following information and materials:
>
> (a) seismic field data in digital format and a description of the data format, together with all supporting information;
>
> (b) fully processed, migrated seismic data in digital format;

[53] Justice Canada, Current to February 26, 2011, *Canada-Nova Scotia Offshore Petroleum Resources Accord Implementation Act (1988, c. 28)*, (http://laws.justice.gc.ca/PDF/Statute/C/C-7.8.pdf).

[54] Justice Canada, Current to February 26, 2011, *Regulations Respecting Geophysical Operations in Relation to Exploration for Oil and Gas in any Area to which the Canada Oil and Gas Operations Act Applies (SOR/96-117)*, (http://laws.justice.gc.ca/PDF/Regulation/S/SOR-96-117.pdf).

> (c) in the case of a magnetic survey, the final digital field data, field analog monitors, diurnal charts, altitude profiles, and all other supporting information;
>
> (d) in the case of a gravimetric survey, the location, elevation, final digital field data, and gravity profiles;
>
> (e) in the case of seabed investigations at well-sites, all sidescan sonar records and mosaics, fathometer records, sub-bottom profile records, grab samples, cores, and seabed photographs; and
>
> (f) all other observations or readings that were obtained during the field operation.

In addition, form and formats are specified:

> Every operator shall retain in Canada on reproducible film the most recent fully processed, migrated seismic sections of the geophysical operation and shall not destroy that film or remove it from Canada without the written approval of the Chief Conservation Officer.

Also, the type of equipment to be used for data gathering and methodological instructions are included with reference to their use, location, electrical charging, storage and so on. The contents of final reports are very prescriptive and in the list there is specific mention of maps, 3D models, high resolution sectional surveys, shotpoint maps, track plots, flight lines with numbered fiducial points, gravity station maps and, for seabed surveys, location maps for core holes, grab samples and seabed photographs, locations where data were gathered, data gathering methodologies, and many other elements associated with the collection and reporting of geospatial data (see part 5 of the *Regulation* entitled Reporting Requirements).

2.17 Canada Petroleum Resources Act[55]

The types of geospatial data, databases and maps that would be officially generated as part of *Canada Petroleum Resources Act* (1985, c. 36 (2nd Supp.)) are similar to those in the *Canada-Nova Scotia Offshore Petroleum Resources Accord Implementation Act* and the *Canada-Newfoundland Atlantic Accord Implementation Act* and their associated Boards.

2.18 Energy Efficiency Act[56]

The *Energy Efficiency Act* (1992, c. 36) makes reference to the collection of statistics as these pertain to energy efficiency as follows:

[55] Justice Canada, Current to February 26, 2011, *Canada Petroleum Resources Act (1985, c. 36 (2nd Supp.))*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/C/C-8.5.pdf).

[56] Justice Canada, Current to February 26, 2011, *Energy Efficiency Act (1992, c. 36)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/E/E-6.4.pdf)

> 22. The Governor in Council may make regulations requiring prescribed persons to file with the Minister, in the prescribed form and manner, at the prescribed time and for each prescribed reporting period, a report setting out prescribed statistics and information respecting
>
> (a) the value, quantity, type and use of energy, including alternative energy, purchased, consumed or sold by that person;
>
> (b) the expenditures of that person on the research, development, acquisition and operation of energy-using equipment and related technology; and
>
> (c) the sales of prescribed energy-using products or classes of energy-using products by that person, including the revenue from, and geographic distribution of, the sales.

In addition, the powers of the Minister include being able to "conduct, or cooperate with persons conducting, research, development, tests, demonstrations and studies;" and "publish information, research or test results", which could include the use of geospatial data.

2.19 Energy Monitoring Act[57]

The following is a subset of statistics that may be georeferenced, which are to be collected from Canadian energy enterprises as part of the *Energy Monitoring Act* (R.S., 1985, c. E-8):

> 5. (1) Unless exempted by the regulations, every energy enterprise shall file with the Minister, in prescribed form and manner, for each prescribed reporting period that applies to that energy enterprise, a return setting out statistics and information relating to
>
> (d) its exploration for, development, production, processing, refining and marketing of energy commodities;
>
> (i) its energy commodity resources, reserves and properties;
>
> (j) its ownership of or interest in any corporation, partnership, trust or organization;
>
> (k) its research and development programs; additional statistics and information
>
> (2) In addition to the statistics and information referred to in subsection (1), every energy enterprise required to file a return under subsection (1) shall set out, in prescribed form and manner, in every return required by that subsection, statistics and information relating to the matters referred to in paragraphs (1)(b) to (l) in respect of every corporation it controls.
>
> 8. Every energy enterprise required to file a return by or under this Act shall submit such additional statistics, information and documentation as may be required by the Minister for any purpose related to the administration or enforcement of this Act.

The Act also includes requirements in terms of record retention, record deposition and the disclosure of this information.

[57] Justice Canada, Current to February 26, 2011, Energy Monitoring Act (R.S., 1985, c. E-8), http://laws.justice.gc.ca/PDF/Statute/E/E-8.pdf

2.20 National Energy Board Act[58]

The *National Energy Board Act* (R.S., 1985, c. N-7) makes explicit reference to the submission of maps with applications for pipeline construction certificates as well as the submission of plans and books of reference. Maps are also to be included with Notices to the Attorney General. Geospatial data must also accompany permit applications for the construction and operation of power lines. In addition, plans, profiles, books of reference, specification and certified copies thereof and other documents are to be submitted to the registrar. The Registrar of deeds has the responsibility to maintain authentic copies of these documents. Authenticity is assessed by the appearance of the seal of the Board on submitted and approved documents. Also, maps and other geospatial data are required as part of the *National Energy Board Electricity Regulations* (SOR/97-130)[59], specifically:

> (e) a map, on a scale sufficient to locate and identify all essential features, showing
>
> > (i) all terminal points, the general route, the international boundary crossover point and the distance in kilometres from the international boundary crossover point to each terminal point of the international power line in and outside Canada,
> >
> > (ii) the provinces, cities, towns, villages, park boundaries, rivers, major roads, railways and navigable waters through, under or across which the international power line is to pass, and
> >
> > (iii) the power line outside Canada;
>
> (f) a plan of survey from which the international boundary crossover point can be accurately determined on the ground

Also, for international power lines greater than 50 kV, along with the above the following map information is required:

> (e) a map on a scale sufficient to locate and identify
>
> > (i) the general route and facility sites being considered,
> >
> > (ii) the alternative route and facility sites under consideration,
> >
> > (iii) the areas subject to physical and environmental constraints, including biophysical and land use or natural resource use constraints, that limit the general route or facility sites, and
> >
> > (iv) the approximate sites of all proposed ancillary facilities.
>
> ...
>
> (t) unless otherwise detailed in the report referred to in paragraph(s),
>
> > (i) a map showing the proposed general route and covering a width of at least one kilometre on each side of the international power line, on a scale sufficient to clearly

[58] Justice Canada, Current to February 26, 2011, *National Energy Board Act (R.S., 1985, c. N-7)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/N/N-7.pdf)

[59] Justice Canada, Current to February 26, 2011, Regulations for Carrying into Effect the Provisions of the National Energy Board Act Respecting International Power Lines and the Exportation of Electricity (SOR/97-130)m

> show the existing environment, including the surface geology, the habitats of wildlife of ecological, economic or human importance, rare and endangered plant species, spawning beds, public recreational areas, parks, historic and archaeological sites, conservation areas, Indian reserves and existing land use, and a description of the environmental components shown on the map

Other geospatial data might be required for environmental impact assessments and for the issuance of permits for ongoing inspections and maintenance.

2.21 Northern Pipeline Act[60]

The *Northern Pipeline Act* (R.S., 1985, c. N-26) does not make specific reference to geospatial data and maps. However, a number of Northern Pipeline Socio-Economic and Environmental Terms and Conditions with British Columbia and Alberta and Saskatchewan require the submission of geospatial data as part of social and environmental impact assessments, consultations, surveillance and monitoring, and to meet stated environmental terms and conditions.

2.22 Nuclear Energy Act[61]

Under the *Nuclear Energy Act* (R.S., 1985, c. A-16), it is presumed that the Minister requires geospatial data, databases and maps in order to fulfill the following powers as stated in the act:

- to conduct research and investigations with respect to nuclear energy;
- to utilize, cause to be utilized and prepare for the utilization of nuclear energy;
- acquire or cause to be acquired, by purchase, lease, requisition or expropriation, nuclear substances and any mines, deposits or claims of nuclear substances and patent rights relating to nuclear energy and any works or property for production or preparation for production of, or for research or investigations with respect to, nuclear energy

[60] Justice Canada, Current to February 26, *2011, Northern Pipeline Act (R.S., 1985, c. N-26)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/N/N-26.pdf).

[61] Justice Canada, Current to February 26, 2011, *Nuclear Energy Act (R.S., 1985, c. A-16)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/A/A-16.pdf)

2.23 Nuclear Fuel Waste Act[62]

The purpose of the *Nuclear Fuel Waste Act* (2002, c. 23) is to "provide a framework to enable the Governor in Council to make, from the proposals of the waste management organization, a decision on the management of nuclear fuel waste that is based on a comprehensive, integrated and economically sound approach for Canada" and it is presumed that geospatial data would be required to fulfill its purpose.

2.24 Nuclear Safety and Control Act[63]

It is implied that geospatial data might be necessary to fulfill the requirements of the *Nuclear Safety and Control Act* (1997, c. 9), although no specific mention is made. However, the related *Uranium Mines and Mills Regulations* (SOR/2000-206)[64] do make specific mention of geospatial data, particularly as it relates to mining license applications:

> (a) in relation to the plan and description of the mine or mill,
>
> > (i) a description of the site evaluation process and of the investigations and preparatory work to be done at the site and in the surrounding area,
> >
> > (ii) a surface plan indicating the boundaries of the mine or mill and the area where the activity to be licensed is proposed to be carried on,
> >
> > (v) a description of the site geology and mineralogy,
>
> …
>
> (c) in relation to the environment and waste management,
>
> > (vii) the proposed location, the proposed maximum quantities and concentrations, and the anticipated volume and flow rate of releases of nuclear substances and hazardous substances into the environment, including their physical, chemical and radiological characteristics,
> >
> > (ix) a description of the anticipated liquid and solid waste streams within the mine or mill, including the ingress of fresh water and any diversion or control of the flow of uncontaminated surface and ground water,
>
> (A) assist off-site authorities in planning and preparing to limit the adverse effects of an accidental release,
>
> (C) report information to off-site authorities during and after an accidental release

---

62 Justice Canada, Current to February 26, 2011, *Nuclear Fuel Waste Act (2002, c. 23)*, Accessed March 2011, (http://laws.justice.gc.ca/PDF/Statute/N/N-27.7.pdf)

63 Justice Canada, Current to February 26, 2011, Nuclear Safety and Control Act (1997, c. 9), Accessed March 2011, (

64 Justice Canada, Current to February 26, 2011,*Uranium Mines and Mills Regulations*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Regulation/S/SOR-2000-206.pdf).

2.25 Remote Sensing Space Systems Act[65]

The *Remote Sensing and Space Systems Act* (2005, c. 45) is not part of the list of *Acts for which the Minister of Natural Resources is Responsible* (see Appendix C). However, data acquisition, management, storage, archiving and dissemination and other remote sensing activities are administered by the Canada Centre for Remote Sensing, which is part of NRCan. The *Remote Sensing Space Systems Act* came into effect in 2007, and is very specific when it comes to data, databases and the licensing thereof. Also, all private or public sector organizations that have a license to operate a remote sensing system are governed by the *Act* and would be considered a licensee. Under the *Act*, the ministers of National Defence, Foreign Affairs and Emergency Preparedness Canada may have priority access to remotely sensed data. Under the conditions of a license, in relation to geospatial information the *Act* states:

> (c) that raw data and remote sensing products from the system about the territory of any country — but not including data or products that have been enhanced or to which some value has been added — be made available to the government of that country within a reasonable time, on reasonable terms and for so long as the data or products have not been disposed of, but subject to any license conditions under subsection (6) or (7) applicable to their communication or provision;
>
> (d) that the licensee keep control of raw data and remote sensing products from the system until they are disposed of in accordance with this Act;
>
> (e) that raw data from the system be communicated only to a government referred to in paragraph (c), the licensee, a system participant or a person to whom they may be communicated under subsection (6);
>
> (f) that the licensee encourage a recipient of raw data or a remote sensing product who has entered into an agreement referred to in paragraph (6)(b) or (7)(b) to comply with the agreement

Under conditions specified by the *Act*, the following provisions relating to geospatial information may also apply:

> (6) In a license, the Minister may authorize the communication of raw data or classes of raw data from the licensed system to any persons or classes of persons other than the licensee or system participants on any conditions that the Minister considers appropriate. The conditions may include requirements that, in specified cases or circumstances, the communication of the raw data
>
> (a) be subject to the Minister's prior approval; or
>
> (b) be done only under a legally enforceable agreement, entered into in good faith, that includes measures respecting their security or their further communication.

The disposition of data is also specified in the Act. Furthermore, in terms of regulation relating to geospatial information, the Act specifies:

> 20. (1) On the recommendation of the Minister, the Governor in Council may make regulations
>
> (a) prescribing a process or series of processes that is or is not to be considered to transform raw data;

[65] Justice Canada, Current to February 26, 2011, *Remote Sensing Space Systems Act (2005, c. 45)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/R/R-5.4.pdf)

(g.1) respecting the archiving of raw data, including the public access to the archived data;

(h) respecting the keeping of records;

(iv) respecting security assessments of proposed or actual licensees or system participants, and

(v) prescribing conditions of licenses

2.26 Remote Sensing Space Systems Regulations[66]

The *Remote Sensing Space Systems Regulations* (SOR/2007-66)[67], includes information on what is and is not considered to be data transformation:

Process not considered to transform raw data

13. (1) Any process that retains the phase information of raw data, or that produces an output from which measurements can be taken to determine the phase response of a remotely sensed surface, including the process to produce the synthetic aperture radar output known as Single Look Complex, is not considered to transform the raw data.

Process considered to transform raw data

(2) Any process or series of processes operating on raw data that rectifies errors, distortions and other artifacts of the system by pixel aggregation, averaging or re-sampling are considered to transform the raw data if the process or series of processes also

(a) radiometrically calibrates the data; or

(b) geocodes the data with respect to features of the Earth by re-sampling

The maintenance of records and data archiving and disposal relating to geospatial information is also clearly stated in the *Remote Sensing Space Systems Regulations* as follows:

Maintenance of records

16. (1) A licensee must maintain the following records for a period of one year:

(a) a record of every sales order placed with it;

(b) a record of every command given to each remote sensing satellite of the remote sensing space system, including the date and time of the command;

(c) a record of all raw data received from each remote sensing satellite, including the date and time of receipt;

(d) a record of raw data being entered into the archives of the licensee and the disposal of raw data, including the date of each entry and disposal;(e) a catalogue that lists the raw data that is accessible to the public, including the date of each entry into the catalogue;

Sales orders

[66] Justice Canada, Current to February 26, 2011, Remote Sensing Space Systems Regulations (SOR/2007-66), Accessed March 2011, (http://laws.justice.gc.ca/PDF/Regulation/S/SOR-2007-66.pdf)

[67] Justice Canada, Current to February 26, 2011, Remote Sensing Space Systems Regulations (SOR/2007-66), Accessed March 2011, (http://laws.justice.gc.ca/PDF/Regulation/S/SOR-2007-66.pdf)

(2) The licensee must keep the records in a manner that enables the ready determination of the following in respect of every sales order:

(a) the date and time of the sales order;

(b) the name and contact information of the person who placed the sales order;

(c) the type of raw data or remote sensing product ordered;

(d) the location sensed in order to fill the sales order; and

(e) the name and contact information of the recipient of the raw data or remote sensing product and the conditions under which the recipient may make use of the data or product.

Archiving raw data

17. (1) A licensee must archive raw data from the remote sensing satellite in a readily retrievable format for a minimum period of 15 months from

(a) the day on which an entry for the raw data was first made in a catalogue accessible to the public; or

(b) if the raw data was not entered in a catalogue accessible to the public, the day on which the raw data was received by a ground station.

Notice of proposed disposal to Minister

(2) A licensee must, before disposing of the raw data, notify the Minister and provide the following information about each scene of raw data:

(a) a unique identifier;

(b) the date and time when the raw data was acquired by the remote sensing satellite;

(c) the geographic boundaries of the scene;

(d) the position of the satellite when the raw data was acquired;

(e) the sensor modes used to acquire the raw data;

(f) the ground station that received the raw data;

(g) the date and time when the ground station received the raw data;

(h) the date on or after which the raw data will be disposed of;

(i) the cost to communicate the raw data; and

(j) the name and contact information of a contact person.

Disposal of raw data

(4) The licensee may not dispose of the raw data until three months after the day on which the notice was sent.

Notice of proposed disposal to others

18. On receipt of a notification under subsection 17(2), the Minister may, and the licensee must on the request of the Minister, inform any person to whom the licensee is authorized to communicate raw data under subsection 8(6) of the Act about the proposed disposal of the raw data.

Request for raw data

19. (1) The Minister or a person to whom the licensee is authorized to communicate raw data under subsection 8(6) of the Act may, at any time before the raw data is disposed of, request the

> communication of the raw data from the contact person referred to in paragraph 17(2)(j), and the licensee must provide the raw data as soon as feasible.

In the application of a license the following must be included, in relation to geospatial information:

> **Data Protection Plan**
>
> 23. The location and function of all facilities, including mobile facilities, to be used to handle raw data and remote sensing products in the operation of the remote sensing space system.
>
> 24. A general description and block diagram of all facilities to be used to handle raw data and remote sensing products, including the longitude and latitude and station mask of each ground station.
>
> 29. Proposed measures to comply with any conditions of the license that restrict the communication of raw data or the provision of remote sensing products related to
>
> (e) the sensed territory;
>
> (f) the location of the recipients; and
>
> **Command and Data Protection Plan**
>
> 30. In lieu of a separate command protection plan and data protection plan, a combined command and data protection plan that contains the information and documents set out in sections 14 to 29 of this Schedule.

2.27 Charts and Nautical Publications Regulations[68]

The *Charts and Nautical Publications Regulations 1995 (*SOR/95-149), pursuant to the *Arctic Waters Pollution Prevention Act* (R.S., 1985, c. A-12)[69] makes specific reference to geospatial data and databases as follows:

- ECDIS, which is an electronic chart display and information system; (SVCEI)
- ENC, which is an electronic navigational chart database
- RCDS, which is a raster chart display system
- RCDS reference catalogue, in respect of an area to be navigated by a ship, means (a) for waters under Canadian jurisdiction, the Catalogue of Nautical Charts and Related Publications, published by the Canadian Hydrographic Service, and (b) for waters outside Canadian jurisdiction, the Catalogue of Admiralty Charts and Other Hydrographic, Publications, published by the Government of the United Kingdom, or

[68] Justice Canada, *Regulations Requiring the Presence on Board Ships of Appropriate Charts, Tide Tables and Other Nautical Documents or Publications and Respecting their Maintenance and Use (SOR/95-149)* which can be referenced as *Charts and Nautical Publications Regulations, 1995,* Accessed March 2011, *(*http://laws.justice.gc.ca/PDF/Regulation/S/SOR-95-149.pdf).

[69] Justice Canada, current to March 9th, 2011, *Arctic Waters Pollution Prevention Act (R.S., 1985, c. A-12),* Accessed March 2011, (http://laws.justice.gc.ca/en/showtdm/cs/A-12)

the Catalog of Charts and Publications, published by the Government of the United States of America; (catalogue de référence)

- RNC, which is a raster navigational chart that is a facsimile of a paper chart issued on the authority of the Canadian Hydrographic Service or a hydrographic office authorized by the government of a country other than Canada

In addition, the *Regulation* directs: how these data and databases are to be standardized as to content, structure and format; how they are to be issued for use with an ECDIS on the authority of the Canadian Hydrographic Service or a hydrographic office authorized by the government of a country other than Canada; and also what they should contain, such as all the chart information necessary for safe navigation. The *Regulation* also states that the most up to date data and catalogue along with the latest edition of the Notices to Mariners are to be used, and indicates the areas the maps, data and databases are to cover, the issuing authority, and the scale of the maps to be used under specific conditions. It stipulates that these charts can only be viewed in electronic form if on ECDIS or in the case of a backup of ECDIS. The *Regulation* is also directive in terms of ECDIS performance standards on board a ship and how backup arrangements are to be managed, electrical standards are to be adhered to, and access to these databases are to be carried out with manufacturers that adhere to quality control standards.

Should an accident occur, it is probable that these data would be called into evidence and it therefore becomes critical to access the geospatial data, databases and maps that were used to inform the decisions which may have led to that accident. Because of the nature of these datasets, their relation to public safety and well being and how they are used, it is also critical that these be well managed, backed up and maintained in such a way that their authenticity can be attested to in order to conform with the *Canada Evidence Act* (R.S., 1985, c. C-5)[70], as previously discussed.

2.28 Summary

This chapter has identified some of the key Canadian acts and regulations that govern geospatial data, databases and maps. The legislation reviewed illustrates the diversity of types of data, their forms, formats and associated software and systems, how they are managed, the context within which they are created, and how they are considered authentic. This accounts to some extent for the heterogeneity of the data, databases and maps discussed in Chapter 7, which explores a few existing geospatial data preservation strategies.

Because of the diverse nature of geospatial and scientific data, it may not be possible to make specific recommendations at the data or database level. However, it is clear that any geospatial

[70] Justice Canada, Current to February 26, 2011, *Canada Evidence Act (R.S., 1985, c. C-5)*, Accessed March 2011 (http://laws.justice.gc.ca/PDF/Statute/C/C-5.pdf).

data preservation and management strategy will require the full collaboration of data creators, maintainers and archivists, especially as seen above. This is so because some geospatial datasets are very highly regulated and are part of Canadian statutes, which makes them a high priority, but also because it would not be possible for any one archivist to have the full range of specializations and knowledge required to ensure these data are preserved within their scientific context. Data without context are not much use to future generations, and context, as seen here, is more than simply metadata; it includes methodology, systems, operations, use, attestations to authenticity, legal frameworks, norms and applications.

3. LAC Consultations

The HAL team and the GeoConnections Project Authority met with representatives of Library and Archives Canada (LAC) on February 11, 2011 to learn more about LAC's work in digital data archiving and their views on archiving and preservation of digital geospatial data (see Appendix A for meeting attendees). This section summarizes the results of this information exchange plus other research of LAC activities relating to geospatial information.

3.1 Introduction

Radical changes in how Canadian heritage content is being created has transformed how Library and Archives Canada (LAC) preserves Canada's works. The geomatics sector has also witnessed change in how geospatial data and maps are created. Within a 50 year period the production of Canada's geospatial content went from specialists creating the first GIS and experimenting with the creation of records in digital formats (e.g. the Canada Land Inventory) to the current context where most records are borne digital in myriad software platforms and many data are captured from sensors in near real-time. LAC and those who produce and manage geospatial datasets in government are challenged with: the pace of this change; incessant technological innovation; the deluge of digital content and the increasing complexity in the nature of the records created.

The large and growing collection of digital maps, geospatial datasets and databases produced at NRCan and elsewhere in the federal government is starting to stabilize in terms of formats, standards, and interoperability while the managing of complex databases and their associated interactive, dynamic, distributed, multi-authored and cross jurisdictional ways of working are becoming normalized. These business practices are well established and the challenge is to find ways to better manage the results of those practices for the long term. LAC on the other hand is expected to keep pace with transformations not only in the geomatics sector but with all content-producing Canadian institutions both within government and external to it. LAC is expected to manage Canada's digital heritage while: experiencing budget constraints; dealing with the dearth of skilled digital archivists; and also having to transform the core practices and theories of their discipline.

LAC is responding by reshaping how the institution is organized, and has just recently created the Digital Resources Division which assembles within one organizational unit digital archival specialists from many disciplines that were once scattered across the institution. In addition, this new Division is in the process of revisiting digital content policies, frameworks, guidelines and practices while concurrently developing the technological capacity and human resources to

manage borne digital content. The current state of flux at LAC is a reflection of the ever changing context within which they are operating. Meeting participants discussed how the institution as well as the field of archiving was undergoing much change, which limited their ability to provide definitive responses to some of the questions posed. However, this did not prevent those in attendance from having a candid conversation about the issues represented by these questions. The meeting was useful, very informative and provided some insight into how NRCan and LAC can work together and provided some ideas for NRCan to consider when managing its records with the goal of preservation.

3.2 LAC Definitions

The definitions upon which archivists once relied to describe the digital objects they manage are changing as the nature and form of Canada's digital content is changing. Irrespective, helpful information was provided. Also, it was noted that the definitions in the ACT are not to be considered as media specific.

A digital map, a digital mapping dataset and a digital mapping database can be a record, a recording, a publication, part of a ministerial record and also be of documentary heritage as defined in the Act. How these are classified is contingent on how it fits in NRCan's and other digital mapping data producers' business practices and how these are valued by the institution. For instance the Atlas of Canada may be a publication, which is also a record that could also be considered to be of documentary heritage. NRCan, the record creator, would appraise[71] it as such, as would Atlas Stakeholders (e.g., school teachers) or potentially the Librarian and Archivist of Canada might do so. How and when to capture snapshots of the Atlas, whether or not that is the best method to adopt, and when to accession[72] it in the Archive is has not yet been determined, and would have to be discussed between LAC and NRCan. NRCan and LAC would then collaborate in deciding how the Atlas is to be made available to future generations of Canadians and how those future generations will view the Atlas. The processes of appraising, how to manage, and how to preserve would be repeated for other datasets, databases and maps.

A digital map, dataset or database could be a ministerial record if it formed part of a particular decision in that office. These digital objects, irrespective of their form, would then be a part of that particular ministerial record set.

[71] n., The process of assessing the value of records for the purpose of determining the length and conditions of their preservation. (InterPARES 2 Terminology Database - http://www.interpares.org/ip2/ip2_terminology_db.cfm)

[72] v., To take legal and physical custody of a body of records and to document it in a register. (InterPARES 2 Terminology Database - http://www.interpares.org/ip2/ip2_terminology_db.cfm)

3.3 Archiving and Preservation Resources

LAC is considering a distributed archive approach, whereby the institutions that create records would also manage their records as LAC would, those records would be catalogued at LAC and that catalogue would point back to the host institution. The issue of trust in this recordkeeping approach remains a key question, which has been discussed at length in archival literature. In the interim, they suggest that record creators adopt good recordkeeping practices and refer to the following documents for guidance to do so:

- TBS Directive on Recordkeeping[73]
- TBS Directive on Information Management Roles and Responsibilities[74]
- TBS Standard for Electronic Documents and Records Management Solutions (EDRMS)[75]
- TBS Policy on Information Management[76]
- TBS Policy Framework for Information and Technology[77]
- Policy on Management of Information Technology[78]
- TBS Multi-Institutional Disposition Authority (MIDA)[79]
- LAC Guidelines: Local Digital Format Registry (LDFR)[80]
- TBS Standard on Geospatial Data[81]
- and the forthcoming Methodology associated with the TBS Directive on Recordkeeping that NRCan Records Management Group contributed to along with LAC (to be released on April 14th).

Karine Burger, Acting Manager, Digital and Accessibility Office is the main contact on the ongoing consultations regarding the recordkeeping methodology and Peter Cowan is her NRCan counterpart on that file. This methodology document, which is being developed by a consultant under contract with LAC in collaboration with NRCan¸ deals with information management and is not specific to data preservation, but does deal with risk.

[73] See http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=16552
[74] See http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=12754
[75] See http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18910&section=text
[76] See http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=12742
[77] See http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12452&section=text
[78] See http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=12755
[79] See http://www.collectionscanada.gc.ca/government/disposition/007007-1062-e.html and http://www.collectionscanada.gc.ca/government/disposition/index-e.html
[80] See http://www.collectionscanada.gc.ca/digital-initiatives/012018-2200-e.html
[81] See http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16553&section=text

LAC believes that archiving has to be backed up by sound data management principles. A key unresolved issue is the frequency with which a snapshot of dynamic digital mapping databases should be taken (e.g., annually, coincident with the Census, when major technological changes are made, etc.). LAC is also looking toward the creation of a Pan-Canadian collaborative network of archiving, record management, librarian and other related experts.

As discussed in Chapter 2, NRCan and other federal geospatial data producers must also adhere to the extant Legal Deposit regulations[82] as stipulated in the *Library and Archives Act*[83]. This means that publications produced by NRCan and other federal geospatial data producers are subject to legal deposit (e.g., reports and other resources on the GeoConnections Web site). Databases have not yet been subject to deposit. Dynamic databases are subject to "on request" via legal deposit (i.e. on the request of the Librarian and Archivist of Canada) but are not absolutely required under these regulations. Since there are concerns about LAC's capacity to handle this type of data, they are not being proactive in ensuring that these requirements are met. Guidelines are provided for Legal Deposit of an Online Publication[84], along with pertinent additional information about legal deposit[85].

Regarding Framework datasets, which are the products of provincial and territorial jurisdictions but are disseminated and assembled by the Federal Government under the Geomatics Accord, there was discussion that these may not be considered as part of LAC's mandate as these may not be Federal records. IP and copyright issues would have to be more closely examined and one recommendation was to include preservation and archiving authority in the Accords. A recommended good place to start is to assess records based on their institutional business value, particularly those specifically mentioned in the Acts and Regulations reviewed earlier, as that list in turn becomes a short list for future preservation.

3.4 Suitability of Geospatial Datasets for Archiving and Preservation

Although LAC would not be directive regarding the suitability of any digital maps, digital mapping datasets and digital mapping databases for archiving and preservation, all government institutions are to manage all of their records in such a way that they could be seamlessly accessioned into the Archive as previously discussed. LAC would work collaboratively with government institutions to help them identify digital maps, datasets and databases to be archived (e.g., the record set of a particular NRCan Scientist's body of work, maps, datasets and databases). Institutions must identify information that has ongoing business value and decide what parts of this information must be preserved. Archiving could be done by the

[82] See http://www.collectionscanada.gc.ca/legal-deposit/index-e.html
[83] See http://laws.justice.gc.ca/eng/L-7.7/page-3.html#anchorbo-ga:s_10
[84] See http://www.collectionscanada.gc.ca/electroniccollection/003008-1000-e.html
[85] See http://www.collectionscanada.gc.ca/legal-deposit/041008-0200-e.html

institutions (distributed archiving model) or by LAC. If operational information, it should be preserved by the institution; if a program or activity is terminated, LAC needs to preserve the information deemed of archival value. Also, an institution like NRCan or LAC itself may have identified records that are at risk of deterioration, and if so, LAC is mandated to assist with their preservation (e.g., early digital editions of the Atlas of Canada might be identified as being at risk). LAC already works with NRCan, in particular the National Air Photo Library (NAPL) and the GeoGratis Program, where LAC maintains and owns the original copies of the digital Canada Land Inventory[86] and associated data while GeoGratis disseminates these data[87]. Such arrangements in the future are possible and most welcomed.

In the case of a program ending suddenly, LAC is able to step in and acquire archival records because of a clause in the LAC Act on records at risk. It would not need to use one or more Multi-Institutional Disposition Authorities (MIDAs) to acquire the records. Where they apply, MIDAs are legal instruments issued by LAC for departments. NRCan or any other federal geospatial data producer can collaborate with LAC on the appraisal and accessioning of those records and to determine how to maintain them for preservation purposes.

3.5 Agreements on Transfer of Geospatial Datasets

There are a number of bilateral agreements, and tools under which these specific agreements are made (e.g., Records Disposition Authorities or RDAs and MIDAs), in place, including for example:

- DSP LIST (Depository Service Program)[88];
- with National Air Photo Library;
- with GeoGratis; and
- with a number of other agencies producing digital geospatial maps, datasets and databases (e.g., DFO Hydrographic Services, Statistics Canada, Elections Canada, National Capital Commission).

Often, specific formats are addressed in the Terms and Conditions that accompany an authority.

It was noted that the agreement currently in place with Mapping Information Branch of NRCan is outdated, although topographic maps are on the DSP list. LAC currently has approx. 100 cadastral map and 50 NTS map TIFF files in their archives. Marc Cockburn, Archivist,

[86] CLI in LAC catalogue: http://collectionscanada.gc.ca/pam_archives/index.php?fuseaction=genitem.displayItem&lang=eng&rec_nbr=133094&rec_nbr_list=133094,3691606,823187,213916,1723948,1785071,825107,2310705,2310704,3683452

[87] GeoGratis CLI Search Results: http://www.geogratis.cgdi.gc.ca/geogratis/en/collection/search.do?search=keyword&keyword=canada+land+inventory

[88] See http://esp-psd.pwgsc.gc.ca/index-e.html

Cartography, Architecture and Geomatics, would be the main contact if further elaboration is required.

LAC is an official member of the IACG but currently does not have an appointed official in that seat. LAC was also a member of the GeoConnections Policy Advisory Node's Archiving and Preservation Working Group and LAC's David Brown was the lead author on the *Archiving, Management and Preservation of the Geospatial Data*[89] report.

3.6 New LAC Policies and Frameworks

LAC has been working on a number of internal policies and frameworks associated with digital data archiving and preservation. The Digital Preservation Policy and the Digital Collection Development Policy are internal documents that are still in discussion and do not yet provide any useful guidance for NRCan or other federal geospatial data producers. The Documentary Heritage Management Framework is also still not fully developed but LAC is developing a construct that will help institutions to decide what information is business critical and help them appraise information resources against this Framework's four principles. Similarly, the Collection Development Framework is still in discussion within LAC. LAC suggested that the producers of digital geospatial data access the documents referred to in Section 2.1.3 for useful guidance.

3.7 LAC Trusted Digital Repository Status

It is projected that the Trusted Digital Repository (TDR) system will be operational by 2017. It is important to note that the TDR is not just the TDR system currently under development but a collection of systems within LAC that will result in LAC being a "Trusted Digital Repository" in and of itself (i.e. the TDR is not just a system, it is a collection of tools, policies, procedures and workflows performed by LAC).

Interoperability and a standards driven approach to content creation are the direction this work is going. The TDR is being re-architected in a modular fashion, in view of resource constraints. The Government of Canada electronic document and records management system (EDRMS) was to be the system to bring government records seamlessly into the TDR Virtual Loading Dock. However, many government departments customized their systems, which makes them currently non-interoperable with the VLD, which has delayed the progress.

[89] Archiving, Management and Preservation of Geospatial Data (http://www.geoconnections.org/publications/policyDocs/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf)

The TDR uses persistent identifiers but there is currently no official stance on what institutions should adopt. However, such use would make the management of digital maps, and geospatial datasets and related databases much easier. Some work has been done in the field of Geomatics on this issue (e.g., the national standards for framework data (e.g., GeoBase) include unique identifiers for each feature). In its *Local Digital Format Registry File Format Guidelines for Preservation and Long-term Access*, LAC is recommending use of the TC 211 ISO 19115 Geographic Information - Metadata (NAP – Metadata) (North American Profile)[90].

[90] See http://www.collectionscanada.gc.ca/digital-initiatives/012018-2200-e.html

4. Canada's Information Management Policies

This chapter describes the research findings and an analysis with respect to information management policies at the federal level in Canada related to data archiving and preservation in general and geospatial information specifically. Primary responsibility for such policies lies with the Treasury Board of Canada Secretariat (TBS).

4.1 Policy Framework for Information and Technology[91]

The TBS *Policy Framework for Information and Technology* came into effect in July 2007 and

> provides the strategic context for the Policy on Information Management and the Policy on the Management of Information Technology. It also takes into consideration the Privacy and Data Protection policy, the Access to Information policy, and the Policy on Government Security. In addition, it provides guiding principles to sound information and technology management practices across government. These principles also serve to support individuals and employees in exercising their rights and performing their duties.

The *Policy Framework* provides guiding principles for the following sound information and technology management practices relevant to information archiving and preservation:

> **Information –** Information supports public reporting, sound planning, and decision making for current and future governments.
>
> **Information technology –** Efficient, effective, and innovative information technology (IT) is a key enabler to achieving well-managed information in support of policies, programs, and services.
>
> **Stewardship –** Information must be rigorously managed throughout its life cycle, regardless of medium or format
>
> **Whole-of-government approach –** A whole-of-government approach means… setting government-wide framework and policy, standards, guidelines, practices, and tools to ensure the quality of information and integration.
>
> **Access and privacy –** Respect for individual privacy applies across the information life cycle in accordance with the *Privacy Act*.
>
> **Security –** Ensuring the confidentiality, integrity, and availability of information is essential to government decision making and the delivery of services.
>
> **Transparency –** Employees document actions and decisions in support of government programs and activities, and maintain information so that it is accessible to anyone who is authorized to have

[91] Treasury Board of Canada Secretariat, 2007. *Policy Framework for Information and Technology* Accessed March 2011 (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12452&section=text)

> access, including those individuals exercising their rights to access information under the *Access to Information Act* and the *Privacy Act*.

The *Policy Framework* includes implicit references to the concepts of data archiving and preservation that apply to geospatial information. For example, the Stewardship guiding principle states, "information must be rigorously managed throughout its life cycle, regardless of medium or format…", and the Access and privacy guiding principle also refers to the "information life cycle". In addition, information management is defined as "a discipline that directs and supports effective and efficient management of information in an organization, from planning and systems development to disposal or long-term preservation"[92].

4.2 Policy on Information Management[93]

The TB *Policy on Information Management (IM Policy)* also came into effect in July 2007, with the objective "to achieve efficient and effective information management to support program and service delivery; foster informed decision making; facilitate accountability, transparency, and collaboration; and preserve and ensure access to information and records for the benefit of present and future generations". Three results are to be obtained through the application of this policy:

- convenient access to relevant, reliable, comprehensive and timely information through government programs and services;
- management of information and records as valuable assets used to support the outcomes of programs and services, as well as operational needs and accountabilities; and
- ensured continuous and effective management of information through the use of effective governance structures, mechanisms and resources.

An informative presentation that contains details of the Government of Canada's information management strategy outcomes framework can be found on the TBS Web site.[94] There are several other TBS documents that are affiliated with this policy, which are dealt with separately below.

While the *IM Policy* does not explicitly mention the requirement for information archiving and preservation, that concept is clearly inferred in several passages. For example, the wording of the last part of the policy objective statement is within that context. In addition, "integrating information management requirements with technology planning ensures that digital

[92] Ibid.
[93] Treasury Board of Canada Secretariat, 2007. *Policy on Information Management* Accessed March 2011, (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12742&section=text)
[94] Treasury Board of Canada, *Chief Information Officer Branch, Treasury Board of Canada Secretariat, RDIMS #794556*, Accessed March 2011. (http://www.gtec.ca/2009/presentations/conference/1940-henry.pdf)

information is accessible, shareable, and usable over time and through technological change"[95] suggests an archiving and preservation intent. The policy also includes Library and Archives Canada's responsibilities as playing a role in its implementation. The definition of information management in Appendix A to the policy includes the words "effective and efficient management of information … from planning and systems development to disposal or long-term preservation"[96].

The *IM Policy* makes no specific reference to geospatial data, databases or maps, but the policy's definition of a record, "information created, received, and maintained by an organization or person for business purposes, legal obligations, or both, regardless of medium or form", encompasses this type of data. Furthermore, the *IM Policy* states that "information management encompasses records, as well as documents, data, library services, information architecture, etc". Table 3 highlights the relevance of the Deputy Head's responsibilities[97] to geospatial information.

Table 3: Relevance of Geospatial Information to Policy Requirements

TBS IM Policy Requirements[98]	Relevance to Geospatial Information
6.1.1 ensuring that departmental programs and services integrate information management requirements into development, implementation, evaluation, and reporting activities	Geospatial information should be part of an integrated information management strategy along with other information, but it is unclear how well this is being accomplished within federal producer organizations.
6.1.2 ensuring that decisions and decision-making processes are documented to account for and support the continuity of departmental operations, permit the reconstruction of the evolution of policies and programs, and allow for independent evaluation, audit, and review;	If geospatial information is used to support operations, policies and programs, it is important to ensure that it is maintained as part of a record set. GeoConnections is assisting in this regard by encouraging that geospatial information is accompanied with metadata, description of data collection methodologies, data quality parameters, contextual information and any other attributes deemed necessary by the creators and maintainers of those datasets to assist with the understanding of them. If used in an experiment, then the parameters of the experiment should also accompany the geospatial information.
6.1.3 ensuring that information is shared within and across departments to the greatest extent possible, while respecting security and privacy requirements;	Another area where GeoConnections is contributing through the creation of data discovery and access mechanisms (e.g., portals, catalogs, etc.)
6.1.4 ensuring that all information is managed to respect user	GeoConnections, ESS and NRCan more broadly have extensive experience in this area as it pertains to geospatial information. The

[95] Ibid.
[96] Ibid
[97] Treasury Board of Canada Secretariat *supra*, note 93
[98] TBS Policy on Information Management, Policy Requirements (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=12742#cha6)

TBS IM Policy Requirements[98]	Relevance to Geospatial Information
agreements, licensing conditions, or both and for ensuring the relevance, authenticity, quality, and cost-effectiveness of the information for as long as it is required to meet operational needs and accountabilities;	*Unrestricted Use Licence Agreement, Best Practices for Sharing Sensitive Environmental Geospatial Data,* and *The Dissemination of Government Geographic Data in Canada - Guide to Best Practices* are excellent examples of data dissemination practices. The *Atlas of Canada* is an example of how multiple licenses are managed within one product, and the adoption of the ISO 19115 metadata standard ensures that data are well described. The field of geomatics and the scientific practices of data collectors generally ensures creation of data to meet user needs according to well established methodologies and practices, although these may not be made obvious in all cases.
6.1.5 ensuring electronic systems are the preferred means of creating, using, and managing information;	Most new geospatial information is born digital although there are many legacy datasets in non-digital formats (e.g., surveyor field notes, mineral collections, etc.).
6.1.6 ensuring departmental participation in setting government-wide direction for information and recordkeeping;	The *TBS Standard on Geospatial Data* is a good first step, as has been the work of the IACG and the CCOG in setting government-wide direction. However, GeoConnections and NRCan do not have an overarching data management and preservation policy, although they have long term data collection, management and dissemination experience. Also, it is uncertain how geospatial information is integrated with recordkeeping.
6.1.7 designating a senior official to represent the deputy head to the Treasury Board of Canada Secretariat for the purposes of this policy;	It is presumed that the divisions within NRCan and other producer organizations with primary responsibility for recordkeeping have representatives to TBS. However, it is probable that geospatial information is not part of those representatives' expertise, particularly since it is uncertain how geospatial information is included in terms of recordkeeping generally.
6.1.8 establishing, measuring and reporting on a departmental program or strategy for the improvement of the management of information; and	It is presumed that the recordkeeping sections of NRCan and other geospatial information producer organizations do this.
6.1.9 informing the Treasury Board of Canada Secretariat of their departments' participation in developing national and international information management standards as those activities relate to this policy.	Since GeoConnections has been very proactive and effective on this front, it is presumed that the recordkeeping sections of NRCan and other geospatial information producer organizations do this, and the *TBS Standard on Geospatial Data* is evidence that this is happening.

4.3 Policy on Management of Information Technology[99]

The objective of the TBS *Policy on Management of Information Technology (IT Policy)* is to "achieve efficient and effective use of information technology to support government priorities and program delivery, to increase productivity, and to enhance services to the public".

The *IT Policy* makes no specific reference to IT requirements related to data or to specific fields such as science or geomatics. The Policy also makes no reference to archiving or preserving IT, interoperability, source code and standards. This could be problematic with respect to geospatial data, databases and maps since these data are often inseparable from the systems or software within which they are collected, maintained, disseminated and rendered. For example, the *Atlas of Canada*, cybercartographic atlases and climate change models are all inseparable from their software. This may also be the case with data collected in real- or near real-time since these data may in fact be inseparable from both the software, system and the hardware (e.g., the sensors) that collect and disseminate the data. In many instances it will be critical to preserve the data, the software and potentially some hardware. However, the adoption of common specifications and standards as it pertains to access, interoperability, open architecture, open source, data formats, metadata, and management can enable IT heterogeneity while ensuring interoperability and hopefully longevity.

4.4 Directive on Information Management Roles and Responsibilities

The objective of the TBS *Directive on Information Management Roles and Responsibilities*[100], which has also been in effect since October 2007, is "to identify the roles and responsibilities of all departmental employees in supporting the deputy head in the effective management of information in their department". This directive obligates the senior executives for IM to ensure that appropriate management direction, processes and tools are in place to retain the quality of information throughout the information life cycle. In this case, information life cycle is defined as encompassing "planning; the collection, creation, receipt, and capture of information; its organization, use and dissemination; its maintenance, protection and preservation; its disposition; and evaluation."[101]

Other clauses of this directive also infer the requirement for information archiving and preservation. For example, managers at all levels are responsible for analyzing information

[99] Treasury Board of Canada Secretariat, 2007. *Policy on Management of Information Technology*, accessed March 2011. (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12755)

[100] Treasury Board of Canada Secretariat, 2007. *Directive on Information Management Roles and Responsibilities*, accessed March 2011. (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12754&section=text)

[101] Ibid.

requirements, "who needs access to it; for what purpose and for how long",[102] and IM functional specialists must collaborate with managers in developing and operating processes, systems, standards and tools to address information life cycle needs. LAC's roles and responsibilities are also referenced in this directive.

4.5 Directive on Recordkeeping

The TBS *Directive on Recordkeeping*[103] came into effect in June 2009 and is to be used in conjunction with the previously described policy and directive. Its objective is to "ensure effective recordkeeping practices that enable departments to create, acquire, capture, manage and protect the integrity of information resources of business value in the delivery of Government of Canada programs and services". This directive also obligates senior IM officials to ensure that methodologies, mechanisms and tools are in place to support departmental recordkeeping requirements throughout the information life cycle. This responsibility includes:

- Identifying, establishing, implementing and maintaining repositories in which information resources of business value are stored or preserved in a physical or electronic storage space;
- Establishing, using and maintaining taxonomies or classification structures to facilitate storage, search, and retrieval of information resources of business value in all formats;
- Establishing, implementing and maintaining retention periods for information resources of business value, as appropriate, according to format;
- Developing and implementing a documented disposition process for all information resources; and
- Performing regular disposition activities for all information resources.[104]

4.6 Standard for Electronic Documents and Records Management Solutions

The TBS *Standard for Electronic Documents and Records Management Solutions*[105] supports the Policy on Information Management and the Directive on Recordkeeping. It took effect in July 2010, with the objectives of supporting efficient and effective management of information through the use of EDRM solutions and reducing their overall cost through standardization and economies of scale. The standard states that "EDRM solutions are automated systems used to manage, protect and preserve information resources from creation to disposition, … maintain

[102] Ibid.
[103] Treasury Board of Canada Secretariat, 2009. *Directive on Recordkeeping*, accessed March 2011. (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=16552)
[104] Ibid.
[105] Treasury Board of Canada Secretariat, 2010. *Standard for Electronic Documents and Records Management Solutions*, accessed March 2011. (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18910&section=HTML)

appropriate contextual information (metadata) and enable organizations to access, use and dispose of records (i.e., their retention, destruction or transfer) in a managed, systematic and auditable way"[106].

4.7 Standard on Metadata

The TBS *Standard on Metadata*[107] became effective in July 2010, with the objective "to increase the use of standardized metadata and value domains in support of the management of information resources". While this directive does not specifically reference archiving and preservation of information, its emphasis is on:

- improving information resources accessibility, sharing, authenticity, reliability, and integrity across departments; and
- increasing the ability of programs and services to share information efficiently and effectively between systems and across departments

The effective use of standardized metadata across government will improve the extent to which digital information records are accessible and useable for long term research. There is a linkage between this standard and the metadata standard prescribed in the following standard on geospatial data.

4.8 Standard on Geospatial Data

The TBS *Standard on Geospatial Data*[108] is the only TB document that relates specifically to the geospatial information domain. Coming into effect in June 2009, the objective of this standard is "to support stewardship and interoperability of information by ensuring that departments access, use and share geospatial data efficiently and effectively to support program and service delivery". While this standard does not explicitly reference digital information archiving and preservation, there is a linkage between it and LAC's *File Format Guidelines for Preservation and Long-term Access*[109]. These guidelines recommend the same format as included in the TB standard (ISO TC 211 ISO 19115 Geographic Information - Metadata (NAP – Metadata) (North American Profile)) for preservation of and long term access to digital geospatial information held by government organizations.

---

[106] Ibid.

[107] Treasury Board of Canada Secretariat, 2010. *Standard on Metadata*, accessed March 2011. (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909&section=text)

[108] Treasury Board of Canada Secretariat, 2009. *Standard on Geospatial Data*, accessed March 2011. (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?evttoo=X&id=16553&section=text)

[109] Library and Archives Canada, *File Format Guidelines for Preservation and Long-term Access*, accessed March 2011. (http://www.collectionscanada.gc.ca/digital-initiatives/012018-2200-e.html)

5. Canadian Consultations and Reports on Digital Data Archiving and Preservation

This chapter provides a summary of work that has been undertaken in Canada to address the issues of digital data archiving and preservation, primarily in the scientific data context. It identifies overall challenges and gaps in current policy and programming and recommendations that have been made for improvement, which apply to geospatial data as well.

5.1 Introduction

A number of Government of Canada studies, reports and committees have made high level recommendations and suggested strategies for the archiving and preservation of digital data in Canada. All recognize the poor state of Canada's digital data resources. A few of the following reports focus on data, some make explicit reference to or are about geospatial data while others are high level reports that discuss national digital strategies, which include the preservation of Canada's digital assets and frame the direction of Government discourse. Of note is the consensus that:

- a national distributed trusted digital repository or infrastructure or system is required;
- good record management practices should occur when and where data are created;
- institutions should have a data management policy in place; and
- the lack of data discovery mechanisms is a major issue as is data access which includes cost, licensing, and copyright among many other issues.

There are also frequent recommendations to adopt a life-cycle model, standards, open specifications, interoperability and to appoint responsible data stewards or custodians within data producing institutions.

5.2 Digital Economy and Canadian Digital Information Strategy Consultations

The 2010 *Digital Economy Consultation*[110] conducted by Industry Canada and the 2007 *Canadian Digital Information Strategy (CDIS)*[111] conducted by Library and Archives Canada (LAC) are two national scale public consultations that remain incomplete, in the sense that neither has consolidated submissions into a cohesive report accompanied with recommendations and a series of strategies to be implemented. The Industry Canada *Consultation Paper*[112] indicates that the Government understands that a pan-Canadian network of trusted digital repositories (TDRs) could be one of the infrastructural strategies to protect Canadian content for the long term. This paper also mentions improving access to research data, data discovery and open access publishing but makes no specific mention of the geomatics sector or geospatial data, while science is mentioned only within an innovation and R&D context.

Alternatively, the LAC *CDIS* report reflects the broad consensus reached on the elements of a national strategy at a *National Summit*[113] held in 2006. While the consultation part was never completed, over 200 specialists contributed to the process, some of which are involved in geospatial data creation, use, dissemination and management. There are numerous suggested actions listed in the report, none of which are specific to geospatial data, although GeoConnections, GeoGratis, the Canada Ice Service Data Archive and GeoBase are lauded for some of their initiatives. Environment Canada, Statistics Canada, Fisheries and Oceans and Natural Resources Canada are mentioned as institutions that produce and contain significant collections of government data.

Some of the assumptions upon which the *CDIS* was premised are particularly relevant to the preservation of geospatial data. For example, that: technological change is constant; a distributed strategy will require support from stakeholders from multiple disciplines; interoperability and open access strategies are key; information access supports Canada's social and economic goals; and the model must be distributed. In addition, the following proposed actions are related:

1. Encourage communities of practice to develop standards-based and interoperable production practices, processes, infrastructure and tools for digital content production;

[110] Industry Canada, 2010, *Digital Economy Consultation*, accessed February 2011, http://de-en.gc.ca/2010/11/22/minister-clement-updates-canadians-on-canadas-digital-economy-strategy/

[111] Library and Archives Canada, 2007, *Canadian Digital Information Strategy*, Accessed February 2011, http://www.lac-bac.gc.ca/obj/012033/f2/012033-1000-e.pdf

[112] Industry Canada, 2010, *Improving Canada's Digital Advantage Strategies for Sustainable Prosperity: Consultation Paper on a Digital Economy Strategy for Canada*, http://de-en.gc.ca/wp-content/uploads/2010/05/Consultation_Paper.pdf

[113] Library and Archives Canada, 2006, *Toward a Canadian Digital Information Strategy: National Summit*,accessed February 2011 http://www.collectionscanada.gc.ca/obj/012033/f2/012033-611-e.pdf

2. Foster the adoption of recognized, open standards, and the development and sharing of best practice guidance and of standards-based tools, within communities of content creators;
3. Ensure effective implementation of a system for persistent identification of digital resources;
4. Develop technical capacity among digital content producers for automated transmission to Trusted Digital Repositories;
5. Develop and implement comprehensive e-records strategies that address policy, regulatory instruments, standards, and systems for government information production and management;
6. Define and establish, on a national basis, roles and responsibilities for digital information capture and long-term preservation by broad category/type of information;
7. Create TDRs and data archives on a national scale and build the requirement to archive to a digital repository, and support the costs of that process, into funding programs that produce digital content;
8. Develop new competencies and positions such as 'digital curators' who would have stewardship responsibility for digital information, whether in an institutional setting or as part of research teams;
9. Encourage development of specialized aggregation services and advanced research and knowledge discovery tools (e.g. for text and data mining); and
10. Review policy and licensing practices for Crown copyright with the view to facilitate access, use and re-use of public sector information and content, to unify licensing policy across the public sector, and to remove cost recovery-based barriers to access.

5.3 Toward a National Digital Information Strategy: Mapping the Current Situation in Canada

The LAC 2006 *Toward a National Digital Information Strategy: Mapping the Current Situation in Canada*[114] states that "the stewardship of digital information produced in Canada is disparate and uncoordinated" and "the volume, diversity and complexity of digital information is growing exponentially" while the "technologies, standards and practices that will better ensure the ongoing accessibility and integrity of digital information are not yet consistently applied". The report includes a brief description of the GeoConnections program and lists a number of

[114] MacDonald, J. and K. Shearer, 2006, *Toward a National Digital Information Strategy: Mapping the Current Situation in Canada*, Ottawa: Library and Archives Canada accessed February 2011 Summary: http://www.lac-bac.gc.ca/digital-initiatives/012018-3200-e.html and full Final Report: http://www.collectionscanada.gc.ca/obj/012018/f2/012018-3200-e.pdf

other geospatial data initiatives at different levels of government in Canada in its Annexes. The report does not provide specific recommendations on how to manage these resources but it does suggest that the:

> preservation of digital information requires a management response, not just a technological response. The response must be active and sustained through time and it must address the preservation of digital information from both the strategic and tactical perspectives and within the context of the management frameworks that govern the businesses of organizations themselves. It must also be comprehensive and address all facets of the required preservation infrastructure: the policies that assign accountability, the standards, practices, procedures and technologies that enable the implementation of preservation strategies, and the people that make it all happen. These issues present a tremendous challenge and must be addressed in an inclusive and collaborative manner.[115]

In this report, MacDonald and Shearer also provide a number of next steps toward the creation of a national digital information strategy, which very much resemble some of the collaborative strategies used to create the CGDI (e.g. multisectoral, departmental and multidisciplinary collaboration, etc.).

5.4 National Consultation on Access to Scientific Data

The 2005 *National Consultation on Access to Scientific Data Final Report (NCASRD)*[116], developed through a partnership between the National Research Council Canada (NRC), the Canada Foundation for Innovation (CFI), Canadian Institutes of Health Research and Natural Sciences and Engineering Research Council (NSERC), expressed the concern that:

> no national data preservation organization exists, nor does Canada have any national data access strategy or policies. Participants at the NCASRD expressed considerable concern about the loss of data, both as national assets and definitive longitudinal baselines for the measurement of changes over time. These losses occur as a result of storage media degradation, media and metadata loss, and software and hardware obsolescence, as well as privacy policies and decisions (e.g., by research ethics committees), and a lack of planning or attention to preservation beyond the individual researcher or organization.[117]

This *NCSARD* report provides a comprehensive list of recommendations that include organizing a Data Force and the creation of a Data Canada secretariat. There are also recommendations related to ethics, education, copyright, human resources, reward structures and resources among others, which would lead toward the creation of a national digital data strategy and archive.

---

[115] MacDonald, J. and K. Shearer, 2006, *Toward a National Digital Information Strategy: Mapping the Current Situation in Canada*, Final Report: http://www.collectionscanada.gc.ca/obj/012018/f2/012018-3200-e.pdf p.50.

[116] Strong, F. D. and p. B. Leach, 2005, *The Final Report of the National Consultation on Access to Scientific Data*, Ottawa: Government of Canada, http://ncasrd-cnadrs.scitech.gc.ca/docs/NCASRDReport.pdf

[117] Strong, F. D. and p. B. Leach, 2005, The Final Report of the National Consultation on Access to Scientific Data, Ottawa: Government of Canada, http://ncasrd-cnadrs.scitech.gc.ca/docs/NCASRDReport.pdf page 2.

5.5 Research Data Canada

One of the tangible outcomes of the *NCSARD* report was the creation of *Research Data Canada*[118] currently housed at the National Research Council of Canada. *Research Data Canada* is a collaborative effort which aims to address the challenges and issues surrounding the access to and preservation of data created by Canadian researchers. It is a multidisciplinary group of universities, institutes, libraries, granting agencies, and individual researchers who recognize the need to deal with Canadian data management issues. Membership in *Research Data Canada* includes several officials involved with the creation, management, dissemination and/or funding of geospatial data, namely: ***John Broome***, the Chair CNC/CODATA and Head, Data Management Policy and Strategy, Earth Sciences Sector Natural Resources Canada; ***Scott Tomlinson***, Data Management Coordinator, International Polar Year Federal Program and ***Chuck Humphrey***, Data Library Coordinator, Libraries, University of Alberta. While the focus is on research data the ideas can be applied to the preservation and management of data more broadly and specifically to geospatial data. The following *Research Data Canada* approach and principles (shown in Table 4) to data stewardship have been echoed in many other initiatives discussed elsewhere in this report:

> Data stewardship as asset management needs to be framed in a life cycle model, embodying all stages of the research process and engaging all with a responsibility for data preservation including researchers, research institutions, funding agencies, and government.[119]

Table 4: Research Data Canada Five Principles of Data Stewardship

Five Principles of Data Stewardship	Description[120]
Principle 1: Roles and Responsibilities	Establishment of a framework of codes of practice for access to digital research data. This framework must define, as clearly as possible, the roles and responsibilities of the various parties involved in data-related activities. Creators and users of research data as well as responsible institutions and funders should be aware of and fulfill their responsibilities in accordance with these codes of practice.
Principle 2: Interoperability, standards and quality assurance	The value and utility of research data depends, to a large extent, on the quality of data and their interoperability with other data. In order to promote international and interdisciplinary use of research data, data should be created and collected in accordance with international standards and where applicable, interoperability of datasets should be a key consideration. If standards do not yet exist, stakeholders should actively contribute to their development. Criteria for selection and long-term retention of data should take into consideration quality assurance issues.
Principle 3:	Digital research data should be easy to find, and access should be provided in

[118] Research Data Canada, http://ncasrd-cnadrs.scitech.gc.ca/eng/index.html
[119] Research Data Canada, http://ncasrd-cnadrs.scitech.gc.ca/eng/about/principles-data-stewardship.html
[120] Research Data Canada, 5 Principles, http://ncasrd-cnadrs.scitech.gc.ca/eng/about/principles-data-stewardship.html

Five Principles of Data Stewardship	Description[120]
Access, usage and credit	an environment which maximizes ease of use; provides credit for and protects the rights of those who have gathered or created data; and protects the rights of those who have legitimate interests in how data are made accessible and used. The fundamental policy objective of making data available and accessible requires that users should have timely, user-friendly access to relevant data, and at the lowest possible cost. There is a need to balance conflicting interests and rights, but wherever possible, free and open access should be the default option.
Principle 4: Benefits and cost effectiveness	The models and mechanisms for managing and providing access to digital research data must: • Safeguard the public investment in the original creation of the data • Provide for cost effective on-going technological and operational management of the stored data
Principle 5: Preservation and sustainability	Digital research data of long-term value arising from current and future research should be preserved and remain accessible for current and future generations. In order to accomplish there must be a collective effort: • Develop criteria for selection and long-term retention of data • Develop a network of sustainable repositories to host and maintain data in the long-term • Ensure access to and re-use of data in research that would comply with regulatory requirements- or should we first create those regulations? • Anticipate adapting and moving data to new electronic environments as they emerge • Develop transparent and practical policies and procedures that would enable storing, sharing, re-use, of stored data.

5.6 SSHRC National Data Archive Consultation

The 2002 *SSHRC National Data Archive Consultation*[121] is *NSCARD's* predecessor and primarily discusses data in the social sciences and humanities and the preservation of data created in the course of state funded research projects. The *Consultation* identified important institutions, infrastructures, management frameworks and data creators and calls for the creation of a national research data archive. The Report's appendices include some useful research: International Models of Data Archiving Services where 36 social science and humanities data preservation and related services are compared and an Outline of Institutional Options for a National Research Data Archive. While the institutional context has changed somewhat in Canada and internationally, these do provide ways of reflecting upon this topic from an institutional perspective.

[121] Social Science and Humanities Research Council, 2002, *Final Report: National Consultation on Research Data Archiving, Building Infrastructure for Access to and Preservation of Research Data*, http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/da_finalreport_e.pdf

5.7 Stewardship of Research Data in Canada: A Gap Analysis

In 2008, the study *Stewardship of Research Data in Canada: A Gap Analysis*[122] was conducted by members of *Research Data Canada*. The life-cycle model is used as the framework to analyze the current state of research data stewardship in Canada as compared to an ideal state. The life-cycle model they refer to includes data production, dissemination, long-term preservation and data discovery and repurposing. The *Gap Analysis* included a review of: policies; funding; roles and responsibilities; repositories; standards; skill and training; rewards and recognition systems; R&D; access; and preservation. Not surprisingly, the analysis concludes that:

> gaps exist for a number of reasons: because of a lack of policies, infrastructure, and funding mechanisms for data stewardship, as well as a research culture that does not recognize the value of data management. Data stewardship in Canada is suffering without a strategic, national vision. The existing piecemeal approach has resulted in serious gaps throughout the lifecycle. This is particularly apparent in the final three stages. Significant amounts of data are rendered inaccessible at the data dissemination stage because of the absence of services and procedures to deliver data to other researchers. The woefully inadequate number of trusted data repositories in Canada contributes to the gap identified in the long-term management stage. Consequently, most research data created in Canada are greatly underutilized and are at a high risk of being lost[123].

Beyond the fact that many but not all research datasets do not have a spatial referent, there are many differences between research data and geospatial data, namely: much of Canada's research data are paid for with public research funds, responsibilities lie with the Primary Investigator, they are often created outside of government, research projects can often be one-off, and they do not have the same temporal issues in terms of frequent updates as seen in many geospatial datasets. Otherwise both geospatial and research data are complex, are in multiple data formats, use non standardized rendering techniques and specialized open and/or proprietary software, are disseminated in portals, catalogues or gateways, or are shared and visualized in distributed systems, datasets can be very large, data can be modeled and there are no uniform metadata and data quality standards.

The similarities between the context of scientific and geospatial data suggest that most of the gaps and conclusions of the *Gap Analysis* and the work of *Research Data Canada* more broadly, can be transferred to the geospatial data environment. Also, issues associated with scientific data in general are analogous to the geospatial data context since geospatial data creators are also scientists who create specialized and complex datasets stemming from the

[122] Research Data Canada, 2008, *Stewardship of Research Data in Canada: A Gap Analysis*, Accessed February 2011, http://ncasrd-cnadrs.scitech.gc.ca/docs/GapAnalysis.pdf

[123] Research Data Canada, 2008, *Stewardship of Research Data in Canada: A Gap Analysis*, page 17, accessed February 2011, http://ncasrd-cnadrs.scitech.gc.ca/docs/GapAnalysis.pdf,

myriad heterogeneous specializations found in the geomatics sector, as discussed in Chapter 7, which shows but a small subset of the types of geospatial data that claim to be archived.

The *Gap Analysis*[124] lists a number of gaps and issues, and only those most relevant to the geospatial data context are listed here:

- that there are large policy gaps throughout the data lifecycle with policies addressing the need for researchers to develop data management plans;
- some government departments have responsibilities for collecting and preserving specific data types: Statistics Canada, Natural Resources Canada, Canadian Space Agency, Environment Canada;
- there are no lines of custodial responsibilities along the data lifecycle and no mechanisms for bringing
- There is a very nascent network of institutional repositories being managed by research libraries, but few are capable of collecting research data in a manner that facilitates easy access and re-use
- Many researchers are unfamiliar with data stewardship processes, including the importance of metadata.
- There is a reticence amongst many to assume responsibility for data stewardship beyond the researchers' immediate interests
- There are insufficient numbers of trained scientists and information professionals with knowledge of data cataloguing, metadata standards and processes, preservation management and assessing the value of data to support researchers.
- A few projects, such as the International Polar Year, require data management plans from researchers in order for projects to receive funds.
- The research cultures in many domains do not embrace data preservation or data sharing. Consequently, researchers do not understand the risk levels associated with their current practices. Reward systems for researchers do not recognize sound data management or data sharing. There is little recognition for leadership in the compilation of, or major contribution to, high value, open access databases and datasets, nor in the development of tools that enhance the value of data (for example, developing methods for mining or combining databases across disciplines).
- Numerous research and development challenges remain in the field of data stewardship. Research into technologies, organizational models, standards and practices, and interoperability are needed on an ongoing basis as the volume of data grows and becomes more complex. R&D projects are not coordinated, nor guided by national priorities.

[124] Research Data Canada, 2008, *Stewardship of Research Data in Canada: A Gap Analysis*, Accessed February 2011, http://ncasrd-cnadrs.scitech.gc.ca/docs/GapAnalysis.pdf

- Researchers often do not have the time or skills to prepare data for dissemination, are reluctant to share data and are unclear about who owns data. Few reward or recognition mechanisms exist for sharing research data.
- Few research organizations have policies regarding preservation of research data. Where policies exist, they do not cover all types of data.
- Where preservation activities do exist, quality control, data storage and backup, and descriptive metadata are the most commonly cited practices

5.8 CODATA Working Group on Archiving Scientific Data and OGC Data Preservation Working Group

Two international groups hold potential for future guidance in digital information archiving and preservation. The Committee on Data for Science and Technology (*CODATA) Working Group on Archiving Scientific Data*[125] has been holding symposia and workshops on the topic and the *Canadian National Committee for CODATA*[126] has been active in documenting and reporting scientific data activities. The Open Geospatial Consortium (OGC) also has a *Data Preservation WG* [127] which is "to address technical and institutional challenges posed by data preservation, to interface with other OGC working groups that address technical areas that are affected by the data preservation problem, and to engage in outreach and communication with the preservation and archival information community". Not much reporting of results has come from this WG.

5.9 GeoConnections Report: Archiving, Management and Preservation of Geospatial Data

In 2005 GeoConnections conducted a study on *Archiving, Management and Preservation of Geospatial Data*[128] which provided a well rounded analysis of geospatial data preservation issues, such as: technological obsolescence; formats; storage technologies; temporal

[125] *CODATA Preservation of and Access to Scientific and Technical Data in Developing Countries*, 2006, Home Page, accessed March 2011. http://www.codata.org/taskgroups/TGpreservation/

[126] *Canadian National Committee for CODATA*, accessed March 2011. http://dac.cisti.nrc.ca/

[127] *OGC, Data Preservation WG*, accessed March 2011. http://www.opengeospatial.org/projects/groups/preservwg

[128] GeoConnections, 2005, *Archiving, Management and Preservation of Geospatial Data: Summary Report and Recommendations*, Produced by David L. Brown, Grace Welch and Christine Cullingworth as an initiative of the Policy Advisory Node. http://www.geoconnections.org/publications/policyDocs/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf

management; and metadata[129]. The Study also provided a list of technological preservation solutions with their associated advantages and disadvantages. This report recommends that "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"[130]. The recommendations included in the report are somewhat outdated and would need to be updated in light of changes in the direction of the GeoConnections program and the Earth Science Sector, and of the new TBS Directives and Guidelines. Irrespective of these shortcomings, the *Study* provides some useful recommendations that merit being listed as shown in Table 5.

Table 5: Recommendations for the Preservation of Geospatial Data, GeoConnections Report

Recommendations for the Preservation of Geospatial Data	
1. Recommendations for Institutional-based Action	**Recommendation 1.1** Organizations should define and implement policies and practices for the creation, use, retention, dissemination, preservation, and disposition of geospatial data. Building a business case for the creation of core geospatial data products is suggested.
	Recommendation 1.2. 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. A custodianship model is an option which 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: **1.2.1.** Create a corporately approved data management strategy and plan that identifies data holdings, and provides standards for data collection, management and preservation practices; **1.2.2.** Committee or role: to act as the organization's data custodian(s) to ensure that policies and procedures for geospatial data management are defined, understood and implemented; **1.2.3.** Create a data assessment fact sheet to assist in the identification of valuable data that requires long-term retention; **1.2.4.** Assign the value of a data object at the time of creation based on operational and preservation requirements; **1.2.5.** Provide detailed metadata based on internationally and national approved

[129] The approach used in the Report adopts the framework created by 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. Accessed February 2011, http://prod.sandia.gov/techlib/access-control.cgi/2002/020107.pdf

[130] GeoConnections, 2005, *Archiving, Management and Preservation of Geospatial Data: Summary Report and Recommendations*, http://www.geoconnections.org/publications/policyDocs/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf Page i .

Recommendations for the Preservation of Geospatial Data	
	metadata standards. **1.2.6.** Inventory existing digital holdings and assess their value for long-term retention and preservation and define a time-table for migration and media refreshing; **1.2.7. S**tandardize data formats and widely disseminate this information within the organization; **1.2.8.** Adopt industry standards such as ISO or open source solutions that liberate the data from proprietary formats.
2. Recommendations for National Action	**Recommendation 2.1** 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. IACG, CCOG, LAC and CGSB should work together to develop the following policy areas: **2.1.1.** Promote the use of the approved metadata standard in the geomatics community through publicity, workshops, and the creation of tools; **2.1.2** Promote the adoption of non-proprietary standards for the creation and exchange of geospatial data such as GML and Open Geospatial Consortium specifications **2.1.3.** Promote the use of LAC guidelines and best practices for the logical and physical storage of digital information; **2.1.4.** Develop checklists, guidelines and other practical tools for the selection of data for preservation **2.1.5.** 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; **2.1.6.** Investigate and recommend sustainable business models for preservation that promote cooperation to share the responsibility and costs associated with preservation activities **2.1.7.** 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 2.2 CCOG should 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.

5.10 Summary

To date none of the above consultations, reports, committees or recommendations have resulted in the creation of a national geomatics data archive nor have new policies yet been implemented. Currently, Library and Archives of Canada (LAC) does not have a digital geospatial data archive, nor is there a national network of trusted digital repositories with the explicit mandate to ingest geospatial data, research or government created. As previously discussed, LAC does expect to have a TDR ready to ingest data sometime in 2017[131].

[131] Interview at LAC, February 11, 2011and *LAC Trusted Digital Repository,* Accessed February 2011, http://www.collectionscanada.gc.ca/digital-initiatives/012018-4000.01-e.html

6. Canadian Research on Digital Data Archiving and Preservation

6.1 Introduction

The technological, institutional and organizational issues related to the long-term preservation of digital maps, atlases and their associated geospatial data remain under-researched and largely unresolved. However, we are beginning to see a few examples of policies (e.g., in DFO) and practices where geospatial data are being preserved. There has been very little archival research done in Canada which focuses solely on digital geospatial data (or digital data in general), while this has been a topic of discussion for quite some time. According to Shearer and MacDonald, "the lack of research is particularly acute in the area of digital information preservation"[132]. Beyond the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project[133] at the University of British Columbia, the only other universities in the country that appear to be addressing this issue are the University of Toronto and the Université de Montréal.

6.2 InterPARES

6.2.1 Research Findings and Analysis

Ongoing since 1999, InterPARES is an interdisciplinary research project that studies the life-cycle of digital records created in complex digital environments, from creation to permanent preservation. The focus is on records produced in the course of artistic, scientific and government activities. InterPARES has produced ground-breaking research on the theory and practice of the preservation of dynamic, interactive, and experiential records. InterPARES has produced reports, guidelines, preservation models, requirements, case studies and general studies. One of the many outcomes of that research is the understanding that today's digital mapmakers, digital cartographers and geospatial data producers must include preservation strategies at the point of creation, not after the fact, to ensure that their digital artifacts will be available to tomorrow's users. If this is not done effectively much of our digital mapping heritage will continue to be lost.

[132] MacDonald, J. and K. Shearer, 2006, Toward a National Digital Information Strategy: Mapping the Current Situation in Canada, Ottawa: Library and Archives Canada accessed February 2011 Summary: http://www.lac-bac.gc.ca/digital-initiatives/012018-3200-e.html and full Final Report: http://www.collectionscanada.gc.ca/obj/012018/f2/012018-3200-e.pdf

[133] InterPARES Website (http://www.interpares.org/)

While this may seem to be very simplistic and obvious, this finding alone is changing the practice of archiving. Archiving used to be an afterthought and archivists were once only involved in managing records once these were deposited in the archive (e.g., after the death of famous scientists). The entire discipline and profession has been turned on its head with the advent of digital content. Archiving today must be trans-disciplinary and include record creators, IT specialists, data managers, scientists, geographers, archivists, librarians, database and software creators, hardware specialists, standards writers, metadata creators and so on. There is a demographic lag in today's archives, since the majority of archivists were not borne into the digital era, and there is also a human resources lag since there are few schools able to train young digital archivists. In addition, archives have not been able to develop the human and technological capacity to ingest the deluge of information being created, particularly, as seen in the case of science and geomatics, data created in ever changing, dynamic, interactive and complex environments.

Fortunately, we are beginning to see TBS record management policies, guidelines and directives discussing the life-cycle of records, which provide the framework within which geospatial data creators can manage their assets in the event that these are to be accessioned into an archive. Also, as previously discussed, the Archive has to reconsider its form. The quantity, complexity and nature of the data being produced is challenging the traditional deposition into the archive model and there is increasing discussion about a distributed archive model. It is uncertain how this will be played out, but it is possible to speculate that it may look like the Statistics Canada Research Data Centre (RDC)[134] model, where mini StatCan offices governed by all its statutes and staffed by StatCan employees are housed in a variety of institutions. The RDCs provide Canadian researchers with access to micro statistical data that are governed by very restrictive access and privacy rules. In the case of a distributed archive, it may very well be a mini LAC preservation office or a satellite trusted digital repository (TDR) within a department such as NRCan, which would be linked to the main TDR via a set of standards, practices, norms, technologies, catalogues, etc. This model would not be unfamiliar to NRCan since it created the Canadian Geospatial Data Infrastructure (CGDI) in a very similar fashion. Also, NRCan and other geospatial data creators have much experience with distributed mapping, the creation of databases where data are collected by multiple external sensors located all across the country, and distributed data portals and catalogues that are standards-based.

While it is not possible to resolve what a distributed archive would look like here, it is possible to look at the outcome of Canadian research projects such as InterPARES, which explored the preservation of digital data and from which important clues as to how to preserve digital geospatial data can be derived. The InterPARES 2 Project (IP2) research focused on themes of accuracy, reliability, authenticity, metadata and the term "record" as they pertain to scientific

[134] Statistics Canada, 2011, *The Research Data Centres*, Accessed March 2011 (http://www.statcan.gc.ca/rdc-cdr/index-eng.htm).

data.[135] In this E-Scan we will rely on the empirical evidence from a selection of IP2 Case Studies (see Table 6) that include geomatics data, and IP2 General Study 10 (GS10)[136], which investigated 32 scientific Data Portals of which 18 are geospatial data portals (see Table 7). The GS10 Survey was undertaken to collect information about the actual practices, standards, and protocols currently used by broadly defined existing data services, archives, repositories, or catalogues in the sciences, and in the case of this E-Scan, geography. These studies help provide a deeper understanding of practices in the natural and physical sciences as these pertain to portals and their associated data. These IP2 Case Studies[137] and General Studies[138] are particularly helpful, as these are available in their entirety on the InterPARES 2 website along with every other survey result and study report.

Table 6: Selected InterPARES 2 Case Studies

IP2 Case Study Name	Description	Observational Data	Computational Data
CS06 - Cybercartographic Atlas of Antarctica (Lauriault and Hackett 2005)	Online interactive and dynamic, open standards, interoperable multimedia, multisensory, multimodal Atlas that renders distributed data from myriad scientific organizations.	The atlas renders distributed observational data from myriad sources in real time. e.g., film, satellite images, tabular data, sounds, etc.	Data are rendered/refined into maps, charts, tables by the Nunaliit Atlas Framework
CS08 - Mars Global Surveyor Data Records in the Planetary Data System (Underwood 2005)	NASA Mars Global Surveyor mission data records at the Planetary Data System (PDS) Space Science Data Archive.	Level 0 and refined planetary spacecraft mission data stored in a database with attributes and metadata	Software used to access or visualize the data
CS14 - Coalescent Communities in Arizona (O'Meara, Pearce-Moses and Preston 2004)	Archaeological Records of the American Southwest rendered in a Geographical Information System (GIS).	Tabulated raw data collected from archeological digs	Raw data are rendered/refined into maps
CS18 - Computerization of Alsace-Moselle's Land Registry (Blanchette 2004)	Electronic registry including digital transcription of 40 000 existing paper registries and new database entries individually signed by a judge using a PKI infrastructure combining biometric access and digital signatures.	Database of digitized paper land registries and attributes	Data and their attributes are accessed via a data base using PKI technology

[135] The International Research on Permanent Authentic Records in Electronic Systems Project (InterPARES) http://www.interpares.org/ip2/ip2_index.cfm (accessed 6 September 2007)
[136] InterPARES 2, General Study 10 Report and associated documents, accessed March 2011 (http://www.interpares.org/ip2/ip2_case_studies.cfm?study=34)
[137] InterPARES 2 Case Studies, accessed March 2011 (http://www.interpares.org/ip2/ip2_case_studies.cfm)
[138] InterPARES 2 General Studies, accessed March 2011 (http://www.interpares.org/ip2/ip2_general_studies.cfm)

IP2 Case Study Name	Description	Observational Data	Computational Data
CS19 - Authenticating Engineering Objects for Digital Preservation (Hawkins 2005)	Examines through an engineering experiment the authentication of digital model (CAD) records using a content/message/semantic-based methodology rather than media, bit-count, or static provenancial attribute-based authentication.	CAD solid model files used in the design and manufacturing of mechanical piece-part assemblies	The files are rendered and stored in a proprietary system
CS24 - City of Vancouver Geographic Information System (VanMap) (McLellan 2005)	An enterprise web-based map system maintained by the City of Vancouver's Information Technology Department.	Land use, social statistics, city infrastructure data both internally collected and acquired from external sources etc.	Enterprise GIS renders data into maps, charts, and tables etc.
CS26 - Most Satellite Mission: Preservation of Space Telescope Data (Ballaux 2005)	Repository of the Microvariability & Oscillations of Stars (MOST) satellite mission data of Canada's first space telescope.	Raw satellite data and their refined counterparts	

While some of these case studies may not be considered to be geospatial in some circles, they are included here because they provide contexts that are very similar to those of the geospatial data creators or they refer to registries as mentioned in some of the acts and regulations discussed in Chapter 2.

Table 7: Geospatial Data Portals Examined as Part of IP2 General Study 10

Portals	Data Sets
British Atmospheric Data Centre (BADC)	Datasets produced by NERC-funded projects; some third party datasets that are required by a large section of the UK atmospheric research community and are most efficiently made available through one location.
World Data Center for Solar Terrestrial Physics	It includes International Geophysical Year (IGY) data, solar and interplanetary phenomena, ionospheric phenomena, flare-associated events, geomagnetic variations, and cosmic rays.
OBIS-SEAMAP (Ocean Biogeographic Information System - Spatial Ecological Analysis of Megavertebrate Populations)	Datasets include marine mammal, seabird, and sea turtle distribution and abundance.
Canadian Institute for Health Information (CIHI)	The data consist of health data provider demographics, health human resources, health spending and health services and health issues, drug, and hospital generated data.
Canadian Geospatial Data Infrastructure (CGDI)	Geospatial data such as DEM, orthophotos, air photos, satellite and radar imagery, data Services, discovery metadata, documents, Internet Maps,

Portals	Data Sets
Access Portal	and Atlases.
Statistics Canada	Under the Statistics Act Statistics Canada is required to collect, compile, analyze, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and conditions of the people of Canada.
Long Term Ecological Research (LTER)	Data of Holocene barrier island geology; salt marsh ecology, geology, and hydrology; ecology/evolution of insular vertebrates; primary/secondary succession; life-form modeling of succession; online maps, photos, webcams; physical; biological, images and geographic data, software, LTER Network Data; and models.
Southern California Earthquake Center (SCEC)	Data sets include seismic waveforms, mostly passive source seismic data collected by broad-band, strong-motion and analog instruments. LARSE I and II seismic survey, data sets from the portable deployments following the Landers and Northridge earthquakes, and data from the Anza network and SCEC borehole stations. Non-seismic data includes "survey-mode" precise GPS measurements made in southern California by various universities. The GPS data archive consists of raw GPS data, RINEX files, indices to the RINEX files, log sheets and site descriptions.
International Comprehensive Ocean Atmosphere Data Set (ICOADS)	Surface marine reports from ships, buoys, and other platform types. Each report contains individual observations of meteorological and oceanographic variables, such as sea surface and air temperatures, wind, pressure, humidity, and cloudiness; monthly summary statistics.
National Geophysical Data Center (NGDC - NOAA)	Geophysical data describing the solid earth, marine, and solar-terrestrial environment, as well as earth observations from space.
Antarctic Digital Database (ADD)	1:200,000/1:250,000 maps and collaborative topographic database compiled from a variety of Antarctic map and satellite image sources.
National Snow and Ice Data Center (NSIDC), NASA	Snow, ice and glaciological Data From satellite images remote sensing instruments, ground measurements, and data models.
U.S. Antarctic Resource Center (USARC)	Maps, orthophotos, satellite imagery, point data, Digital Elevation Models, Digital Raster Graphics, and aerial photography,
British Antarctic Survey - Antarctic Environmental Data Centre	Antarctic research results, climate modeling and predictions. Also, oceanic, environmental, atmospheric, geoscience, water, and earth observation data from a variety of sensors and in many forms.
Global Change Master Directory –Global Change Data Center	Data inform climate change research, and datasets include models, instruments and services. It is a work space as well as a place to store these working operational models. Disciplines include atmospheric science, oceanography, ecology, geology, hydrology, and human dimensions of climate change.
Community Data Portal at NCAR	CDP catalogs observational and computer simulation datasets, sources of data are related to sciences in the areas of: oceanic, atmospheric, space weather, and turbulence.
Earth Systems Grid (ESG) portal	Climate change models include High-resolution, long-duration simulations, data simulations, derived from UCAR/NCAR projects (particularly the Community Climate System Model (CCSM) and Parallel Climate Model (PCM) and the Intergovernmental Panel on Climate Change (IPCC). Also physical" datasets that are generated directly by

Portals	Data Sets
	climate simulations.
USGS Data Portals - GEO-DATA Explorer (GEODE)	Geologic discipline's programs including coastal and marine geology, earth surface dynamics, earthquake hazards, integrated natural resource sciences, mineral resources, National Cooperative Geologic Mapping, volcano hazards. Also scientific and energy related data.

6.3 Data Complexities

According to the National Science Foundation's 2005 Report of the National Science Board, data can be distinguished by how they were collected, such as: (1) observational; (2) computational; or (3) experimental.139 Observational data can be direct observations of ocean temperature on a specific date, the film footage of an Antarctic ice sheet breakup, pre-election polls, or photographs of a meteorite cloud, and can also be historical recordings of particular events that cannot be replicated nor recollected. Many of the data portals examined in IP2's GS10 include observational data, such as the British Atmospheric Data Centre or the World Data Centre for Terrestrial Physics. Computational data, such as the results from a climate change model that includes comprehensive information about the model (e.g., descriptions of the hardware, software, original input data, and metadata), may be reproducible and in this case they may be less important than the model itself. Experimental data "such as measurements of patterns of gene expression, chemical reaction rates, or engine performance, present a more complex picture."[140] These data may be reproducible; however, the cost of doing so is prohibitive, and it may not be possible to reproduce the exact same experimental conditions.

The National Oceanic and Atmospheric Administration (NOAA), which produces data for emergency preparedness and coastline management, organizes its data according to the following criteria: 1) Original Data; 2) Synthesized Products; 3) Interpreted Products; 4) Hydrometeorological, Hazardous Chemical Spill, and Space Weather Warnings, Forecasts, and Advisories; 5) Natural Resource Plans; 6) Experimental Products; and 7) Corporate and General Information. Since the data NOAA produces are very influential and the decisions based on them involve risk to human health and safety, each of these categories is very well described, data undergo rigorous error checks, there are specific data quality parameters, and NOAA adheres to the Office of Management and Budget (OMB) guidelines.[141] Data produced

[139] National Science Foundation, 2005, *Report of the National Science Board: Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*, Accessed March 2011, (http://www.nsf.gov/pubs/2005/nsb0540/nsb0540_4.pdf), p. 18.

[140] National Science Foundation, 2005, *Report of the National Science Board: Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*, Accessed March 2011, p.19.

[141] Federal agencies in the United States have responsibilities under the *Data Quality Act (Public Law 106- 554; H.R. 5658, Sec. 515)*. In accordance with this Act, the Office of Management and Budget (OMB) has issued guidelines that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by Federal agencies." Office of Management and Budget (OMB), *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility,*

in Canada under the Charts and Nautical Publications Regulations, 1995 would have the same degree of importance.

None of the IP2 Case Studies included experimental data, all included observational data in a variety of forms, while five of the studies also included computational data or models. All but the Cybercartographic Atlas[142] have their data rendered or stored in a proprietary system. All the data are stored in a variety of databases or in a searchable data portal. The Engineering project data[143] (CS19), the Land Registry[144] (CS18), and some of the NASA data (CS08) in particular are inseparable from the systems within which they have been created and/or stored.

This small sample of case studies and the surveyed data portals illustrate the complexities involved with the preservation of geospatial data and the particularities of each scientific practice.

6.3.1 Portals

Portals provide a framework within which geospatial data archivists can work and from which they can expand with policies, standards and metadata. It is important for archivists and data preservationists to remember that data producers and users have already appraised these data as being valuable enough to be paid for, collected, described, licensed, endorsed, organized and disseminated.

Geospatial data are increasingly being discovered and accessed via data portals. Portals have a variety of names, such as data repositories, clearinghouses, catalogues, archives, geolibraries and directories. In this E-Scan, the term portal is used to encompass all of these. Portals make it possible for users to "gather data germane to their own needs more readily, extract data from online and other electronic repositories, develop the information product they need, use the products for decision making, and contribute their locally gathered geoinformation and derived products to libraries or other repositories."[145]

As is commonly known in the field of geomatics, portals can provide all or some of the following services: search and retrieval of data; item descriptions; display services; data processing; the platform to share models and simulations; and the collection and maintenance of data. Much but not all of the data derived from portals are raw in nature and require the user to interpret, analyze and/or manipulate them. The reasons for portal creation are one-stop-

and Integrity of Information Disseminated by Federal Agencies; Notice; Republication, Federal Register, vol. 67 no. 36 (Washington DC, Friday, February 22, 2002), www.whitehouse.gov/omb/fedreg/reproducible2.pdf (Accessed March 2011).

142 Tracey P. Lauriault and Yvette Hackett, *CS06 Cybercartographic Atlas of Antarctica* (Vancouver, 2005).

143 Kenneth Hawkins, *CS19 Authenticating Engineering Objects for Digital Preservation* (Vancouver, 2005).

144 Jean-François Blanchette, Françoise Banat-Berger and Geneviève Shepherd, *CS18 Computerization of Alsace-Moselle's Land Registry* (Vancouver, 2004).

145 National Research Council, 1999. Spatial Information Resources, *Distributed Geolibraries* (Washington DC), accessed March 2011 (http://www.nap.edu/openbook.php?isbn=030906540)2, p.36.

shopping, distributed responsibility over data sets, discoverability, and reduction in cost as data are stored or described once and used many times.[146]

Clearinghouses, directories and catalogues are the technical embodiments of data-sharing policies. Individuals within organizations, research projects, or scientific collaborations register their data holdings in the portal via an online form organized according to a metadata standard, and then choose to make their data available for free or for sale, and viewing or downloading.[147] Metadata standards "establish the terms and definitions to provide a consistent means to describe the quality and characteristics of geospatial data,"[148] and the ISO 19115 metadata[149] standard has become an international standard in the field of geomatics. Most of the portals examined in IP2's General Study 10 include either very detailed metadata or rudimentary header information that only contains lineage information.[150]

The architecture of data portals varies. The *National Research Council, Spatial Information Resources, Distributed Geolibraries*[151] report indicates that portals can be a single enterprise sponsored portal (like a national library), a network of enterprises (like a federation of libraries) or a loose network connected by protocols (like the Web).[152] *Distributed data portals* have datasets described according to a given standard, and when a request is sent to them by a given site a search is executed by a search agent[153] to access or render the data into a map or some other form. The Cybercartographic Atlas of Antarctica, for example, adheres to interoperable Open Geospatial Consortium standards and specifications. When a user accesses a particular atlas module, a call is made to the British Antarctic Survey data portal in the United Kingdom, the data are accessed and then rendered into a map in real time by the Atlas Framework in Ottawa and delivered directly to the user's computer. Other examples of this type of portal are those which use Web mapping services, such as the British Antarctic Survey. A *Collection level catalog/portal* identifies a data custodian's holdings and uses them to direct searches.[154] Z39.50, a server sharing standard that enables searching multiple data holdings, is an example of a collection portal mechanism as is the Ocean Biogeographic Information System – Spatial Ecological Analysis of Megavertebrates Populations. A *unified catalogue* exists in one place: data custodians submit metadata for each data set to a central site which makes them available

146 Tracey P. Lauriault, 2003. *A Geospatial Data Infrastructure is an Infrastructure for Sustainable Development in East Timor*, (Master's Thesis, Carleton University).
147 Ibid.
148 Nancy Tosta and Michael Domaratz,1997. The U.S. National Spatial Data Infrastructure, *Geographic Information Research: Bridging the Atlantic*, ed. Massimo C. Craglia and Helen Couclelis, London, p. 22.
149 International Standards Organization, 2003. *Geographic information – Metadata*, accessed March 2011, (http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35&ICS2=240&ICS3=70).
150 See Table 3 Metadata on the InterPARES 2 website at http://www.interpares.org/display_file.cfm?doc=Archivaria64_Todays_Data_supplementary_tables.pdf
151 National Research Council, 1999, *Spatial Information Resources, Distributed Geolibraries*
152 National Research Council, 1999, *Spatial Information Resources*, p. 65-66.
153 Ibid.
154 Ibid.

for searching, and the record directs the user to the data set.[155] The GeoConnections Discovery Portal (IP2SF15) is an example of this type of portal.

Digital collections/portals can be housed in a single physical location (e.g., Statistics Canada), and they may be virtual, housed in a set of physical locations and linked electronically to create a single, coherent collection (e.g., Global Change Master Directory, International Comprehensive Ocean Atmospheric Dataset). The distinction between centralized, distributed or unified portals may have funding, policy and preservation implications.

Data collections may also differ because of the unique policies, goals, and structure of the funding agencies. Collections created and maintained by government data centres such as the NASA Space Mission Data (CS08), GeoConnections, and specific research projects such as the MOST Satellite Mission[156] (CS26) each pose unique challenges for policy makers.

There are three functional data collections/portal categories are: 1) research data collections; 2) resource or community data collections; 3) reference data collections. These are not rigid categories, particularly since some research data collections are also resource collections. These collections are also indicative of the long standing practice of collaboration in the sciences.[157]

Research data collections[158] or portals contain the results of one or more focused research projects and data that are subject to limited processing. Data types are specialized and may or may not conform to community standards, adhere to metadata standards, and content access policies. Data collections vary in size but are intended to serve a specific scientific group, often limited to immediate participants. These collections are supported by relatively small budgets, often through research grants funding a specific project, and therefore do not have preservation as a priority. The Microvariability & Oscillations of Stars (MOST) Satellite Mission online database Case Study (CS26) falls into this category, as do a number of the portals in the GS10 study.

Resource or community data collections[159] serve a single science, geomatics or engineering community. These digital collections are often large enough to establish community-level standards, either by selecting from among pre-existing standards or by bringing the community together to develop new standards where they are absent or inadequate. The CanCore Learnware metadata standard[160] is an example of this type of community standard. The budgets for resource or community data collections are moderate and often supported by a

[155] Ibid.
[156] Bart Ballaux, *CS26 Most Satellite Mission: Preservation of Space Telescope Data* (Vancouver, 2005).
[157] American Institute of Physics (AIP), 2001, *AIP Study of Multi-Institutional Collaborations: Final Report. Highlights and Project Documentations* Melville, accessed March 2011, http://www.aip.org/history/publications.html
[158] National Science Foundation, 2005, *Report of the National Science Board*, p. 20
[159] Ibid.
[160] CanCore, *CanCore Metadata Initiative*, http://www.cancore.ca/en/. (accessed 27 January 2007)

government agency. Preservation is contingent on departmental or agency priorities and budgets. Some of the GS10 Portals that fit this description are: Canadian Institute for Health Information (CIHI); Southern California Earthquake Center (SCEC); National Geophysical Data Center (NGDC – NOAA) and the Earth Systems Grid (ESG) portal.

Reference data collections[161] are intended to serve large segments of the scientific, geomatics and education community. These digital collections are broad in scope; serve diverse user communities including scientists, students, policy makers, and educators from many disciplines, institutions, and geographical settings. Normally they have well-established and comprehensive standards which often become either de jure or de facto, such as the Geomatics ISO 19115 Metadata standard and the Federal Geographic Data Committee (FGDC) Metadata standard. Budgets supporting reference collections/portals are often large and come from multiple sources in the form of direct, long-term support; the expectation is that these collections will be maintained indefinitely but not necessarily archived. Examples from the GS10 study include: Canadian Geospatial Data Infrastructure (CGDI); Statistics Canada; and the Global Change Master Directory – Global Change Data Center.

6.3.2 Data Quality

Geospatial data creators give primacy to data quality, which includes authenticity, normally articulated as provenance or lineage as discussed in relation to the *Canada Evidence Act* (R.S., 1985, c. C-5) in Section 2.7. Data accuracy is critical and data need to be reliable. Data quality is normally articulated in a dataset's metadata; without metadata or data quality parameters, a geomatics practitioner may not use, trust or rely on that data.

Each geographic sub-discipline and science differs in how it defines data quality, as is demonstrated from the Case Studies and the Portal Survey results. However, most include some or most of the following elements: positional accuracy; attribute and thematic accuracy; completeness; semantic accuracy; and temporal information, reliability, lineage, logical consistency, and objectivity.[162] These quality elements are normally captured in metadata, and geomatics researchers argue that digital data archivists must consider data quality if they are to acquire their data. Indeed, the data quality of a record may be an important factor in the decision of what scientific data to archive. Clearly it will be very difficult for archivists to make appraisal decisions about scientific data on their own and they may not necessarily be able to do so according to typical archival practices, as Ken Thibodeau points out:

> The relevant framework of appraising scientific data sets, thus, is not defined by the business activities or the need for corporate memory of the sponsoring agency, but by the research community. Seeking the input of scientists in the appraisal of the data recognizes that the roles and

[161] GeoConnections *supra*, note 128, p. 21.
[162] Anders Ostman, 1995, *The Specifications and Evaluation of Spatial Data Quality* (Paper read at Proceedings from the 18th ICA/ACI International Conference at Stockhom, 1997) and Stephen C. Guptill and Joel L Morrison, eds. *Elements of Data Quality* (Oxford, 1995).

> the actions of academic researchers are at least as important as the functions of the agency that funded the research or launched the satellite.[163]

Data quality will be one of the important elements that will need to be factored into the archival appraisal process and will require the assistance of data creators and scientists themselves. In the best of all worlds, data quality would have been included in metadata at the beginning of a dataset's life-cycle. Archivists have a role to play with funding agencies, scientific institutions and scientists in ensuring that archival practices are part of the research process from the very beginning as discussed in a number of studies on this topic.[164]

There is a widespread need for digital preservation strategies and procedures that are purposely designed to assure people that the data and records they rely upon are what they purport to be, and are free from tampering or corruption. Statistics Canada, for example, states that "the confidence of clients in the quality of that information is critical to the Agency's reputation as an independent, objective source of trustworthy information."[165]

The US OMB *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies* document is insightful. Quality is considered to be "an encompassing term comprising utility, objectivity, and integrity."[166] As the guidelines state:

- *Utility* refers to the usefulness of the information to the intended users.
- *Objectivity* focuses on whether the disseminated information is being presented in an accurate, clear, complete, and unbiased manner, and as a matter of substance, is accurate, reliable, and unbiased.
- *Integrity* refers to security—the protection of information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.
- *Utility* is closely related to the concept of fit-for-use; *objectivity* includes all of the elements of scientific data quality, while *integrity* in this case is analogous to authenticity in archival science.

Furthermore, these guidelines state that

> agencies shall develop a process for reviewing the quality (including the objectivity, utility, and integrity) of information before it is disseminated. Agencies shall treat information quality as integral to every step of an agency's development of information, including creation, collection, maintenance, and dissemination. This process shall enable the agency to substantiate the quality of the information it has disseminated through documentation or other means appropriate to the

[163] Kenneth Thibodeau, 2005, "Preserving Scientific Data on Our Physical Universe," *IASSIST Quaterly* Winter, accessed March 2011. http://www.iassistdata.org/downloads/iqvol194thibodeau.pdf, p. 26.

[164] AIP, 1992, *Study of Multi-Institutional Collaborations* and the National Research Council, *Preserving Scientific Data on Our Physical Universe*, accessed March 2011. http://www.aip.org/history/publications.html

[165] Statistics Canada, 2002, *Statistics Canada's Quality Assurance Framework*, Ottawa, Accessed March 2011 http://www.statcan.ca/bsolc/english/bsolc?catno=12-586-X&CHROPG=1 , p.1.

[166] Office of Management and Budget (OMB), *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*: Federal Register, vol. 67 no. 36, (Washington DC, Friday, February 22, 2002) www.whitehouse.gov/omb/fedreg/reproducible2.pdf, p. 8453.

> information [and] agencies shall adopt specific standards of quality that are appropriate for the various categories of information they disseminate.[167]

Certainly, data preservationists who are mandated to preserve these data for the long term will also have to include data quality among the values they assess in the course of their appraisals. The National Geophysical Data Center portal makes explicit reference to the OMB data quality guidelines in its data management strategy when referring to influential scientific data. There is apparently no equivalent policy or guideline in Canada.

Authenticity

The responses to the InterPARES 2 Case Study questions regarding authenticity[168] revealed that the prevalent concept of reliability that is discussed later is closely tied to reproducibility and accuracy. In many cases, reliability is associated with faith in the technological systems in place or security measures related to access to the system. Trust in data sources was found to be important in both the Cybercartographic Atlas of Antarctica (CS06) and the Archeology records study (CS14),[169] while technological integrity, validity procedures and system checks were implemented in the Mars Global Surveyor Data Study (CS08), the Computerization of Alsace-Moselle's Land Registry (CS18) and the Most Satellite Mission project (CS26). Some of the studies controlled access to their datasets via the appointment of responsible agents (VanMap and the Alsace-Moselle Land Registry), or specialists (Cybercartographic Atlas), while peer review of data was a method used in the Mars Global Surveyor Data study. In the Engineering Objects Study (CS19), the creators did not believe their records to be authentic as they have no assurance system that they are. Based on their responses, it would seem that the case study respondents do not think of authenticity in the same way that archivists do, as there seems to be more emphasis on measures to ensure data quality. Most creators are trying to ensure that they have good quality data and not necessarily authentic data in an archival sense. The Land Registry may be the exception, as the system is specifically designed to ensure that each registration is authenticated in the system.

The GS10 portals that were surveyed provided a rich array of information on the topic of authenticity in the geospatial sciences. Ensuring that data are of good quality and can be trusted is critical to these data portals, or else users would not rely on them. For many of the portals, the process of data control begins at the time the data are ingested into the system, made accessible via Web server sharing protocols, or described in a metadata description form. Some portals only ingest data that are derived from peer review journals (National Virtual Observatory). Others only allow certain groups, organizations or individual researchers to contribute, such as modelers (Earth Systems Grid), designated scientists (OBIS-SEAMAP Ocean Biogeographic Information System - Spatial Ecological Analysis of Megavertebrate

[167] Office of Management and Budget (OMB), 2002, p. 8459.

[168] See Table 1: Authenticity Statements in the supplementary file on the InterPARES 2 website at http://www.interpares.org/display_file.cfm?doc=Archivaria64_Todays_Data_supplementary_tables.pdf

[169] Erin O'Meara, Richard Pearce-Moses and Randy Preston, *CS14 Archaeological Records in a Geographical Information System: Research in the American Southwest,* (Vancouver, 2004).

Populations) and approved government programs (National Snow and Ice Data Center, NASA). Some acquire data from purchased sources (Canadian Institute for Health Information). Finally, others restrict data only on the basis of whether they fit the mandate of the community that the portal serves.

Once the data are in a particular portal, there is a wide variety of security measures in place to ensure they are not tampered with. Many portals include user authentication and/or registration mechanisms (USGS Data Portals - GEO-DATA Explorer, Community Data Portal at NCAR); others only have trained personnel working the system, some have Public Key Infrastructure (PKI) and Department of Environment (DOE) GRID Certification Authority access (Earth Systems Grid), and authorized users (Long Term Ecological Research). Data users can access metadata that include lineage information to help them determine if the data are fit for use. Before data are made available to the public or to specific research communities, most have validation processes in place to attest to their authenticity and quality (NASA Life Sciences Archive; Southern California Earthquake Center). However, there did not seem to be any mechanisms in place that discussed how users can assess that the datasets that they have downloaded or received are authentic beyond metadata and file headers.

As previously discussed in the context of data being used for evidence in a judicial process, the Presumption of Authenticity is important. The creators of Science Case Studies, the VanMap Case Study, as well as most of the data portals surveyed, do not use the term authenticity. Where metadata exist, along with secure access to the data, many of the InterPARES 1 authenticity benchmark requirements are met. However, where there are no metadata, or in the case where a user wishes to assess if a data set has been tampered with during transfer, the concept of presumption of authenticity is introduced and defined as "an inference as to the fact of a record's authenticity that is drawn from known facts about the manner in which that record has been created and maintained"[170]. This being the case, the context, practices, associated documentation, validation processes and authentication, and access measures would suggest that the Case Studies discussed in this paper and the data in the GS10 portals are presumed authentic from an archival perspective.

Accuracy

The measure of accuracy, error or distance from the truth is a critical data quality element. There are numerous sources of error in any given data set, and data users are well aware of these; therefore, they count on metadata descriptions that include data lineage and accuracy measures in order to enable them to make the decision to trust and use those data. Alternatively, scientists will refer to peer reviewed papers that discuss the scientific model applied and the research methodology directing data collection. There is no one size fits all accuracy measure. Each scientific community and each specific data set includes its own accuracy parameters and particularities.

[170] InterPARES 2, *Terminology and Glossary*, accessed March 2011, http://interpares.org/ip2/ip2_terminology_db.cfm.

For geomatics practitioners, "accuracy can be defined as the difference between a measurement, or attribute, and some comparable measurement known to be of higher accuracy."[171] This argument, however, is circular and relativist, therefore data can "never be more accurate than the most accurate source."[172] Accuracy is the most pervasive and common metadata element; "to a purist, no number has meaning unless it is accompanied by an estimate of uncertainty … [and] at a minimum, the metadata should include general comments on the maximum expected errors, even if a quantitative measure such as standard deviation cannot be given."[173] In the sciences, errors are a given, therefore, a measure of the margin of that error is imperative.

In the archival community, accuracy is less specific than it is in the sciences, and is defined as "the degree of precision to which something is correct, truthful, and free of error or distortion, whether by omission or commission,"[174] or "the degree to which data, information, documents or records are precise, correct, truthful, free of error or distortion, or pertinent to the matter [Archives]."[175]

As an example, in the field of geomatics, "positional accuracy represents the nearness of those values to the entity's 'true' position in that system,"[176] position being defined by a coordinate system or a Projection Grid Coordinate System. On a map, land areas and points will adhere to different systems, such as a geodetic coordinate systems to store positional entities (e.g., wiggle of a river, a telephone pole). A number of transformations can occur from the point of collection, from one grid system to another and from one Geographic Information System (GIS) to another, and for a variety of different visualizations. This, of course, introduces errors. Users want data that are very near the positional truth, and in some cases will accept a certain distance from the truth depending on the use of the data.

A dataset's life cycle from acquisition to compilation and derivation, comprises important areas of concern to accuracy[177]. *Acquisition* is most important, since it is the point where the original observations are collected and where "fundamental assumptions, calibrations and corrections are made."[178] *Compilation* is the part where a database is created; it occurs when the facts are assembled into some sort of comprehensive arrangement or into a scientific dataset, and it is a phase where many errors can be introduced. *Derivation* is the stage where data are being manipulated; the output of this process is a representation, interpolations, averaging, and any number of manipulative techniques that may change the form, format or structure of the data.

---

[171] Michael F. Goodchild, 1995, "Attribute Accuracy," in *Elements of Data Quality*, ed. Stephen C. Guptill and Joel L. Morrison, Oxford, p. 66.
[172] Michael F. Goodchild, 1995, "Attribute Accuracy,", p. 66.
[173] National Research Council, *Preserving Scientific Data*, p. 37.
[174] Pearce-Moses, http://www.archivists.org/glossary.
[175] InterPARES 2, *Terminology and Glossary*.
[176] Jane Drummond, 1995, "Positional Accuracy," *Elements of Data Quality*, eds. Stephen C. Guptill and Joel L. Morrison, p.32
[177] See Derek G. Clarke and David M. Clark, 1995, "Lineage" in *Elements of Data Quality*, ed. Stephen C. Guptill and Joel L. Morrison, Oxford, pp. 13-30.
[178] Tosta *supra*, note 148, p. 18.

This may or may not be a reversible phase and is a diversion point from the original observations. For this reason, keeping the raw data as well as derived data is important.

Reliability

In the sciences, the concept of reliability is closely associated with the concepts of reproducibility and accuracy. It can be related to the degree to which a forecast's or model's probabilities or results match the observed frequencies of an occurrence in the environment or consistently produce the same result. More generally, reliability is a quality that can be attributed to a person, as in a reliable person; to a device, such as a reliable machine; or to a system that is organized to accomplish certain ends, as in a reliable computer or records system. It is the individual assessor who determines what attributes are required before reliability can be reasonably inferred.

The concept of reliability is similar for archivists. According to the Society of American Archivists' definition, reliability is "the quality of being dependable and worthy of trust. – 2. The quality of being consistent and undeviating. – 3. Diplomatics: Created by a competent authority, according to established processes, and being complete in all formal elements."[179] Reliability is considered to be one of the foundations of trustworthiness.[180] Trustworthiness thus has qualitative dimensions (reliability and authenticity) and quantitative dimensions (accuracy and completeness). If the record's integrity appears to be compromised in some way, or if its lineage is not clear and complete, knowledgeable users would have grounds for withholding trust.

In cartography, for example, a reliability diagram includes the authority that produced the map and the quality of the source material. Some rely on reliability diagrams that speak to the probability that a particular model, data or experiment is accurate. Reliability measures are also closely associated with measures of error in a system or data set, which of course is in turn associated with a degree of accuracy as previously discussed. Measures of reliability are statistically complex and designed to test the probability of forecasts or a models outcome.

Data Quality Disclaimers

Ironically while most organizations aim to ensure their data are accurate, reliable and authentic, the Case Studies and the Data Portals that we examined demonstrated that many of these same organizations will add disclaimers to absolve themselves of any responsibility for damages that may result from the use of their data. In the VanMap study (CS24), a special disclaimer is used in connection with utility data stating that "the City of Vancouver assumes no responsibility for the accuracy or completeness of the field information shown in VanMap. All work carried out is done wholly at the risk of the party undertaking the work who agrees, as a condition of such undertaking, to release the City of Vancouver from all liability. Location of underground

[179] Pearce-Moses, http://www.archivists.org/glossary
[180] Heather MacNeil, 2000, *Trusting records: Legal, Historical, and Diplomatic Perspectives*, Dordrecht, p. xi

utilities should always be confirmed by manual digging."[181] The Cybercartographic Atlas explicitly states that it is to be used solely for educational purposes. The Antarctic Digital Database (ADD) warns that its maps, when combined, may reflect some inconsistencies, particularly when older datasets are included. The National Geophysical Data Center and the World Data Center for Solar Terrestrial Physics indicate that the Government of the United States and its employees cannot be held accountable for any data quality warranties, and also ask that if errors are identified they would appreciate to be notified. The IU (Indiana University) Bio Archive reminds users that data contain errors, and the British Atmospheric Data Centre (BADC) absolves itself from responsibility of data on its site and once downloaded onto the user's computer.

Data Quality in General Study 10 Portals

The preservation of accurate and authentic data over the long term is best seen as a complex problem that needs to be addressed on several levels. There are three groups of issues surrounding portals and data quality:

- those related to the portal itself, its operation and its design, management, and long-term viability;
- those related to the accuracy of the individual datum and data sets; and
- those related to the relationship between the portal, its data and services, and the individual or corporate user – essentially those issues that emerge from a history of interaction that builds trust and comfort with the user.

The issues that are related to the portal itself are those that are linked to maintaining an authentic memory, especially of the sources of the data, their management or changes over time, and their connections to contributors or sources. Building sites and services that continue to be what they purport to be, and whose changes and transitions over time are visible and knowable to a user build conditions of trust. The InterPARES 1 project developed benchmarks that could be used by portals to ensure that their data continue to be authentic over time.

The portals investigated for General Study 10 represent the heterogeneity of geospatial data. Data quality, as we have seen, has many dimensions; authenticity, as it is customarily viewed within archives, is only one of these dimensions

Amongst the portals examined, there are several different examples of how data quality issues are addressed. The British Atmospheric Data Centre (BADC) follows the terms and conditions of its host organization, the Natural Environment Research Council Terms and produced the document *Terms and Conditions for data and information provided by the NERC/BADC.*[182]

[181] Evelyn McLellan, *CS24 City of Vancouver Geographic Information System (VanMap)*, (Vancouver, 2005).
[182] Natural Environment Research Council (NERC), 2007, *Terms and Conditions for data and information provided by the NERC/BADC*, Oxfordshire, accessed March 2011. http://badc.nerc.ac.uk/conditions/badc_anon.html

Combined with disclaimers on use discussed below, the document includes discussions of data limitations:

- Issues related to scientific models that evolve overtime which affect how data are collected;
- Data errors introduced during transcription and transformation;
- Details related to scale differences;
- Third party data which may not have been reviewed and
- Some datasets collected to serve particular purposes and may be incomplete for other uses.

The Ocean Biogeographic Information System – Spatial Ecological Analysis of Megavertebrate Populations portal includes a list of major data gaps that mostly address completeness of its maps, such as:

- the deep sea is the least surveyed part of the planet;
- coastal areas have not been adequately sampled for the distribution of wildlife;
- northern oceans are more sampled then those in the south;
- many marine species are not named, there have been naming changes over time, some species data have not been published, and many have not been entered into databases; and
- many databases are not connected to OBIS –SEAMAP.

The Canadian Institute for Health Information includes in its data quality framework 5 parameters: accuracy; comparability; timeliness; usability and relevance.

While the Long Term Ecological Research portal suggests that responsibility lies with data providers, it contains a number of Quality Assurance (QA) Controls in place, such as General Guidelines for QA, Parameter specific guidelines, and Parameter specific default threshold values and checks, which are also included in the metadata.

Some portals will provide assurances for their own data but not for data from third party organizations (World data Centre, OBIS-SEAMAP), or indicate that the data quality rests with those who provide or submit their data to the portal (GCDI, Long Term Ecological Research, World Data Centre – Global Change Master Directory).

Data Quality in InterPARES 2 Selected Case Studies

We now turn to data quality, particularly accuracy, within the InterPARES 2 Case Studies[183], where results show a varied approach to addressing the issue:

- The Cybercartographic Atlas (CS06) relies on the professional practices and authority of the institutions from which data are derived, and adheres to cartographic professional practices to choose the right level of data accuracy and to select cartographers for the right representation, a process which is very much reliant on metadata and professional practices.
- The Mars Global Surveyor Data Records in the Planetary Data System (CS08) includes data processing plans, manuals, specifications and workbooks to guide processing, transferring and data preparation. Further, the data are peer reviewed for accuracy and reliability and are validated through a system that also conducts checksums.
- Accuracy in the Coalescent Communities GIS (CS14) is more subjective and specific to the person who decides which data sets will be used; data sourcing is less formalized and there is a range of error acceptable to the profession.
- For the Alsace-Moselle Land Registry (CS18), a rigorous system including data verification, validation processes, PKI signatures and cross referencing, along with a well designed architecture, ensures that the registries are accurate and authenticated within a legal evidential framework of property ownership.
- In Authenticating Engineering Objects (CS19), accuracy is in the hands of the designer and his/her adherence to design modeling standards and CAD system geometry checks.
- For VanMap (CS24) there is a distributed system for establishing data accuracy based on from whom and where the data come from. There is much reliance on professional practice, the authority of organizations from which external data sources are derived and, before work is conducted that relies on the data, the data are ground truthed.[184]
- The Most Satellite Mission (CS26) has put in place a technical validation process with a series of check sums, and scientists verify the data. Some data are processed through a model; any inaccuracies are addressed at that stage, and the data are potentially reprocessed.

The analysis of these few case studies suggests that accuracy is associated with the risk of having inaccurate data: the more legal requirements there are, the more rigorous are the quality checks, as in the case of the Alsace-Moselle Land Registry (CS18). Also, the more automated

[183] See Table 2: Accuracy Statements in the supplementary file on the InterPARES 2 website at http://www.interpares.org/display_file.cfm?doc=Archivaria64_Todays_Data_supplementary_tables.pdf

[184] Ground truthing is a process by which a feature on a map or a satellite image is compared to what is there in reality (at the present time). It can be used to verify the accuracy of a classified image, or to calibrate the pixels of a satellite images to real features and materials on the ground. This is done to minimize errors in the classification, such as errors of commission and errors of omission, or to validate a feature labeled in an image.

the process is, the more technical the check sums are and the more reliant the creators are on the technical systems in place and the less reliant on human checks; this is the case with the NASA Mars Surveyor Data, the Engineering Drawing study and the MOST satellite data. Professional practice, however, is very important in the Cybercartographic Atlas of Antarctica and the VanMap studies, as is a reliance on the trust associated with the integrity and authority of external data providers. The Atlas however relies heavily on good metadata provided by external data sources and on the metadata related to the atlas modules themselves. Finally, the Archeology study included the widest margin of error in its practices and the most subjective quality checks.

6.3.3 Metadata

The US National Science Board sums up the importance of metadata as follows:

> To make data useable it is necessary to preserve adequate documentation relating to the content, structure, context, and source (e.g., experimental parameters and environmental conditions) of the data collection – collectively called metadata. Ideally, the metadata are a record of everything that might be of interest to another researcher. For computational data, for instance, preservation of data models and specific software is as important as the preservation of data they generate. Similarly, for observational and laboratory data, hardware and instrument specifications and other contextual information are critical. Metadata is crucial to assuring that the data element is useful in the future. The use of metadata and their accuracy have increased over the past several decades.[185]

Metadata are essential for the dissemination of geospatial data whereby "a data set without metadata, or with metadata that do not support effective access and assessment of data lineage and quality, has little long-term use."[186] Authenticity in the sciences is linked to a clear lineage recorded in the accumulating metadata surrounding data. Both data and their cumulative and related metadata must be present, clear, unambiguous, and un-compromised. Lineage information supports assessments of the probability of error, either in the data, or in its collection, compilation, aggregation or derivation.

Data portal discovery services rely on metadata descriptions. Metadata is like a form of truth in labeling and it is considered "axiomatic that a database has limited utility unless the auxiliary information required to understand and use it correctly – the metadata – is included in the record."[187] The data quality elements discussed throughout this text are captured in metadata. Scientists will not trust a data set that does not come with a description, and one cannot determine if a data set is fit for a particular application without metadata. The major uses of metadata include "(1) managing and maintaining an organization's investment in data, (2) providing information to data catalogs and clearinghouses, (3) providing information to aid data transfer and use, and (4) providing information on the data's history or lineage."[188] Metadata are also a means of attesting to and assessing a data set's authenticity. In the absence

[185] National Science Foundation, 2005, *Report of the National Science Board*, p.20.
[186] National Research Council, 1999, *Preserving Scientific Data*, p.36.
[187] Ostman *supra*, note 162, p. 31
[188] Ostman *supra*, note 162, p. 62.

of metadata, it is possible to gain some understanding of a scientific data set if there are associated peer review papers and reports that describe them; however, this would be a more laborious process.

Metadata schemas and standards in the sciences, particularly geographic information science, are well developed, and are essential for data discovery and fit-for-use decisions. However, many metadata standards remain housed in communities of practice and domain specific classification systems and data structures.[189] This is very apparent in the selected case studies and the portals examined in General Study 10. It was also observed that existing archiving metadata could be expanded to include metadata standards from other disciplines; the InterPARES 2 Description Team explored this issue by analyzing the ISO 19115 *An International Metadata Standard for Geographic Information*. Therefore, information models (i.e., ontologies) derived from formal and informal methods could be used to assess the feasibility of maintaining knowledge of data use context over time.

The Selected Case Studies demonstrate that each scientific and geomatics project adheres to its own standards, some of which are developed by a community of practice as in the case of the Cybercartographic Atlas (CS06), which adheres to ISO 19115, DIF and FGDC metadata standards.[190] The Mars Surveyor NASA Data Study (CS08) adheres to a NASA institutional and data type specific Planetary Science Metadata standard, while Authenticating Engineering Objects (CS19) adheres to strict corporate and vendor standards. The MOST Satellite Mission Study (CS26) developed its own very basic standard to meet the needs of its project, VanMap (CS24) does not have a clear standard, and the Archeology study (CS14) simply refers to the source of the data ingested into the GIS. The Land Registry study (CS18) indicates there are no metadata, and this may be related to the secure and encrypted access protocols and the architecture of the system. However, it is assumed that some basic catalogue type of metadata elements exist in order to find and retrieve registrations in the system. This small sample once again demonstrates the different specificities inherent in the geo-sciences.

Most but not all of the Data portals examined include metadata. Some are very minimalist and include only header files; others refer to associated peer review articles; some were designed specifically for a particular data set; while others adhere to the metadata standards of their discipline (e.g., Canadian Geospatial Data Infrastructure), Access Portal or institutions (e.g., World Data Center for Solar Terrestrial Physics).

[189] A. Gupta, B. Ludascher, M.E. Martone, 2002, "Registering Scientific Information Sources for Semantic Mediation," *Lecture Notes in Computer Science* no. 2503, pp.182-198.

[190] Federal Geographic Data Committee, *Geospatial Metadata Standard* (2007), accessed March 2011. http://www.fgdc.gov/metadata/geospatial-metadata-standards#whatstandard

6.3.4 Geospatial Data Records in Archival Terms

InterPARES 1[191] stated that five characteristics are required for a digital entity to be a record: stable content and fixed form; embedded action; archival bond; three persons (i.e. author, addressee, writer); and an identifiable administrative and documentary context. For some case studies, particularly Cybercartographic Atlas of Antarctica (CS06)[192] and VanMap (CS24)[193], which are explicitly designed to allow for data to change and information to be added, this means that they are not or do not contain records in archival terms. To become records, they must be fixed in time and space. InterPARES 2, VanMap and San Diego Centre for Supercomputing have collaboratively designed a research study to determine whether it might be feasible to introduce fixity into the system by changing the system's architecture so that each time a layer is updated the layer is saved and set aside. This would allow composite views of VanMap to be assembled for any given date, consisting of layers that had been saved on that date or most recently prior to that date.[194] This is however a far from perfect solution and is both expensive and beyond the capacity of most institutions creating and using these dynamic products.

One of the most serious problems in this respect is that for most geospatial data creators the term "record" means data, databases, and related information. For many archivists, these are not considered records except in very special and limited circumstances, where the concept of "bounded variability"[195] may be applied. This is not simply a matter of semantics. It is a fundamental difference in perspective between creators and preservers, compounded by the emergence in all disciplines of ephemeral interactive information which exists only in cyberspace. This is particularly the case for many of the InterPARES 2 Case Studies (like the Cybercartographic Atlas and VanMap) which are interactive, experiential and dynamic. Duranti and Thibodeau argue that

> interactions between humans and computer systems, experiences enabled or mediated by experiential systems, and processes which are carried out with at least some degree of spontaneity by dynamic systems are not the residue of action. They are not means of remembering either what was done or what is to be done. In short, they are not records.[196]

[191] InterPARES 1, accessed March 2011. http://interpares.org/ip1/ip1_index.cfm
[192] Sherry Xie, *Diplomatic Analysis CS06 Cybercartographic Atlas of Antarctica (revised)*, (Vancouver, 2006).
[193] Jennifer Douglas, *CS24 Diplomatic Analysis Template Preservation of the City of Vancouver GIS database (VanMap)* (Vancouver, 2006).
[194] Evelyn McLellan, *CS24 City of Vancouver Geographic Information System (VanMap)*, (Vancouver, 2005). See also the article "From Data to Records: Preserving the Geographic Information System of the City of Vancouver," by Glenn Dingwall, Richard Marciano, Regan Moore and Evelyn Peters McLellan in this issue of *Archivaria.*
[195] Evelyn McLellan, *CS24 City of Vancouver Geographic Information System (VanMap)* CS24.
[196] Luciana Duranti and Kenneth Thibodeau, 2006, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science* vol. 6 no.1, pp.13-68 p. 59.

This archival position is entirely defensible from the perspective of the theory of diplomatics[197] but is problematic in many scientific situations, such as for computational data where a model or a simulation is the primary result. The nature of the "record" is changing dramatically and traditional archival science will have to adapt to these changes in both theoretical and practical terms if they are to preserve this new information environment in the archives of the twenty first century.

6.3.5 The Role of Library and Archives Canada

Although the problems and challenges of archiving these dynamic data sets and digital artifacts are being identified, the institutional environment is often not conducive to the systematic action required to address the problem. For example Library and Archives Canada (LAC) is not ready to systematically archive born digital geospatial data, databases and maps let alone complex artifacts such as the Atlas of Canada.

Currently, there are a few LAC guidelines for cartographic material, but these are general in nature and primarily address paper maps. The 2006 LAC *Managing Cartographic, Architectural and Engineering Records in the Government of Canada*[198] made only passing reference to digital maps such as "the National Archives acquires geomatic systems" and "geomatic records include geomatic systems, discs, CD-ROMs and other cartographic material in electronic formats" (LAC 2006a). This has since been updated[199]; however, the focus remains primarily on paper content and digitized paper content. The 2011 document still refers the reader to the 2001 Canadian Committee on Archival Description (CCAD) *Rules for Archival Description Chapter 5*[200] for information pertaining to standards and practices for cartographic records. The *Rules* primarily address paper maps while general issues pertaining to digital databases and programs are covered in the 2003 *Chapter 9: Records in Electronic Form*[201]. While both of the CCAD documents are designed for archivists who manage these records, these documents could be useful references for government record managers.

---

[197] Defined as n. ~ The study of the creation, form, and transmission of records, and their relationship to the facts represented in them and to their creator, in order to identify, evaluate, and communicate their nature and authenticity. http://www.archivists.org/glossary/term_details.asp?DefinitionKey=21

[198] Library and Archives Canada (LAC), accessed 2006, *Managing Cartographic, Architectural and Engineering Records in the Government of Canada, Ottawa: Government of Canada*, http://www.collectionscanada.ca/information-management/002/007002-2050-e.html (***NOTE – this link is no longer live***)

[199] Library and Archives Canada (LAC), accessed 2011, *Managing Cartographic, Architectural and Engineering Records in the Government of Canada, Ottawa: Government of Canada*, http://www.collectionscanada.gc.ca/government/002/007002-2050-e.html.

[200] Canadian Committee on Archival Description, 2001, *Rules for Archival Description Chapter 5 Cartographic Materials*, accessed February 2011. http://www.cdncouncilarchives.ca/rad_ch5.pdf

[201] Canadian Committee on Archival Description, 2003, *Rules for Archival Description Chapter 9 Records in Electronic Form*, accessed February 2011. http://www.cdncouncilarchives.ca/RAD_chap9_revised_Aug2003.pdf

The LAC Local Digital Format Registry (LDFR) *File Format Guidelines for Preservation and Long-term Access Version 1.0*[202] includes *Geospatial* guidelines. The *Geospatial*[203] section recommends TC 211 ISO 19115 Geographic Information - Metadata (NAP– Metadata) (North American Profile). This is a dynamic document that will improve as knowledge grows and as digital geospatial data record creators and managers contribute to it. While these documents are lauded as a good start, they fall short of adequate guidelines for the kind of digital geospatial artifacts discussed in the InterPARES 2 Case Studies, the portals in the InterPARES 2 General Study on Data Portals, and the types of data being ingested into the nine geospatial data archiving and preservation examples discussed in next chapter.

[202] Library and Archives Canada (LAC), *File Format Guidelines for Preservation and Long-term Access Version 1.0*, accessed February 2011. http://www.collectionscanada.gc.ca/digital-initiatives/012018-2200-e.html.

[203] Library and Archives Canada (LAC), *File Format Guidelines for Preservation and Long-term Access Version 1.0* 2.9 Content Category: Geospatial, accessed February 2011. http://www.collectionscanada.gc.ca/digital-initiatives/012018-2220.09-e.html

7. Digital Geospatial Data Archiving and Preservation in Canada and Abroad

This chapter describes the findings and analysis related to the environmental scan of geospatial data archiving and preservation initiatives in Canada and other key countries with spatial data infrastructure programs.

7.1 Introduction

The objectives of the geospatial data archive environmental scan (E-Scan) were to: seek Canadian and international initiatives where digital geospatial data, databases and maps are being archived; briefly describe them; and provide analysis and recommendations for archiving and preserving the CGDI's geospatial data assets. As expected, there are very few geospatial initiatives that are explicitly referred to as archives. The initiatives that were found are summarized in Table 9 and are more fully described in Appendix B. These initiatives were assessed based on the examination of their websites, by testing their data access interfaces and, wherever possible, downloading their data. These findings are an overview; a comprehensive assessment would require interviews with curators in each case.

A web-search yielded only nine (9) Canadian and two (2) international geospatial initiatives which are explicitly called an archive, as shown in Table 8. In Canada there are a number of geospatial data portals that are archive-like, and do point to or provide access to collections of digital maps, digital datasets and digital databases (e.g., Discovery Portal, GeoGratis, GeoBase). These are however not explicitly called archives by their host institutions. Geospatial data portals have many, but not all, of the attributes of the geospatial data archives described in this E-Scan.

Table 8: E-scanned Geospatial Data Archives and Their Custodians

Archive	Custodian
Canadian	
1. National Climate Data and Information Archive	Environment Canada (EC)
2. National Water Data Archive	Environment Canada (EC), Water Survey of Canada (WSC) w/Archived Sediment Data example

Archive	Custodian
3. Integrated Science Data Management (ISDM) Wave Data Archive	Department of Fisheries and Oceans (DFO)
4. IOS/OSD Data Archive	Department of Fisheries and Oceans (DFO)
5. Lithoprobe Data Archive	Natural Resources Canada (NRCan), Earth Science Sector (ESS), Geological Survey of Canada (GSC)
6. The National Soil DataBase (NSDB)	Agriculture and Agri-Food Canada (AAFC); Canadian Soil Information Service (CANSis)
7. Earth Observation Data Services (EODS)	Natural Resources Canada (NRCan), Earth Science Sector (ESS), Canada Centre for Remote Sensing (CCRS)
8. National WaveForm Archive (NWFA)	Natural Resources Canada (NRCan), Earth Science Sector (ESS), Earthquakes Canada
9. Geomagnetism Summary Plots: Archives	Natural Resources Canada (NRCan), Earth Science Sector (ESS), Geological Services of Canada (GSC)
10. The Canadian Ice Service Archive (CISA)	Environment Canada, Canadian Ice Service
11. System of Agents for Forest Observation Research with Advanced Hierarchies (SAFORAH)	Canadian Forest Service (CFS), University of Victoria and other academic and government partners.
International	
12. The National Geospatial Digital Archive (NGDA)	University Libraries of University of California at Santa Barbara (UCSB) and Stanford University
13. UK Data Archive and British Library	United Kingdom National Archive and The British Library

7.2 Canadian Examples

The Canadian geospatial data archives discussed here are all housed in Federal Government departments which have a science stream such as: Environment Canada (EC), Department of Fisheries and Oceans (DFO), Natural Resources Canada (NRCan) Canadian Forest Service, and Agriculture and Agri-Food Canada (AAFC). These were mostly databases centered around a theme such as climate, an environmental element such as water, soil or forest, a water phenomenon such as spectral waves, geological phenomena such as earthquakes and geomagnetism, or a geological specific earth layer such as the lithosphere. The SAFORAH data archive was forest data, only in remote sensing formats. All seem to fit within the business function and reporting responsibilities of their custodial institution and some are research collaborations (e.g., SAFORAH, Lithoprobe). While all are grounded in established scientific data collection methods and presumably according to related standards, and all but *The National Soil DataBase (NSDB)* contain data curated from networks of sensors, only the *Lithoprobe Data Archive* was specific to an ongoing collaborative science project. It can be

argued that all would be part of the scientific work of their curatorial organizations, yet only the DFO archives made explicit reference to a departmental science strategy. Finally, while these archives reside in specific departments, the data they contain are accessioned in partnership with others, are collected as part of a cost-recovery arrangement with other institutions, are the holdings of partner institutions, are part of major collaborative research projects, are derived from sensors housed and managed in many institutions, or are derived from myriad private and public sector satellite networks. These archives are therefore all collaborative endeavours.

The *National Soil DataBase (NSDB)* stands out as the only legacy dataset, while all of the other archives include a combination of data from active and inactive sensors and are continuously being updated, in many cases with near-real time data. They are therefore growing collections of data, data that have not necessarily been set aside for disposition purposes into an archive and which may remain part of ongoing business practices. The *Earth Observation Data Services (EODS)* in particular stands out in this regard. This would imply that these are not archives but are collections of data that are well managed with preservation into a good recordkeeping system in mind. Data may be set aside for preservation and only copies are disseminated. Interviews with curators would be required to assess if this is in fact the case.

Many of the archives examined disseminate their holdings and provide the means to view them using specialized software. While none of the initiatives discussed the preservation of software, hardware and associated specialized file formats, they may however do so. For these to be archives and more than backup systems they would also have to form part of a preservation strategy, if the intent is to ensure future access. Finally, it is important to consider that each of these initiatives is unique and reflects the practices, methods and norms of its respective science. These are deeply ingrained practices that are embedded into the tools from which data are sensed, how quality is assessed, and in how these data are organized described, formatted, disseminated and used. A preservation strategy for geospatial data will have to be attentive to and respect this multidisciplinary and specialist context. These reflect very much the findings of the InterPARES 2 portal and case studies previously discussed.

7.3 International Examples

The InterPARES 2 Portal study included some international examples of geospatial data portals but not necessarily archives. When geospatial data are archived internationally, they do not differ in context to the initiatives in Canada with the exception of the National Geospatial Digital Archive (NGDA) and the British Library archive of large-scale mapping of the nation in digital format. The NGDA is part of a large US national project from the Library of Congress and funded by the National Digital Information Infrastructure and Preservation Program (NDIIPP). The University libraries of University of California at Santa Barbara (UCSB) and Stanford University are implementing the NGDA. Initial NDIIPP funding was for preservation only and not for access to the records. Stanford is currently building this access

component and it should be online by the end of 2011. The NGDA has created a number of excellent Collection Policies[204] documents, including Content Provider Agreements and Content Node Agreements. The Content Development Policy[205] for instance includes:

- Collection purpose and description of the users
- Selection, evaluation and prioritization
- Scope of coverage
- Metadata recommendations
- Sources for digital geospatial data
- Coordination and cooperation with other collections

This clearly delineates the appraisal policies of the archive, its scope, the type of data, areal coverage and so on. This is particularly useful as a means to guide archiving and preservation decisions. The NGDA Content Provider Agreement[206] can be used as a reference for development of agreements within and between institutions in Canada, in some cases as an addition to or complement to currently existing data sharing agreements (e.g., MOUs or Accords). Finally, the National Geospatial Digital Archive: NGDA Content Collection Node Agreement 3_19_09[207] is the type of agreement that could be developed between nodes in a distributed archive system in Canada.

The British Library[208] is a partner institution of the UK Data Archive[209], and it archives an annual snapshot of a subset of the UK and Northern Ireland Ordinance Survey's digital mapping datasets, which are deposited with all of the UK Legal Deposit Libraries (LDL). These are only accessible in the UK Library's special reading room from a dedicated terminal equipped with a map viewer. The commercial version of the OS MasterMap includes many layers, however, the UK Library only receives two: Topography and ITN. LAC also has a Legal Deposit[210] process and published Canadian topographic and other other NRCan maps are deposited in the Depository Services Program[211]. These are discoverable via the Depository Service Catalogue which either points directly to a publication or to a library where the publication is available. The UK Data Archive does not specifically collect geospatial data

---

204 NGDA, 2006-2009, *Collections Policies Pages*, accessed March 2011. http://www.ngda.org/policies.html

205 NGDS, 2006, *Content Development Policy*, accessed March 2011. http://www.ngda.org/docs/NGDA_Collection_Development_Policy.pdf

206 NGDA, *NGDA Content Provider Agreement,* accessed March 2011. http://www.ngda.org/docs/Stanford_NGDA_Contentprovider_102307final-1.pdf

207 NGDA, 2009, *National Geospatial Digital Archive : NGDA Content Collection Node Agreement 3_19_09*, accessed March 2011. http://www.ngda.org/docs/NGDA_NodeAgreement_March2009.pdf

208 British Library, *Ordinance Survey digital mapping*, Accessed March 2011, http://www.bl.uk/reshelp/findhelprestype/maps/digitalmapping/ordnancesurvey/osdigitalmaps.html

209 See http://www.data-archive.ac.uk/about/partners-collaborators

210 LAC, 2009, *Legal Deposit*, Accessed March 2011, http://www.collectionscanada.gc.ca/legal-deposit/index-e.html

211 PWGSC, 2010, *Depository Service Program*, Accessed March 2011, http://dsp-psd.pwgsc.gc.ca/index-e.html

unless they are part of a social science and humanities collection of data, but it assists its partner organizations with a range of services supporting creators and users of social and economic data for research and teaching. The UK National Archive is an example of an institution that has an explicit mandate to preserve digital data and to support other institutions to manage their specialized records. There is no such equivalent in Canada. The UK Data Archive provides a number of useful guidelines[212] and services to enable data providers to deposit their data and to help them manage their data when collected. In addition, the UK Data Archive has a well developed Preservation Policy[213] that NRCan could build upon along with the TBS, LAC, NGDA and other documents at the UK Data Archive.

7.4 Are These Initiatives Archives?

As previously alluded, based on the information available strictly through an E-scan it is difficult to assess if these initiatives are archives, according to the definition, "an agency or institution responsible for the preservation and communication of records selected for permanent preservation[214]" or a "place where records selected for permanent preservation are kept[215]". Permanence is the unknown variable in this assessment. It is also very difficult to evaluate if the data and databases that are 'archived' are kept in a trusted recordkeeping system which is "a type of system where rules govern which documents are eligible for inclusion in the recordkeeping system, who may place records in the system and retrieve records from it, what may be done to and with a record, how long records remain in the system, and how records are removed from it[216]".

For instance, it is possible that the institutions within which the *Earth Observation Data Services (EODS), The National Soil DataBase (NSDB)* or the *Integrated Science Data Management (ISDM) Wave Data Archive* are housed, respectively NRCan, AAFC or DFO, have a data preservation policy in place together with the necessary human and technological resources required to ensure the permanent preservation of these data and databases. It is also possible that these are data and databases which are simply stored in these so called 'archives' because they are considered important enough to be backed up and shared. For instance, the cost recovery potential of the *EODS* data might warrant investment in a good storage, retrieval

[212] Life-Cycle Management (http://www.data-archive.ac.uk/create-manage/life-cycle), Consent and Ethics which includes: legal, ethical, anonymization, access restrictions (http://www.data-archive.ac.uk/create-manage/consent-ethics), Guides for researchers (http://www.data-archive.ac.uk/create-manage/starting-research) and Advice and Training (http://www.data-archive.ac.uk/create-manage/advice-training)

[213] UK Data Archive, 2010, *Preservation Policy*, Accessed March 2011, http://www.data-archive.ac.uk/media/54776/ukda062-dps-preservationpolicy.pdf

[214] Def. Archive: InterPARES 2 Terminology Dictionary http://www.interpares.org/ip2/ip2_terminology_db.cfm

[215] Def. Archive: InterPARES 2 Terminology Dictionary http://www.interpares.org/ip2/ip2_terminology_db.cfm

[216] Def. Trusted Recordkeeping System: InterPARES 2 Terminology Dictionary (http://www.interpares.org/ip2/ip2_terminology_db.cfm). LAC expects to have its Trusted Digital Repository operational sometime in 2017 (http://www.collectionscanada.gc.ca/digital-initiatives/012018-4000.01-e.html).

and backup system, while the *NSDB* could simply be a collection of agricultural data considered of importance by a group of dedicated soil scientists within the institution and who have the skill and will to be their custodians. The *ISDM Wave Data Archive*, however, may be the closest to being a 'real archive' since DFO has a *Management Policy for Scientific Data*[217] that includes a mandate, model, infrastructure, and human resources dedicated to the archiving and preservation of their data resources. It is presumed that the DFO *IOS/OSD Data Archive* would fall under the same policy. This too would have to be validated in communication with custodians.

7.5 Are These Initiatives Preservation Repositories?

It was beyond the scope of this analysis to conduct a complete and thorough preservation evaluation of these initiatives based on the ISO *Open Archival Information System (OAIS)*[218] standard. Briefly, a preservation strategy involves "a coherent set of objectives and methods for protecting and maintaining (i.e., safeguarding authenticity and ensuring accessibility of) digital components and related information of acquired records over time, and for reproducing the related authentic records and/or archival aggregations[219]". Nonetheless, these initiatives do have elements that are characteristic of a preservation repository, such as

- the means to access the data and metadata;
- reference and context information;
- provenance information;
- licensing and terms of use;
- file format information; and
- in some cases are created within a data management policy (e.g. DFO).

Here again we see that aspects of these reviewed repositories resemble the geospatial data portals that store and manage the data they disseminate (e.g., GeoBase and GeoGratis). It is harder to assess file transfer mechanisms, preservation strategies such as data migration, emulation, etc., file name conventions, unique identifiers, adherence to standard vocabularies, backup schedules and the technology used. However, it was possible to make a series of useful

[217] DFO Management Policy for Scientific Data (http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm)

[218] ANNEX A of the 2002 *Consultative Committee for Space Data Systems Reference Model* based on the ISO *Open Archival Information System (OAIS).* The document provides an excellent case study of an existing archive. (http://public.ccsds.org/publications/archive/650x0b1.pdf). The Document has been updated to a 2009 PINK Book Version (http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf). which does not include the Case Study but does provide updated and useful model information.

[219] Def. Records Preservation Strategy: InterPARES 2 Terminology Database (http://www.interpares.org/ip2/ip2_terminology_db.cfm)

observations by looking at the Web sites of the initiatives discovered in this E-Scan, by using their data dissemination interfaces, and wherever possible, by downloading some of their data.

7.6 Observations on Findings

The following observations are based on the examination of the geospatial data archives discussed in Sections 7.2 and 7.3, according to a set of 10 Principles considered to be core criteria to evaluate and assess digital preservation repositories. These 10 Principles[220] were developed at the Centre for Research Libraries in Chicago, in consensus with the Digital Curation Center (U.K), Digital Preservation Europe, NESTOR (Germany), and the Center for Research Libraries (North America), to guide efforts on auditing and certifying repositories. These 10 Principles are used here simply to shape the examination of these geospatial data archives from a preservation perspective. In Table 9, the 10 Principles are used as a way to organize and summarize observations. Qualifying information related to these observations can be found in Appendix B.

Table 9: E-Scan Geospatial Data Archive Observations

10 Principles of a Preservation Repository	E-Scan Geospatial Archives
1. The repository commits to continuing maintenance of digital objects for identified community/communities.	The DFO initiatives are managed within a *Management Policy for Scientific Data*[221], which suggests a coordinated and supported commitment. The *Lithoprobe* initiative is part of a major collaborative research project, which could imply that the project is as stable as its research funding and the stakeholders involved. No specificities were provided and the curators would have to be interviewed for a more thorough assessment. Websites did not include mandates and plans for the other geospatial data archives examined. NGDA, UK Data Archive and the British Library are the only examples that satisfy this principle.
2. Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfill its commitment.	All the Canadian initiatives examined are housed in Federal Government Departments and it is hoped that they follow Government of Canada (GoC) disposition terms and conditions[222]. In some cases (e.g., with dedicated programs such as Water Survey of Canada (WSC) or Earthquakes Canada, a research program such as *Lithoprobe*, or a service such as the *Earth Observation Data*

[220] Center for Research Libraries in Chicago consensus developed 10 Principles digital preservation repositories (http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re).

[221] DFO Management Policy for Scientific Data (http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm)

[222] LAC Multi-Institutional Disposition Authority (MIDA) (http://www.collectionscanada.gc.ca/government/disposition/007007-1062-e.html) and TBS Directive on Recordkeeping (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=16552)

10 Principles of a Preservation Repository	E-Scan Geospatial Archives
	Services), this presumably provide clues of organizational fitness. However, programs do disappear with changing priorities. NGDA, UK Data Archive and the British Library are the only examples that satisfy this principle. However, the NGDA still relies on government grants.
3. Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.	It is presumed that, since the Canadian examples are housed in GoC departments, that GoC record management practices are in place and therefore they also maintain contractual and legal rights, and that the documents associated with these rights and responsibilities are part of the record set and practices. However, the formal preservation arrangements for these data and databases is uncertain. Most of these data also are accompanied by explicit licenses and terms of use, such as *License Agreement for Use of Environment Canada Data*[223], or copyright notices, such as *The National Soil DataBase (NSDB)* or *GSC END-USER AGREEMENT FOR DIGITAL DATA*[224]. Other datasets fall under multiple licenses, such as the *Earth Observation Data Services,* while still others come with a variety of disclaimers in terms of data quality and liability, such as the *Archived Sediment Data,* the *National Water Data Archive* or the *IOS/OSD Data Archive*. Most of the initiatives are conducted in partnership or in collaboration with other Federal, Provincial, or Territorial research institution or monitoring stations, which implies responsibilities, albeit the legal nature of these are not explicit. The Cost Share arrangements for the *Archived Sediment Data* of the *National Water Data Archive* is the most obvious legal arrangement. NGDA, UK Data Archive and the British Library also satisfy this principle and provide guidelines and policies related to licensing, node agreements, data sharing agreements, etc.
4. Has an effective and efficient policy framework.	The DFO *Integrated Science Data Management (ISDM).Wave Data Archive* and the *IOS/OSD data archive* are the most explicit in this regard as they are managed within a *Management Policy for Scientific Data*[225] that includes archiving. More broadly, it is presumed that these would also be part of the *DFO International Science Strategy*[226]. Consultation with the officials involved would be required to evaluate if the other initiatives have similar policy frameworks in place or if they adhere to the TBS information management policy and guidelines and recordkeeping directives.

[223] License Agreement for Use of Environment Canada Data (http://climate.weatheroffice.gc.ca/prods_servs/tables/attachment1_e.html)
[224] GSC END-USER AGREEMENT FOR DIGITAL DATA (http://earthquakescanada.nrcan.gc.ca/stndon/licence-eng.php)
[225] DFO Management Policy for Scientific Data (http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm).
[226] DFO Science Strategy (http://www.dfo-mpo.gc.ca/science/Publications/intss-ssint/index-eng.html).

10 Principles of a Preservation Repository	E-Scan Geospatial Archives
	NGDA, UK Data Archive and the British Library also have policy frameworks.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.	All of the Canadian initiatives seem to be preserving data that fit well within their mandates, and these also seem to have business value to their institutions and stakeholders and fit within reporting functions. Data are acquired from distributed networks of sensors (e.g., *National Climate Data and Information Archive* and *National WaveForm Archive (NWFA))*, gauges (e.g., *National Water Data Archive)*, satellite and buoy networks (e.g., *Wave Data Archive)*, receiving stations (e.g., *Earth Observation Data Services)*, and observatories (e.g., *Geomagnetism Summary Plots Archives)* which are grounded in scientific methods and well established technological practices. How these data are ingested, and the criteria which correspond to institutional commitments and capabilities are unknown variables which could be better understood by communicating with curators. The NGDA, UK Data Archive and the British Library satisfy this principle and the NGDA Collections Policy and the UK Data Archive Preservation Policy are notable in their specificities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.	The technical specifications of the Canadian repositories within which these data are contained, the security measures that govern them and how data are managed are not known. However, these initiatives all include good data documentation, file format information, and most of have metadata or the record set is accompanied by a wide body of reference and context material. In addition as it pertains to data quality, many of the initiatives include elements within their metadata, have made explicit statements on their websites (e.g., *IOS/OSD Data Archive*) or have discussed this in papers, reports and notes. Issues of provenance would be addressed in descriptions and metadata. It is uncertain how changes to the data and related software and hardware are managed across media, time, or with updates. Assessing authenticity would require a benchmarking[227] exercise which is beyond the scope of this E-Scan. NGDA, UK Data Archive and the British Library are the only examples that satisfy this principle.
7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts	See response to Principle 6 with respect to metadata and documentation. Without communicating with curators, it is not possible to assess actions taken, production, support and usage before and after preservation. There was no mention of software on these Canadian sites or mention of issues pertaining to proprietary formats. NGDA, UK Data Archive the British Library are

[227] See the InterPARES 2 PRESERVER GUIDELINES Preserving Digital Records: Guidelines for Organizations (http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_21.pdf)

10 Principles of a Preservation Repository	E-Scan Geospatial Archives
before preservation.	the only examples that clearly satisfy this principle.
8. Fulfills requisite dissemination requirements.	All of the Canadian initiatives have excellent access, dissemination and data viewing interfaces. Data catalog interfaces, maps, charts, and tables are the tools used to access data or to render these in file formats, and according to methodologies that would be known to the scientific communities who use these data. Most of the data are publicly accessible *(e.g., National Water Data Archive, Archived Sediment Data of the National Water Data Archive, Integrated Science Data Management Wave Data Archive, The National Soil DataBase, National WaveForm Archive,* and *Geomagnetism Summary Plots : Archives)*, some are cost-recovered (e.g., *National Climate Data and Information Archive, Lithoprobe Data Archive,* and *Earth Observation Data Services*) and some require special access permissions (e.g., *IOS/OSD Data Archive*). While the data are well disseminated, the software and hardware dependencies are not well described, nor were data viewers. These too would have to be preserved to ensure data can be viewed and used in the future, especially if disseminated and viewed in proprietary formats and tool. UK Data Archive satisfies this principle while the NGDA is building this access functionality at the moment. The British Library provides restricted access to some OS digital maps.
9. Has a strategic program for preservation planning and action.	The DFO *Integrated Science Data Management Wave Data Archive* and the *IOS/OSD Data Archive* are the most explicit in this regard as they are managed within a *Management Policy for Scientific Data* and more broadly within the *DFO International Science Strategy*. Specificities were, however, not available for these two nor any of the other Canadian initiatives. NGDA, UK Data Archive and the British Library are the only examples that clearly satisfy this principle.
10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.	It is presumed that all the Canadian examples do, since the data are accessible and disseminated, but the details of these from an archival perspective would have to be assessed in communication with curators. NGDA, UK Data Archive and the British Library are the only examples that clearly satisfy this principle.

7.7 GeoConnections Contributions

GeoConnections has been prescient in providing guidelines for geospatial data curators and public health data managers. For example, the GeoConnections guide, *Best Practices for Sharing Sensitive Environmental Geospatial Data*[228] was developed for data contributors,

[228] GeoConnections, 2010, *Best Practices for Sharing Sensitive Environmental Geospatial Data*, Prepared by:

owners, custodians, stewards and consumers of geospatial data who are "concerned with the issues and concepts associated with protecting, sharing and utilizing sensitive geospatial data, with a focus on supporting programs, services, businesses and / or applications related to the Environment and Sustainable Development (E&SD) community". This document provides practical guidance to those interested in developing sensitive environmental geospatial data sharing policies and protocols. The *Guide* provides a series of metrics along with a framework to aid in the assessment of the determination of data considered to be environmentally sensitive. The *Guide* also provides the following environmentally sensitive data sharing principles:

> Unless the dataset is classified as sensitive it can be provided free of restrictions;
>
> Information cannot be considered sensitive if it is readily available through other sources or if it is not unique;
>
> The Data Custodian of the information is the only agency that can determine whether an environmental geospatial dataset is to be classified as sensitive;
>
> Data Consumers of sensitive environmental geospatial datasets must honour the restrictions accompanying the information in the form of an agreement, license and/or metadata; and
>
> Organizations should document and openly publish their process, criteria and decisions.

GeoConnections has also produced some general guidelines for public health managers who may wish to disseminate sensitive demographic and health data. In *A Manager's Guide to Public Health Geomatics*[229], section 4.4 Working with Sensitive Data provides an excellent overview of public health data access issues and points on approaches and instruments that managers can use to share data, explains how to anonymize data, introduces metadata, and provides input on data access training.

AMEC Earth & Environmental a division of AMEC Americas Limited, Accessed March 2011, (http://www.geoconnections.org/publications/Key_documents/Sensitive_Env_Geo_Data_Guide_EN_v1.pdf)

[229] GeoConnections, 2010, *A Manager's Guide to Public Health Geomatics*, AMEC Earth & Environmental a division of AMEC Americas Limited, Accessed March 2011, (http://www.geoconnections.org/publications/Key_documents/ManagerGuide_PubHealthGeomatics_EN.pdf).

8. Archiving and Preservation Business Models

This chapter describes the findings and analysis with respect to business models[230] for digital data archiving and preservation, starting with a brief discussion of theoretical models of records archiving and preservation. The business models that were discovered and are described hereunder come primarily from the research and science communities.

8.1 Theoretical Archiving and Preservation Models

There are two theoretical models or metaphors that archivists use to think about the records archiving and preservation processes. The *Lifecycle Model* can be described as a model of records management and archival science that characterizes the life span of a record as comprising eight sequential stages: creation or receipt; classification; maintenance and use; disposition through destruction or transfer to an archival institution or agency; description in archival finding aids; preservation; reference and use.[231] The *Continuum Model* can be described as a model of archival science that emphasizes overlapping characteristics of recordkeeping, evidence, transaction, and the identity of the creator.[232] In contrast with the lifecycle model, which is based on clearly definable stages in recordkeeping and creates a sharp distinction between current and historical recordkeeping, the continuum model emphasizes the integration of recordkeeping and archiving processes.

The lifecycle model is based on the premise that records managers and archivists play separate and distinct roles in recordkeeping, with the role of the first group ending upon records disposition. One of the first to note that this model has limitations and that it should be replaced by a new paradigm, which he called the "continuum model", was Jay Atherton. From 1978 to 1986, Mr. Atherton served as senior manager responsible for the work of the National Archives of Canada in promoting and facilitating effective collaboration between the professions of records management and archives, and he has been honoured for his contributions to the

[230] "Business models" in this context refers to information management lifecycle processes that integrate archiving and preservation into ongoing operational environments

[231] An alternative definition used in the InterPARES research project is "the theory that records go through four distinct stages of change in activity, including creation or receipt, use and maintenance, in-active storage, and disposition (destruction or transfer to an archives)". (see InterPARES 2 Dictionary, accessed March 2011, http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf&CFID=1645881&CFTOKEN=65820933)

[232] Ibid.

archivist profession.[233] He argued that "effective management of recorded information … requires ongoing cooperative interaction between the records manager and the archivist in order to:

- ensure the creation of the right records, containing the right information, in the right format;
- organize the records and analyze their content and significance to facilitate their availability;
- make them available promptly to those (administrators and researchers alike) who have a right and a requirement to see them;
- systematically dispose of records that are no longer required; and
- protect and preserve the information for as long as it may be needed (if necessary, forever).[234]

While the model proposed by Atherton has gained some support, it has also been criticized because it retains the concept that there is a linear sequence of stages that occurs in the history of a single record. Other continuum models have been proposed by the influential Australian archivist Frank Upward[235] and by Brien Brothman[236], who has proposed a third metaphor that he called the helical model. Recent work at LAC on the impact of Web 2.0 have produced several interesting Thought Papers, including one by Greg Bak that touches on records archiving and preservation models, *Impacts of Web 2.0 on Information Models: Life Cycle and Continuums*.[237] Bak argues that the life cycle model (or metaphor) breaks down in Web 2.0 environments, where the following questions are germane:

- What is the moment of creation of a mashup of data sets?
- What is classification, and when is it effected, in an environment better suited to folksonomies than controlled vocabularies, and in which terms may be changed, added or deleted at any point?
- What is the right trigger for retention in a digital environment — creation, last edit, last use?
- How does an organization effect disposition on materials, such as Facebook profiles, over which they have ceded ownership?

[233] See Association of Canadian Archivists Award Recipient Biographies, accessed March 2011. http://archivists.ca/content/aca-award-recipient-biographies

[234] Atherton, Jay 1985. "From Life Cycle to Continuum: Some Thoughts on the Records Management-Archives Relationship", *Archivaria 21* (Winter 1985-86), pp. pp. 43-51.

[235] Upward, Frank (1996). "Structuring the records continuum part one: post-custodial principles and properties." *Archives and Manuscripts* 24/2 (Nov 1996), pp. 268-285; and Upward, Frank (1997). "Structuring the records continuum part two: structuration theory and recordkeeping." *Archives and Manuscripts* 25/1 (May 1997), pp. 10-35.

[236] Brothman, Brien 2006. "Archives, Life Cycles, and Death Wishes: A Helical Model of Record Formation", Archivaria 61, pp. 235-269.

[237] Bak, Greg 2010. *Impacts of Web 2.0 on Information Models: Life Cycle and Continuums*, accessed March 2011. (http://www.collectionscanada.gc.ca/digital-initiatives/012018-3403-e.html)

- How is the owner or steward of collaboratively authored resources identified?[238]

These are very challenging questions, with which producers of digital geospatial data are also now grappling, and the last two questions are particularly meaningful within a VGI/crowdsourcing context.

In a subsequent presentation at a meeting of the ARMA-NCR Chapter[239], Bak built upon his Thought Paper to propose a new metaphor, called the "Canadian Continuum"[240], which is based upon Atherton's continuum but eliminates its linearity and stages[241], and addresses some of the criticisms that have been made of Upward's continuum model. Bak identified the following requirements for a new, Web 2.0 and eRecords friendly metaphor:[242]

- no prescription of a linear series of stages;
- no restriction of roles among creators, records managers and archivists (e.g., classification may start with a user adding freetext tags, which a records management system might crosswalk against a controlled vocabulary, with an archive later adding a new layer of classification to ensure that the classification reflects current language use);
- broad understandability by recordkeeping specialists as well as non-specialists, and by all members of an organization that interact with information resources; and
- a goal of maximal utility of the information resources that it describes.

While this appears to be a forward-looking and reasonable new model or metaphor for digital information archiving and preservation, it has not yet been formally adopted or endorsed by LAC.

8.2 Useful Business Models

From those assessed, the following examples of business models (within the context of this study) are considered worthy of highlighting and of consideration for future emulation in the geospatial information archiving and preservation domain.

[238] Ibid.
[239] ARMA used to be the acronym for the Association of Records Managers and Administrators.
[240] Bak, Greg 2011. "Towards a Canadian Continuum: Recordkeeping, the Records Continuum and Web 2.0", paper presented at ARMA-NCR Chapter meeting, January 2011.
[241] It is a single-phase continuum, which he refers to as "information management and use"
[242] Ibid.

8.2.1 DFO Integrated Science Data Management (ISDM) Wave Data Archive[243]

The ISDM Wave Data Archive is a good example of an initiative that integrates preservation and archiving of data from multiple organizations into an operational information management environment. The active field program of wave acquisition started in 1971, and since the program was discontinued in 1996, ISDM has continued to process delayed-mode wave data submitted by researchers, universities, regional institutes, and the oil and gas industry. In addition, ISDM acquires daily wave data from buoys operated by the Meteorological Service of Canada (MSC) at Environment Canada.

As discussed in Chapter 7, the ISDM is a digital data archive that manages scientific data from a wide range of sources. The ISDM databases contain over 10 million hourly sea state and swell measurements from some 500 locations in Canada's lakes and surrounding oceans. While this information is being used in real-time to produce weather and sea state forecasts, its archiving program ensures that the information can be used in a variety of applications requiring data over long timeframes, such as hindcast models of wave climatology used in ocean maritime navigation, engineering and climate change studies.[244]

While wave data is mentioned here as an example of DFO's digital data archiving work, the archiving and preservation commitment extends across all kinds of scientific data for which DFO is responsible, as evidenced by the department's *Management Policy for Scientific Data*[245], which came into effect in June 2001. This *IM Policy* includes several references to the requirement for data archiving and preservation, such as the following:

> **Basic Principles**
>
> 1. Fisheries and Oceans Canada (DFO) scientific data sets… are irreplaceable, and must be protected and managed to ensure long-term availability
> 2. … it is essential that DFO Science/Oceans maintain responsibility for their quality control, management, archiving and dissemination
> 3. … all scientific data collected by the Department must be migrated to a 'managed' archive immediately after the data have been processed
>
> **Data Archiving**
>
> All DFO scientific data must be managed as part of an integrated system accessible through regional, zonal and national data centres… The responsibilities of the integrated system of data centres will be to:
>
> - Ensure long term accessibility and documentation in the event of organizational changes, retirements, etc.

[243] See http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/index-eng.htm
[244] Canadian Hydrographic Service, 2008. Integrated Science Data Management
[245] DFO, 2001, *Management Policy for Scientific Data*, accessed March 2011. (http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm)

- Protect data against loss resulting from error, accident, technological change, degradation of media, etc.

Data Submission

[It] is important to ensure that data are quickly migrated into a 'managed' environment where they are properly backed up and secured from accidental or circumstantial loss, and where the supporting metadata are integrated with the data to preserve the long-term usefulness of a data set

Data Rescue

DFO Science and Oceans sectors will develop a national data rescue program to locate and preserve scientific data that are of value to departmental programs and may be in danger of being lost.[246]

8.2.2 International Polar Year (IPY) Data and Information Service[247]

The International Polar Year was a large Arctic and the Antarctic scientific programme from March 2007 to March 2009, organized through the International Council for Science (ICSU) and the World Meteorological Organization (WMO).[248] The International Polar Year Data and Information Service (IPYDIS) is a global partnership of data centers, archives, and networks working to ensure proper stewardship of IPY and related data. The IPYDIS activities stem from recommendations made at an IPY data workshop in March 2006, which focused on overall data management planning.[249] The National Snow and Ice Data Center at the University of Colorado is a coordination office for the IPYDIS that ensures the long term preservation and access to IPY data.

The IPYDIS received guidance from the IPY Data Policy and Management Subcommittee[250], which developed the *IPY Data Policy*[251]. Regarding data preservation and archiving, the *Policy* states

> …it is essential to ensure long-term preservation and sustained access to IPY data. All IPY data must be archived in their simplest, useful form and be accompanied by a complete metadata description. An IPY Data and Information Service (IPYDIS—http://ipydis.org) should help projects identify appropriate long-term archives and data centers, but it is the responsibility of individual IPY projects to make arrangements with long-term archives to ensure the preservation of their data. It must be recognized that data preservation and access should not be afterthoughts and need to be considered while data collection plans are developed.[252]

---

246 Ibid.
247 IPY, *International Polar Year Data and Information Service (IPYDIS)*, accessed March 2011. (http://ipydis.org/)
248 IPY, 2007-2008, About IPY Page, accessed March 2011. (http://ipy.arcticportal.org/about-ipy)
249 Mark Parsons et al, 2009. *IPY Data Management Strategy, Status, and Roadmap*, Prepared for the 8th Meeting of the ICSU/WMO IPY Joint Committee, 23-24 February 2009, accessed March 2011. (http://ipydis.org/documents/jc8report_feb09.pdf)
250 IPY, 2007-2008 Data Management Page, accessed March 2011. (http://classic.ipy.org/international/joint-committee/data-management.htm)
251 IPY, 2008, International Polar Year 2007-2008 Data Policy, accessed March 2011. (http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf)
252 Ibid.

The IPYDIS is a good example of a distributed or "virtual" archive of scientific data. The Service's Web site guides digital scientific data users to a number of portals and centers that are currently providing access to IPY and related data. All such data registries and repositories are required to adhere to the IPY Metadata Profile[253] requirements. This profile was based on the Global Change Master Directory (GCMD) Directory Interchange Format, and is compliant with and has been mapped to the following geospatial metadata standards:

- Global Change Master Directory (GCMD) Directory Interchange Format (DIF);[254]
- Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) (FGDC-STD-001-1998)[255] and Remote Sensing Extensions (RSE) (FGDC-STD-012-2002);[256] and
- THREDDS Dataset Inventory Catalog Specification Version 1.0.[257]

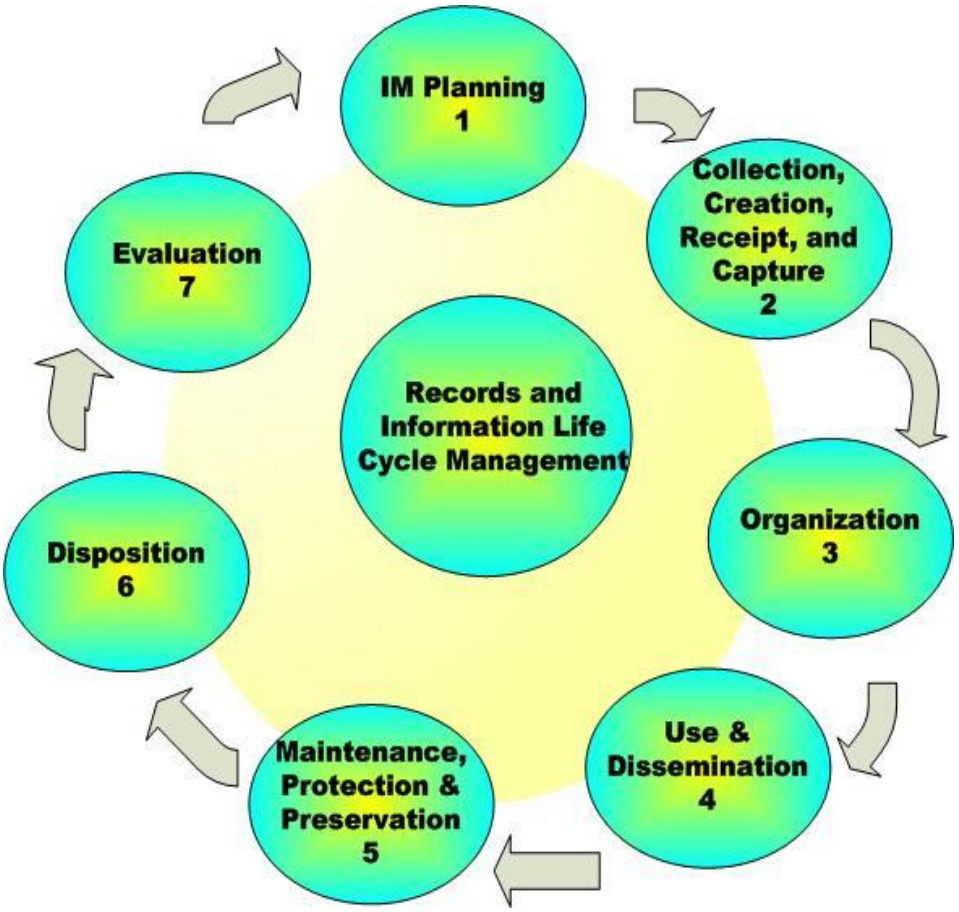
In addition, several other international standards are recommended for use by IPY researchers, including Open Geospatial Consortium (OGC) standards, and the ISO 19115:2003 metadata standard.

8.2.3 LAC Records and Information Life Cycle Management Tool[258]

This is an online guide to help government information managers improve their knowledge and understanding of the seven steps in the records and information life cycle, as illustrated in Figure 1.

[253] IPY Data and Information Management Service, *IPY Metadata Profile version 1.0*, accessed March 2011. (http://ipydis.org/data/metadata.html)
[254] Global Change Master Directory. National Aeronautics and Space Administration, 2010, *Directory Interchange Format (DIF) Writer's Guide, 2010*, accessed March 2011. (http://gcmd.nasa.gov/User/difguide/) See http://gcmd.gsfc.nasa.gov/User/difguide/difman.html http://gcmd.gsfc.nasa.gov/User/difguide/difman.html
[255] FGDC, 2006, *Content Standard for Digital Geospatial Metadata*, accessed March 2011. (http://www.fgdc.gov/metadata/csdgm/)
[256] FGDC, 2002, *Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata*, accessed March 2011. (http://www.fgdc.gov/standards/projects/FGDC-standards-projects/csdgm_rs_ex/MetadataRemoteSensingExtens.pdf)
[257] UNIDATA, 2004, *Dataset Inventory Catalog Specification Version 1.0*, accessed March 2011. (http://www.unidata.ucar.edu/projects/THREDDS/tech/catalog/InvCatalogSpec.html)
[258] See http://www.collectionscanada.gc.ca/government/products-services/007002-2012-e.html

Figure 1: Records and Information Life Cycle Management



Source: Library and Archives Canada Web Site (http://www.collectionscanada.gc.ca/government/products-services/007002-2012-e.html)

The guide describes each stage of the life cycle under the following headings:

- **Stage Name** – a brief introduction to each stage
- **Why is the Stage Important?** – explains the importance of each stage and what is involved
- **Critical Notes About the Stage** – explains the most important aspect(s) of each stage
- **Inputs to the Stage** – describes where IM practitioners can get key inputs to each stage
- **Outputs from the Stage** – identifies the key outputs from each stage
- **LAC's Role in the Stage** – describes how LAC can assist IM practitioners with each stage
- **Resources that Support the Stage** – provides links to other guidance and tools that IM practitioners can use at each stage of the life cycle

While not exclusively designed to support the archiving and preservation of digital information, this guide is a good resource that demonstrates that archiving and preservation are fundamental parts of an overall information management process and must be built into that process from the beginning. From this perspective, the most important stages of the life cycle are Stage 5: Maintenance, Protection and Preservation and Stage 6: Disposition. The guide expands on LAC's key role in Stage 5 and provides links to a wide range of resources.[259] Regarding Stage 6, the guide emphasizes the importance of dealing with records as soon as they no longer have business value to the organization, to avoid the costly buildup of stored records backlogs, and of determining whether they will be destroyed, alienated or transferred to LAC for archiving. The role of LAC at Stage 6 is clearly laid out and an extensive list of additional resources[260] is provided. A good complement to this material is Research Data Canada's five principles of data stewardship discussed in Section 5.5.

8.2.4 TBS Guideline for Employees of the Government of Canada: Information Management (IM) Basics[261]

TBS published these guidelines to support the implementation of the *Policy on Information Management*, the *Directive on Information Management Roles and Responsibilities* and the *Directive on Recordkeeping*, which were discussed in Chapter 4. They were designed to help Government of Canada employees to better understand their roles and responsibilities in managing information resources. The guide is divided into the following sections:

- **The Value of Managing Information** – introduces the purpose and value of information management (IM), the related responsibilities, and key IM practices
- **Managing Information – In Practice** – describes key basic practices and activities for managing government information organized according to the stages of the information life cycle
- **Manage Information Based on Its Value** – identifies the value of information and the related retention and disposition issues for employees
- **Managing Electronic Information and Email** – provides guidance on managing, filing, and sharing electronic information, along with tools that may be available within the institution
- **Safeguard and Protect Information** – provides information and guidance on the protection and security of various forms of information
- **Job Change Can Affect Information Management Practices** – identifies what an employee should know and do when arriving at or leaving a job

[259] See http://www.collectionscanada.gc.ca/government/products-services/007002-2067-e.html
[260] See http://www.collectionscanada.gc.ca/government/products-services/007002-2073-e.html
[261] See http://www.sct.gc.ca/pol/doc-eng.aspx?id=16557&section=text

- **Where You Can Find More Help** – identifies other sources of information including experts to consult and links to other online resources

The Managing Information – In Practice[262] section of the guide deals with the need to keep relevant information long-term to support operational needs or to preserve information of enduring value, and emphasizes that there are strict rules governing the disposal of government information. It also draws the distinction between records of enduring value (i.e., requiring archiving) and transitory records, described as "those information resources that are required only for a limited time to ensure the completion of a routine action or the preparation of a subsequent record"[263], the destruction of which must be authorized by LAC.

8.2.5 InterPARES Preserving Digital Records: Guidelines for Organizations[264]

Produced by the International Research on Permanent Authentic Records in Electronic Systems (InterPARES), these guidelines are intended to highlight a number of areas that are particularly important to the preservation of authentic digital records. The guidelines are based on the premise that

> management of digital records must proceed from a comprehensive understanding of all phases or stages of records' existence, from the time they are generated, through their maintenance by their creator, and during their appraisal, disposition, and long-term preservation as authentic memorials of the actions and matters of which they are a part. From the perspective of long-term preservation, all the activities to manage records throughout their existence are linked, as in a chain, and interdependent… If certain activities and actions are not undertaken on records, their integrity (that is, their reliability and authenticity) and preservation are imperiled.

The contents of the guidelines are divided into the following sections:

- **Manage Chain of Preservation Framework** – determining framework requirements, and designing, implementing and maintaining a chain of preservation framework, which includes all the elements of policy, strategy, methodology, etc. needed to manage digital records
- **Appraise Records for Permanent Preservation** – regularly making decisions on the disposition of records as part of the management of a recordkeeping system, which in some cases means conducting appraisals when it is determined that records in a longstanding system need to reach a disposition
- **Acquire Selected Records for Permanent Preservation** – acquiring selected records, and preserving them thereafter, with the goal of continued authenticity and accessibility

[262] TBS, 2009, Managing Information-In Practice, *Guideline for Employees of the Government of Canada: Information Management (IM) Basics*, accessed March 2011. (http://www.sct.gc.ca/pol/doc-eng.aspx?id=16557&section=text#cha5)
[263] Ibid.
[264] InterPARES 2, *Preserver Guidelines*, accessed March 2011. (http://www.interpares.org/display_file.cfm?doc=ip2(pub)preserver_guidelines_booklet.pdf)

of those records that are selected for continuing preservation; moving records from the creator's (or legitimate successor's) to the preserver's custody with great care to ensure that nothing goes awry in the transfer process

- **Preserve Accessioned Records** – keeping the authentic copies of the creator's records in a trusted preservation system, which should include in its design description and retrieval systems, and rules and procedures for the ongoing production of authentic copies as the existing system becomes obsolete and the technology is upgraded
- **Output Records** – providing access to preserved records, managed by the preserver with the same sense of responsibility and degree of technical and professional competence imparted to records appraisal, acquisition/transfer, description and storage

Each section of the document provides preservers of digital records with detailed, step by step guidance on how to ensure that those records can be properly preserved for long term future use. The guide includes three appendices: the first provides benchmark requirements for an assessment of the authenticity of the creator's digital records; the second outlines the minimum conditions necessary to enable the preserver to attest that copies of inactive electronic records are authentic; and the third offers a framework of preservation strategies to protect and maintain the accessibility of authentic copies of digital records throughout the chain of preservation.

9. Conclusions and Recommendations

This chapter provides conclusions drawn from the analysis of the research findings, and based on those conclusions, recommendations for proposed actions by GeoConnections in additional research and in the development of specific archiving and preservation operational policies of benefit to Canada's geospatial community.

9.1 Legislation Review

The review of Canada's federal legislative requirements for archiving and preservation of geospatial data covered overarching acts and regulations that relate to the topic, and those that relate to geospatial data specifically with a primary focus on legislation for which the Minister of Natural Resources is responsible. This review demonstrated that there is considerable complexity in the archiving and preservation requirements in such legislation and that even the legislation for which one minister is responsible (albeit one with a lead responsibility in this domain) has a broad range of geospatial information implications. We conclude that the archiving and preservation provisions of the *LAC Act* clearly apply to digital geospatial datasets, databases and maps, and that the provisions of several other key pieces of legislation (e.g., *Copyright Act, Privacy Act, Evidence Act*) must be taken into consideration when developing archiving and preservation policies and plans. We also conclude that any broad-based action to address digital geospatial information archiving and preservation at the federal level would benefit from the application across all producer organizations of the type of focused research and analysis conducted here for Natural Resources Canada. A good place to start would be the identification of records with high institutional business value, particularly those specifically mentioned in the Acts and Regulations.

Recommendation 1: **That GeoConnections consider conducting a thorough review and analysis of all federal legislation that relates to geospatial information to determine the type and extent of data assets that may require archiving and preservation treatment, as a basis for any comprehensive strategy for addressing geospatial information archiving and preservation within the Government of Canada.**

One area of particular complexity, which is becoming more commonplace in the geospatial information domain, involves the creation of digital objects by many record-creating organizations, whose business practices may be different from those of the Government of Canada even though the assembly and dissemination of the dataset is within the purview of the Government of Canada (e.g., framework datasets such as GeoBase). Given the federal

legislative requirement for archiving and preservation of geospatial information that has been clearly revealed by this research and analysis (and we suspect is also required at the provincial level), we conclude that future federal-provincial-territorial agreements dealing with shared creation of datasets should specifically address data archiving and preservation. In addition, in the context of archiving and preservation, associated copyright and intellectual property issues would also have to be dealt with.

Recommendation 2: **That GeoConnections consider promoting the incorporation, within any future federal-provincial-territorial agreements for the co-creation of digital geospatial datasets, of specific provisions for data archiving and preservation, including the processes for examination and resolution of any associated intellectual property and copyright issues.**

9.2 LAC Consultations

The consultations with Library and Archives Canada (LAC), supplemented with additional research, were very productive. While it is clear that the dynamics of digital information are creating considerable challenges for archivists, LAC has made some structural changes to address these challenges and is willing to work with digital data creators on their archiving and preservation efforts. We conclude that there is considerable merit in NRCan and other geospatial information producers considering LAC's tools and guidelines and engaging proactively with LAC to develop their archiving and preservation strategies and plans.

Recommendation 3: **That GeoConnections consider promoting, within the federal geospatial information producing community, the proactive engagement of Library and Archives Canada's Digital Resource Division in the development of their digital geospatial information archiving and preservation strategies and plans.**

An area of LAC work that holds potential promise is their project to develop methodology associated with the TBS Directive on Recordkeeping, in which NRCan recordkeeping officials are partnering. Since this new methodology (to be released in April 2011) may help geospatial data producers address data archiving and preservation needs, we conclude that it merits close attention and that it would be beneficial for GeoConnections to consult with their recordkeeping officials in NRCan to determine what other types of agreements and collaborations are in place and how best to employ the new methodology.

Recommendation 4: **That GeoConnections consider collaborating with recordkeeping officials in Natural Resources Canada's Enterprise IM division on the application of the new Methodology for Recordkeeping to digital geospatial information archiving and preservation.**

9.3 Canada's Information Management Policies

The review and analysis of the federal government's information management policies has clearly demonstrated that effective policy instruments, directives and standards are in place to assist geospatial information managers with management of their information resources throughout their lifecycles, including the preservation and archiving stages. What is less clear is the extent to which such policies are being implemented by geospatial information producing agencies, but anecdotal evidence suggests that there is room for improvement. We conclude that the full implementation of all of the Treasury Board Secretariat's IM policies would position the Government of Canada's critical digital geospatial datasets, databases and maps well for subsequent archiving and preservation, but that a broader assessment of current practices would be essential as a precursor to the development of any comprehensive strategy for addressing geospatial information archiving and preservation.

Recommendation 5: **That GeoConnections consider conducting a thorough assessment of federal geospatial information producers' applications of the TBS IM policies, directives and standards, as a basis for any comprehensive strategy for addressing geospatial information archiving and preservation within the Government of Canada.**

9.4 Canadian Digital Data Archiving and Preservation Consultations and Reports

National discussions in Canada have primarily focused on the creation of digital strategies and/or the preservation of scientific data in the social and physical sciences. Only the 2005 GeoConnections study on *Archiving, Management and Preservation of Geospatial Data* addresses geospatial data, while others subsume geospatial data within discussion of scientific data. Although there are differences regarding the specificities of recommendations arising from this work, there is general consensus on the following:

- a national distributed trusted digital repository or infrastructure or system is required;
- institutions should have a data preservation management policy in place;
- good record management practices should occur when and where data are created; and
- the lack of data discovery mechanisms is a major issue as is data access which includes cost, licensing, and copyright among many other issues.

All the reports are suggesting, and LAC are leaning towards, the creation of a distributed data archive. NRCan has acquired significant technical, policy, partnership and operational knowledge with the building of the Canadian Geospatial Data Infrastructure (CGDI) including direct experience with distributed systems, the stewardship of large complex and dynamic

geospatial datasets from many domains, and the discovery and dissemination of these assets. Also, some units within NRCan are currently engaged in the preservation of geospatial data. We therefore conclude that a key additional component of the CGDI should be archiving and preservation, particularly since many of the existing components (e.g., Portal, metadata, standards, GeoBase, GeoGratis, interoperability, data disseminating and sharing guidelines, etc.) are key components of a distributed archive.

Recommendation 6: **That GeoConnections consider making the archiving and preservation of geospatial data a key new thrust of the CGDI and that the processes and practices adopted to create the CGDI be adapted and focused toward the creation of a Geospatial Data Distributed Archive Hub.**

Furthermore, national discussions point to the need for the development and implementation of a data preservation management policy. Also, national discussion reports, LAC and TBS guidelines point toward the adoption of a lifecycle records management model. There is braod consensus that records need to be managed where and when they are created. This ensures that geospatial dataset, database or map creators factor in, from the very beginning, the potential to preserve their records, which eases the work of data curation and ingestion into an archive at a later date. We conclude that NRCan is in a strong position to lead consultations between the Departments' records management unit, other federal geospatial information producers, InterPARES researchers, data librarians and archivists, and LAC to develop a geospatial data archiving and preservation policy that would be part of an overarching geospatial records management policy based on a lifecycle model.

Recommendation 7: **That GeoConnections consider leading, in consultation with the key stakeholders and experts, a collaborative effort to develop a geospatial data archiving and preservation policy as part of an overarching geospatial data records management policy based on a lifecycle model.**

This research and analysis suggests that the conclusions reached and the recommendations made by the authors of the 2005 GeoConnections study on *Archiving, Management and Preservation of Geospatial Data* continue to be valid. We conclude in particular that leadership is required if the federal-provincial-territorial family of geospatial data producers are serious about tackling this challenge, and the Canadian Council of Geomatics (CCOG) is perhaps best positioned to provide such leadership.

Recommendation 8: **That GeoConnections consider requesting the Canadian Council of Geomatics (CCOG) to lead efforts to develop and implement geospatial information archiving and preservation policies and practices, and support the CCOG in those efforts.**

9.5 Canadian Digital Data Archiving and Preservation Research

InterPARES 2 research has demonstrated that practices related to data access, licenses, disclaimers and data quality vary depending on the science and the organizations that produce and share these data, which are issues GeoConnections is very familiar with. The heterogeneity of data being produced further reinforces the important role that varied participants (e.g., scientists, developers, software vendor, etc.) in the life-cycle management of geospatial data records, particularly preservation, play since only they have the necessary expertise to inform the context within which these are created. We conclude, based on the reaffirmation of the InterPARES 2 research methods, that the preservation of dynamic, interactive and experiential records, which include geospatial data, can only be carried out in collaboration with multiple stakeholders and experts. These findings very much support the rationale for Recommendation 7 above.

In addition, the InterPARES 2 Portal Study showed that some of these portals have data stored in one location, some access data in a distributed fashion, others point to a data custodian, and some provide the means to render and view their data holdings. A key finding of the Portal Study is that these provide a framework within which geospatial data archivists can work and from which they can expand with policies, standards and metadata. Also, the data made accessible via portals have already been appraised as being valuable. The characteristics of the NRCan portals linked to the CGDI are similar to those investigated by InterPARES 2. We therefore conclude that NRCan could benefit from inventorying its data portals, creating a common window into all of these, making a decision to preserve those data worthy of preservation and building data archiving, preservation and management practices within and around these existing initiatives. This could form the beginning of the access portion of a Geospatial Data Distributed Archive Hub.

Recommendation 9: **That GeoConnections consider creating a single window into all of its geospatial data assets, appraising those data collections in the existing portals for preservation worthiness, and building geospatial data archiving and preservation practices into these technologies and within the institutions that create these data.**

In consideration of requirements of the *Evidence Act* discussed in Chapter 2, we also conclude that the authenticy work of InterPARES (in particular their Benchmark Requirements for Authenticity) should be a key component of the preservation of geospatial data and this process should be included in a geospatial data archiving and preservation policy (see Recommendation 7).

9.6 Digital Geospatial Data Archiving and Preservation

While it is difficult to fully qualify them as archiving and preservation initiatives without direct consultation with the officials responsible, we were able to discover several Canadian and international examples of geospatial information systems that mention archiving as part of their mandates. Our analysis of these examples revealed that there is considerable variation in how well they comply with the set of core criteria to evaluate and assess digital preservation repositories known as the 10 Principles[265]. However, several good examples of tools were revealed, such as the National Geospatial Digital Archives' Collection Policies documents, including Content Provider Agreements and Content Node Agreements, and the UK Archives' Preservation Policy. We conclude that there are some good examples of digital geospatial data archiving and preservation initiatives in Canada and internationally that warrant more detailed investigation through direct contact with responsible officials. In addition, we conclude that guidance for further initiatives in Canada is available from other jurisdictions, which could reduce the level of effort and improve the likelihood for success of such initiatives.

Recommendation 10: **That GeoConnections consider conducting a thorough assessment of the geospatial archiving and preservation initiatives and tools identified through this project, as a basis for any comprehensive strategy for addressing geospatial information archiving and preservation within the Government of Canada.**

9.7 Archiving and Preservation Business Models

Many of Canada's current recordkeeping and archiving policies and practices are based upon the Lifecycle Model, which subdivides the life of information assets into a series of stages, and creates a sharp distinction between current and historical recordkeeping. Newer paradigms like the Continuum Model emphasize the integration of recordkeeping and archiving processes. Given the dynamic nature of geospatial information resources and the complexities of deciding when and how to preserve information snapshots for future generations, we conclude that the newer models, which emphasize the interaction between recordkeepers and archivists at the earliest stages in a dataset's or database's existence, are better suited to today's environment. However, we have found no evidence that such a paradigm is in current use in the geospatial community.

We have examined several types of business models that integrate archiving and preservation into ongoing information management operational environments. While we have discovered several excellent examples of guidelines and tools to achieve this objective, only two examples

[265] Center for Research Libraries in Chicago *supra*, note 218

of fully operational business models that are founded in a formal policy have been found – Department of Fisheries and Oceans' Integrated Science Data Management (ISDM) system and the International Polar Year Data and Information Service (IPYDIS). We conclude therefore that these two business models are worthy of further investigation and consideration as models for Government of Canada geospatial information producers to emulate.

Recommendation 11: **That GeoConnections consider developing a business model for the use of geospatial information producers in integrating archiving and preservation into ongoing information management operational environments, based on detailed investigation of the ISDM and IPYDIS business models.**

Appendix A: LAC Consultation Participants

Susan Haigh Acting Director, Library and Archives Canada Digital Resource Division 550 de la Cité Blvd Gatineau, Québec K1A 0N4 Canada Telephone : 819-997-6207 susan.haigh@lac-bac.gc.ca	***Karine D Burger*** Acting Manager Library and Archives Canada Digital and Accessibility Office Digital Resource Division 550 de la Cité Blvd Gatineau, Québec K1A 0N4 Canada Telephone : 819-953-4953 Fax : 819-934-7534 karine.burger@lac-bac.gc.ca	***David Brown*** Manager Library and Archives Canada Digital Resource Division (Formerly – Cartography, Architecture and Geomatics) 550 de la Cité Blvd Gatineau, Quebec K1A 0N4 Canada Telephone : 819-934-6821 Fax : 819-934-6818 david.brown@lac-bac.gc.ca
Heather L Tompkins Geomatics Technician Library and Archives Canada Digital Resource Division (Formerly – Cartography, Architecture and Geomatics) 550 de la Cité Blvd Gatineau, Quebec K1A 0N4 Canada Telephone : 819-934-6824 Fax : 819-934-6818 heather.tompkins@lac-bac.gc.ca	***Marc Cockburn*** Archivist Library and Archives Canada Acquisitions Division (Formerly – Cartography, Architecture and Geomatics) 550 de la Cité Blvd Gatineau, Quebec K1A 0N4 Canada Telephone : 819-934-6823 Fax : 819-934-6818 marc.cockburn@lac-bac.gc.ca	***Cindy Mitchell*** Program Advisor CGDI Operational Policies Team Natural Resources Canada 615 Booth Street, 06Ath Floor, Room: 613A Ottawa, Ontario K1A 0E9 Canada Telephone : 613-995-4876 Cindy.Mitchell@nrcan-rncan.gc.ca
HAL: ***Ed Kennedy*** 613-237-2220, ext. 334 ekennedy@hal.ca ***Tracey P. Lauriault*** 613-234-2805 tlauriau@gmail.com	Regrets: ***Pam Armstrong*** Manager Library and Archives Canada Digital Repository Services and Standards	

Appendix B: Geospatial Data Archive Initiatives in Canada and Abroad

Table 10 provides a more detailed analysis of the ten Canadian and two international geospatial data archive initiatives discussed in Chapter 7.

Table 10: Geospatial Data Archive Initiatives in Canada and Internationally

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
Canadian					
1. National Climate Data and Information Archive (http://climate.weatheroffice.gc.ca/Welcome_e.html)	*Environment Canada* EC maintains the archive, and data are contributed by weather stations from across the country. These are organized by province & territory, service or database.	A variety of climatological data are archived at different time intervals. (http://climate.weatheroffice.gc.ca/prods_servs/index_e.html) Databases/datasets include: - Climate Data Online - Almanac Data - Climate Normals & Averages - Climate Summaries - WMO Volume A Report - Canada - IDF Files: Short-duration rainfall intensity-duration-	Public access; some cost recovery. Data are accessed via an interactive map (http://climate.weatheroffice.gc.ca/climateData/canada_e.html), or an online customized search tool (http://climate.weatheroffice.gc.ca/advanceSearch/searchHistoricData_e.html) and can be special ordered at a cost recovered rate. Download formats are csv and xml. Data are ordered per site, and the acquisition of data from multiple sites requires a special order.	Descriptions of the data are available on the site in a schema specific to this science (http://climate.weatheroffice.gc.ca/prods_servs/documentation_index_e.html) The database contains latitude and longitude of each observation (Degrees & minutes). (http://climate.weatheroffice.gc.ca/prods_servs/documentation_index_e.html). Weather Station information descriptions are also provided (http://climate.weatheroffice.gc.ca/prods_servs/	Not all data are quality controlled. Some Partner stations provide data to the Archives which do not undergo Archive review processes. These are marked with a "‡" symbol next to the date of the observation. A data quality explanation and disclaimer is provided when data are accessed (http://climate.weatheroffice.gc.ca/climateData/dataQuality_e.html). The data are also accompanied by a data completeness legend.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
		frequency (IDF) More specificity is provided in the FAQ (http://climate.weatheroffice.gc.ca/FAQ_e.html) - Historical Publications-National Climate Archives. - Historical Radar - Canadian Weather Energy and Engineering Datasets (CWEEDS) - Canadian Weather year for Energy Calculation (CWEC)	License Agreement for Use of Environment Canada Data (http://climate.weatheroffice.gc.ca/prods_servs/tables/attachment1_e.html). In addition, the following is provided when the data are accessed: "Information presented on this web site is considered public information and may be distributed or copied. An appropriate byline acknowledging Environment Canada is requested." (http://climate.weatheroffice.gc.ca/climateData/dataQuality_e.html)	metstat_e.html)	
2. National Water Data Archive (http://www.ec.gc.ca/rhc-wsc/default.asp?lang=En&n=9018B5EC-1)	*Environment Canada, Water Survey of Canada (WSC)* in partnership with the provinces, territories and other agencies (http://www.ec.gc.ca/rhc-wsc/default.asp?lang=En&n=2472CB39-1). WSC operates over 2,500 active hydrometric gauges across the country.	Data Products and Services: - Archived Hydrometric Data - Water Level and Streamflow Statistics - Sediment Data Station – Information Downloads - Environment Canada Data Explorer - HYDAT Database - Hydrological Modelling	Public access. Dbases can be downloaded directly from the site at no cost. Data are searched from a basic or an advanced interface (http://www.wsc.ec.gc.ca/applications/H2O/index-eng.cfm?stype=station). Data can also be downloaded with Environment Canada's Data Explorer (http://www.ec.gc.ca/rhc-	The data are documented in English and French in a PDF. Standard metadata schema were not found on the site; however, the data are associated with a description in an explanation with the data formats and the search interface is indicative of what would be considered of value to scientists in this field.	The data are quite well documented in reference guides and reference indexes. When data are accessed the following disclaimer is found at the base of the data table: "*In no event shall Environment Canada be liable for damages whatsoever (including, without limitation, damages for loss of business profits, business interruption, loss of*

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
	Data are collected and compiled by WSC's eight regional offices. WSC is the national authority responsible for the collection, interpretation and dissemination of standardized water resource data and information in Canada. Centrally managed & housed in HYDEX and HYDAT dbases. The WSC operates the network of hydrometric monitoring stations on behalf of most provinces and all territories, under federal-provincial or federal-territorial agreements. In the case of Quebec, the province collects water quantity data under a similar agreement.	Tool Water information collected through the National Hydrometric Program. These data include: daily and monthly mean flow, water level and sediment concentration for over 2,500 active and 5,500 discontinued hydrometric monitoring station across Canada. HYDEX is the inventory and HYDAT is the relational dbase which contains the data.	wsc/default.asp?lang=En&n=0A47D72F-1) and a trademarked Modeling Tool (http://www.nrc-cnrc.gc.ca/eng/ibp/chc/software/kenue/green-kenue.html). Data are available in different formats: CSV Format No license agreement was found with these data; however, the following License Agreement for Use of Environment Canada data was later discovered (http://climate.weatheroffice.gc.ca/prods_servs/tables/attachment1_e.html).	Further, the data are accompanied with a comprehensive online Reference Index: - Station Name Index - Station Number Index - Reference Index Online These are also accompanied with Training Material	*business information, or other pecuniary loss) arising out of the use of, or inability to use this Environment Canada product, even if Environment Canada has been advised of the possibility of such damages.*"
2.1 Archived Sediment Data Example of an archived dataset of the National Water Data Archive.	*Environment Canada, WSC* Public access to historical suspended sediment data, which were collected at over 800 locations in Canada through	Sediment data, including information about the station and other field specific attributes.	Public Access. In addition to what is indicated above: Data are accessed via a basic or advance search tool. Users must search by station. Searches are	In addition to what is described above: Data are described with a data description field when data are downloaded (ex.:http://www.wsc.ec.gc.ca/sedat/sedflo/index_	In addition to what is described above: Data quality not indicated but a disclaimer is provided: "*In no event shall Environment Canada be liable for damages whatsoever (including,*

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
(http://www.wsc.ec.gc.ca/sedat/sedflo/index_e.cfm?cname=mainStation_e.cfm)	federal-provincial and federal-territorial cost-sharing agreements.		by province & territory, drainage basin, regional offices, lat & long. The data can be viewed in a browser and can be downloaded in printed formats. These can be viewed in a table or a chart. No license agreement was found with these data however the following License Agreement for Use of Environment Canada data was later discovered (http://climate.weatheroffice.gc.ca/prods_servs/tables/attachment1_e.html).	e.cfm?cname=WEBfrmInstReport_e.cfm&RequestTimeout=300)	*without limitation, damages for loss of business profits, business interruption, loss of business information, or other pecuniary loss) arising out of the use of, or inability to use this Environment Canada product, even if Environment Canada has been advised of the possibility of such damages"*
3. Integrated Science Data Management (ISDM) Wave Data Archive (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/index-eng.html)	*Department of Fisheries and Oceans (DFO)* ISDM has been collecting, archiving and providing spectral wave data on behalf of networks within the Canadian area of interest (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/overview-apercu-eng.htm).	Spectral wave data. On a daily basis data are processed and reported on the GOES (satellite) and ARGOS buoys networks within the Canadian area of interest. 6 million observed wave spectra from 500 locations in the Canadian area of interest (35 to 90 degrees north and 40 to 180 degrees west), all of which are available	Public Access. Data are available directly from the site and are searched with a simple interface tool. Data searches can be viewed in tables, in a browser or downloaded in CSV formats and can also be viewed in a map or plotted into a chart. Data can also be ordered via an online order form (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-	See Data Quality for qualifications. No metadata accompany data downloaded in zip files nor in the browser view. The Search tools do provide some insight into what elements are considered of value to this scientific community. Historical MSC Buoy Status Reports are provided and give	Data Quality and data elements are described by downloaded file format types. Files are fully described (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/formats/csv-eng.htm; http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/formats/FormatB-

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
	ISDM's mandate is to manage and archive ocean data collected by DFO, or acquired through national and international programmes conducted in ocean areas adjacent to Canada, and to disseminate data, data products, and services to the marine community in accordance with the policies of the Department. These data are part of the DFO International Science Strategy (http://www.dfo-mpo.gc.ca/science/Publications/index-eng.htm) which is managed by Canadian Science Advisory Secretariat (CSAS) (http://www.dfo-mpo.gc.ca/science/advice-avis/index-eng.html). These data are also managed within a Management Policy for Scientific Data (http://www.dfo-	for direct download from this web site.	gdsi/request-commande/form-eng.asp). Data formats: - Format B for Non-Directional Spectral Wave Data (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/formats/FormatB-eng.htm) - Format of the Co and Quad File (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/formats/co-quad-eng.htm)	context, (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/index-eng.htm) while the data format descriptions are also helpful.	eng.htm, and http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/formats/co-quad-eng.htm) ISDM performs a quality inspection (QC) of each observed wave spectra prior to update into the database. Flags are assigned to the observed and derived parameters reflecting data quality. QC is performed by examining the energy distribution of the power spectrum and comparing relative values of significant wave height and peak period between neighboring buoys.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
	mpo.gc.ca/science/data-donnees/policy-politique-eng.htm).				
4. IOS/OSD Data Archive (http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/index-eng.htm)	*Department of Fisheries and Oceans* (DFO). Contains the holdings of oceanographic data generated by the Institute of Ocean Sciences and other agencies and laboratories, including the Institute of Oceanography at the University of British Columbia and the Pacific Biological Station. ***Note:*** It is presumed that these data would fall under the DFO Management Policy for Scientific Data as previously described. (http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm)	The directory contains: (http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/index-eng.htm#DataArchiveOverview): - Water property profiles - Moored instrument data - Satellite Images - Seasonal mean T and S - BC Lighthouse data - Buoy data - Line P data - SST data - Gulf of Alaska data	Some databases are restricted, others are public. Some data can be viewed online in plots, charts, tables, and maps. "Local" (DFO employees) can access these data by connecting to the DATA_LIBRARY disk service. Permission to attach to this disk service is not granted by default - contact with the Oceanographic Data Manager is required to arrange access. When access is granted, users will be allowed READ-ONLY access; users are unable to modify, delete or create files in this area. ASCII files in IOS HEADER format. A searchable database of "header" information (time, date, latitude, longitude, parameters sampled etc.) for all the data in the DATA_LIBRARY. This database may be	Data descriptions vary according to the datasets being described. e.g. The ***Line P Program data*** are accompanied by a historical paper, maps & illustrations, sampling methods, collaborators and scientists, and publications (http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/line-p/linep/index-eng.htm).	Disclaimers and explanations are provided with and are specific to the varied datasets. e.g. ***Data from BC Lighthouses Warning***: "*The daily sampling strategy at the BC Lighthouse Stations was designed long ago by Dr. John P. Tully. We have chosen not to change the strategy in the interests of a homogeneous data set. Sampling occurs at or near the daytime high tide. This means, for example, that if an observer starts sampling one day at 6 a.m., and continues to sample at the daytime hightide, as instructed, then on the 2nd day he/she will take samples at about 06:50 the next day, 07:40 the day after etc. When the daytime high tide gets close to 6 p.m. then it snaps back to 6 a.m. and the cycle starts again. Since there is a diurnal*

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
			searched and "pointers" to the location of the data in the DATA_LIBRARY will be returned. Simple queries can be carried out using this WEB-based application. The Satellite Images 'Quick Look' Archive for instance is available to the public online: IOS/Ocean Sciences Data Inventory for Satellite images (AVHRR and SeaWiFS) (http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/search-recherche/satellite-eng.asp).		*signal in sea-surface temperature the sampling creates a 14-day signal as an artifact)*" (http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/lighthouses-phares/index-eng.htm). The ***Line P Program data*** are accompanied by data quality flags (http://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/line-p/linep/index-eng.htm).
5. Lithoprobe Data Archive (http://www.gdr.nrcan.gc.ca/seismtlitho/archive/index_e.php)	*Natural Resources Canada, Earth Science Sector (ESS), Geological Survey of Canada* Lithoprobe is a major national research project in the earth sciences. It combines multidisciplinary earth science studies of the Canadian landmass and surrounding offshore margins. It integrates modern	Seismic and magnetotelluric (MT) data Seismic Transect data are available in: - maps, - data tape listings, - images of seismic sections, - SEGY files of seismic sections, and - geometry Magnetotelluric data by	Public Access and some cost recovery. - Data can be accessed directly from the site. - Data can be purchased (http://www.gdr.nrcan.gc.ca/seismtlitho/archive/datarequest_e.php#dr) by filling out a data request form. The data are in 'edi' file format and are kept in zip files. Once permission is granted data can be	Data collected as part of the Lithoprobe project (http://www.lithoprobe.ca/). Seismic data and magnetotelluric data collected according to transects (aka study areas). More details are provided in accompanying papers (http://gdr.nrcan.gc.ca/seismtlitho/archive/transe	There is no information available on the site. Many publications and some explanation of the data are available on the Lithoprobe site and it is presumed that the accompanying papers would discuss methodological issues, which would in turn include explanations of data quality.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
	geophysical, geological and geochemical concepts, methods and technology to extend to depth, and back in time, knowledge of the lithosphere (the upper rigid 100 to 250 km of the planet) in various key transects or study areas. Ten such transects form the scientific bases of the project.	transect: - station numbers and the corresponding co-ordinates - viewed in a static site map	accessed via an ftp site	ct_synthesis_e.php)	
6. The National Soil DataBase (NSDB) (http://sis.agr.gc.ca/cansis/nsdb/intro.html)	*Agriculture and Agri-Food Canada; Canadian Soil Information Service* NSBD is the set of computer readable files which contain soil, landscape, and climatic data for all of Canada. It serves as the national archive for land resources information that was collected by federal and provincial field surveys, or created by land data analysis projects.	- National Ecological Framework (EcoZones, EcoRegions, and EcoDistricts) - Soil Map of Canada / Land Potential DataBase (LPDB) - Agroecological Resource Areas (ARAs) - Soil Landscapes of Canada (SLC) - Canada Land Inventory (CLI) - Detailed Soil Surveys	Data are freely available online in a variety of formats: - Web maps - Printed maps - GIS spatial data - GIS attribute data - Web Map Service (WMS) - Web Feature Service (WFS) - Table Joining Service (TJS) - online mapping services - Arc Export - ASCII format - ARC/INFO export and shapefile formats in Decimal Degrees (DD).	Data descriptions vary dependent on the type of data. The following types of descriptions were found: - Narrative Descriptions - Reports - User's Handbook - Dataset Descriptions - Technical Bulletins - Data Model - General Overview - Legacy Papers There is also a Canadian Soil Publications Archive which contains digitized historical documents (http://sis.agr.gc.ca/cans	Data Quality Disclaimer "*Although every effort has been made to ensure the accuracy, currency and reliability of the content, Agriculture and Agri-Food Canada assumes no liability deemed to have been caused directly or indirectly by this data.*" (http://sis.agr.gc.ca/cansis/nsdb/slc/v3.1.1/intro.html).

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
			- dbf format. Copyright: "*Anyone may use these data free of charge, provided they acknowledge Agriculture and Agri-Food Canada's authorship. AAFC retains exclusive rights, title, and ownership of this data. Copying and redistribution of part or all of this dataset is acceptable, provided that the contents of this file are included with the redistribution. AAFC may be unable to answer queries about these data if they were obtained through a third party.*"	is/publications/intro.html).	
7. Earth Observation Data Services (http://eods.nrcan.gc.ca/about_e.php)	*NRCan, ESS, Canada Centre for Remote Sensing, EODS* EODS provides satellite data to NRCan programs and other governments of Canada. EODS provides data reception through two receiving stations, one in Prince Albert, Saskatchewan, and the other in Gatineau, Quebec. "There is a state-of-the-	Data from several satellite sensors in the National Earth Observation archives dating back to 1972. Imagery from the following sensors: MOS RADARSAT ENVISAT SPOT NOAA LANDSAT	Public and cost recovered. Data are accessed through the Satellite Acquisition Services (SAS) group at the Canada Centre for Remote Sensing (CCRS) in Ottawa. Some data can be acquired from a number of catalogs. Regarding data distribution licensing, a variety of licenses are	Metadata for the Earth Observation Data Catalogue, enabling discovery and access to National EO data archive. Looks like ISO19115 but unsure as not specified.	Dataset quality is explained with accompanying metadata.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
	art archiving, processing and data dissemination systems".	JERS ERS SEASAT Earth Observation data in support of Canadian public good as well as private sector requirements.	available contingent on whether or not data are purchased, or acquired via the different catalogs where a subset of these remotely sensed data are disseminated (e.g., GeoBase and GeoGratis Unrestricted User license).		
8. National WaveForm Archive (NWFA) (http://earthquakescanada.nrcan.gc.ca/stndon/NWFA-ANFO/index-eng.php)	*NRCan, ESS, Earthquakes Canada* A public Web-accessible seismic data archive.	- Event-related waveform - Digitized events - Continuous data from the Yellowknife Array (YKA) - Continuous data from the Canadian National Seismograph Network (CNSN) - Event-related CNSN - Quasi-continuous data from various special deployments - Continuous data from the POLARIS network - Events from the Global Infrasound Archive (GINA) - Numerical acceleration data from Strong Motion deployments	Public Access Data can be downloaded directly from the site. These data are under a license, a distribution policy and attribution is requested: - GSC END-USER AGREEMENT FOR DIGITAL DATA (http://earthquakescanada.nrcan.gc.ca/stndon/licence-eng.php) - CNSN Waveform Data Distribution Policy (http://earthquakescanada.nrcan.gc.ca/stndon/NWFA-ANFO/dist_policy-eng.php) - Citation Request (http://earthquakescanada.nrcan.gc.ca/cite-eng.php) "Canadian Archive" (CA)	Data are described by data type and descriptions are available in notes in the NWFA Archive (e.g. of GSC http://earthquakescanada.nrcan.gc.ca/stnsdata/nwfa/formats/shi_msgs.pdf).	Could not find.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
			format, a compressed binary format developed by the GSC originally for the CNSN, and now used more widely as a common archive format.		
9. Geomagnetism Summary Plots : Archives (http://geomag.nrcan.gc.ca/common_apps/ssp-2-eng.php)	*NRCan, ESS, Geological Survey of Canada (GSC)* The Geomagnetic Monitoring Service of the Geological Survey of Canada (GSC) operates a network of 13 Magnetic Observatories across Canada.	Summary plots of the one minute variations of the geomagnetic field are available for all Canadian magnetic observatories. These plots show the X (northward), Y (eastward) and Z (vertical downward) components of the magnetic field. Universal Time (UT) is used. daily plot of activity at individual Canadian observatories: - Magnetic field (B) - Rate of change (dB/dt) - Geoelectric field - 24 hour summary plots of activity at all Canadian observatories: archive - other dates - digital data from Canadian magnetic observatories - magnetic activity	Public Access. Available from the website with a simple online search tool. Can be viewed on the site or downloaded in xml format. Data are available in tables, and plots.	Data are described by regional observatory (e.g. http://gsc.nrcan.gc.ca/geomag/obs/blc_e.php). Once data are accessed these are accompanied by explanatory notes (e.g. http://geomag.nrcan.gc.ca/common_apps/plt/mp-i-eng.php#b_plot) Metadata standards were not found. Sensor technology are also explained (e.g. http://gsc.nrcan.gc.ca/geomag/obs/canmos_e.php) A license was not found with the data, however there is a policy on the GSC Site for Geoscience Data Repository (http://edg.rncan.gc.ca/terms_e.php)	??

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
		indices produced from Canadian observatories			
10. *The Canadian Ice Service Archive (CISA)* (http://www.ec.gc.ca/glaces-ice/default.asp?lang=En&n=0A70E5EB-1)	Environment Canada, Canadian Ice Service	Historical ice data that is updated daily and Environment Canada/CIS staff often draw upon this data to meet the specific needs of their ice clients. Ice and iceberg charts in several types of media (paper, microfilm, and digital formats); A large collection of aircraft and satellite imagery; and Bulletins and other text documents. Available but not online: Daily analysis charts (BW since 1999 also available online) Regional ice analysis charts (BW Gif and .e00 also available online) Daily iceberg analysis charts Image analysis charts Observation charts	Directly from the website in an easy 3 step process. Data are accessed in zip files. Data can be viewed in maps and atlases.	No metadata made available with the document except in the map legend.	Undefined.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
		Historical ice charts Ice forecast bulletins			
11. *System of Agents for Forest Observation Research with Advanced Hierarchies (SAFORAH)* (http://www.saforah.org/)	Canadian Forest Service (CFS), University of Victoria and other academic and government partners.	Provides support to collaborative research and sharing of remote sensing data and information products within the Canadian government, universities, scientific institutions and the public. The Earth Observation for Sustainable Development (EOSD) Over 870 circa 2000 orthorectified Landsat images were used to product 630 EOSD land cover products based on the 1:250,000 NTS (National Topographic Series) map sheets. Canada Landsat Mosaics Full resolution compressed Canada image mosaics are available for the circa 1990 and 2000 Landsat coverage on a CFS FTP site. Remote sensing data from spaceborne and	Users can obtain access to remote sensing data and information products in SAFORAH by using the Catalogue and User Data Ordering System (CUDOS) or an Open Geospatial Consortium (OGC) portal (http://mars.csc.uvic.ca:8080/ogc/). RADARSAT-1 End User License Agreement (http://mars.csc.uvic.ca:8080/ogc/radarsat-1_licenseAgreement.html) CSW, WPS services. Some data may be restricted to specific projects or because of vendor licensing agreements.	Catalogue links broken, unable to ascertain	Catalogue links broken, unable to ascertain

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
		airborne sensors			
International					
12. The National Geospatial Digital Archive ***University libraries of University of California at Santa Barbara (UCSB) and Stanford University*** (http://www.ngda.org/) See Collection Policies (http://www.ngda.org/research/Collections/NGDA_Collection_Development_Policy_11_06_final.doc)	Library of Congress formed The National Digital Information Infrastructure and Preservation Program (NDIIPP). The UCSB and Stanford are leading the formation of the National Geospatial Digital Archive (NGDA), a collecting network for the archiving of geospatial images and data. The Objectives of the Project: Create a new national federated network committed to archiving geospatial imagery and data. Investigate the proper and optimal roles of such a federated archive. Collect and archive major segments of at-risk digital geospatial data and images.	US digital geospatial data," defined as digital items, displayed as graphics, that are georeferenced or are geographically identified. These are primarily composed of: digital maps; remotely sensed images (e.g., aerial photographs; data collected by satellite sensors); datasets (e.g. shapefiles, layers, geodatabases, etc.); atlases; globes (celestial and terrestrial); aerial views (e.g., panoramas); block diagrams; geologic sections; topographic profiles; etc." See Collection Policies (http://www.ngda.org/research/Collections/NGDA_Collection_Development_Policy_11_06_final.doc)	NGDA received funding starting in 2004 to preserve data but not to make it accessible. Stanford is currently constructing the access piece to this project. It is not expected to be available for another 6 months to a year. NGDA's mission is to be a collecting network. As such, collaboration with other institutions is expected and necessary. Because digital geospatial data sets require large amounts of server space, the cooperation of many institutions will be necessary to build an extensive collection. Cooperation agreements written specifically to govern the collecting areas of the partners should include: The collecting areas for each participating institution.	Metadata are considered essential and the Collections Policies provides a series of required fields.	Data Quality information e.g. FGDC metadata elements such as attribute accuracy and completeness report.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
	Develop best practices for the presentation of archived digital geospatial data. Develop partner communication mechanisms for the project and then ongoing. Develop a series of policy agreements governing retention, rights management, obligations of partners, interoperability of systems, and exchange of digital objects.		The frequency of updates and versioning for each dataset. The length of the agreement. The type and level of access to be provided to the collected materials. A set interval to review the collection agreement. A glossary.		
13. UK Data Archive and the British Library (http://www.data-archive.ac.uk/home) and (http://www.bl.uk/reshelp/findhelprestype/maps/digitalmapping/ordnancesurvey/osdigitalmaps.html)	**UK Data Archive:** Helps creators manage data, provide the means for data to be deposited, provide users the means to discover and access data. Important Guidelines & Policies: - **Preservation Policy** (http://www.data-archive.ac.uk/media/54776/ukda062-dps-	**UK Data Archive:** UK's largest collection of social and economic data. Centre of expertise in the areas of acquiring, curating and providing access to data. Designated a Place of Deposit by the National Archives to curate public records. Acquisition of high quality data from the academic, public, and commercial sectors host key	**UK Data Archive:** Data are searched via the Economic and Social Data Service Data Portal (http://www.esds.ac.uk/Lucene/Search.aspx) . Users must be registered to access data. **Catalogue search:** Help on searching	**UK Data Archive:** Data cataloguing metadata are used for discovery purposes (e.g. http://www.esds.ac.uk/findingData/snDescription.asp?sn=4632) **British Library:** Presumed to be the metatada standard adopted by the national OSs.	**UK Data Archive:** Presumed to be available when data are accessed. The discovery data provides lineage and methodology information but not elements on data quality. **British Library:** Presumed to be the data quality standards adhered to by the national OSs.

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
	preservationpolicy.pdf) - ***Life-Cycle Management*** (http://www.data-archive.ac.uk/create-manage/life-cycle) - Consent and Ethics which includes: legal, ethical, anonymization, access restrictions (http://www.data-archive.ac.uk/create-manage/consent-ethics) - ***Guides for researchers*** (http://www.data-archive.ac.uk/create-manage/starting-research) - **Advice and Training** (http://www.data-archive.ac.uk/create-manage/advice-training) British Library is a partner and collabotor instition of the UK Data Archive which archives UK and Northern Ireland OS Digital Mapping.	national and international survey data and qualitative data. Also host a number of data services such as the Census portal and the History Data Service. Providing secure access to disclose sensitive data through a Secure Data Service. Engaged in a number of data management initiatives, running the Rural Economy and Land Use Programme (RELU) Data Support Service. Also provide data curation for other organisations. **British Library:** Large-scale mapping of the nation in digital format, since 1998 for Ordnance Survey (OS) and 2004 for OS Northern Ireland (OSNI) - Annual snapshots of a subset of UK OS Land-Line™, 1998 mapping data product in NTF format, and 2006 GML-based	Variables search Browse by subject Major studies Major depositors New releases HASSET thesaurus Other archives Terms and conditions are specific to record sets: http://www.esds.ac.uk/orderingdata/termsandConditions.asp ***British Library OS Maps:*** The digital maps may be viewed in LDL reading rooms using mapping applications designed specifically for this set of historical and current OS and OSNI digital data. In the British Library, it is available on the public terminal located closest to the Enquiry Desk in the Maps Reading Room. The web application is launched via the link "Ordnance Survey digital data" in the Library's listing of Electronic resources (http://www.bl.uk/eresou		

Name of Initiative	Organization	Type of data	Data Access	Data Description	Data Quality
		mapping product, OS MasterMap™ - Topography and Inegrated Transport Network layers - Northern Ireland Land & Property Services' online map tool, Geohub NI.	rces/dbstptitles/eresourceso.html#O).		

Appendix C: Acts for Which the Minister of Natural Resources is Responsible

Arctic Waters Pollution Prevention Act
An Act to prevent pollution of areas of the arctic waters adjacent to the mainland and islands of the Canadian arctic

Canada Foundation for Sustainable Development Technology Act
An Act to establish a foundation to fund sustainable development technology

Canada Labour Code
An Act to consolidate certain statutes respecting labour

Canada Lands Surveyors Act
An Act respecting Canada Lands Surveyors

Canada Lands Surveys Act
An Act respecting the surveys of public lands of Canada

Canada-Newfoundland Atlantic Accord Implementation Act
An Act to implement an agreement between the Government of Canada and the Government of Newfoundland and Labrador on offshore petroleum resource management and revenue sharing and to make related and consequential amendments

Canada-Nova Scotia Offshore Petroleum Resources Accord Implementation Act
An Act to implement an agreement between the Government of Canada and the Government of Nova Scotia on offshore petroleum resource management and revenue sharing and to make related and consequential amendments

Canada Oil and Gas Operations Act
An Act respecting oil and gas operations in Canada

Canada Petroleum Resources Act
An Act to regulate interests in petroleum in relation to frontier lands, to amend the Oil and Gas Production and Conservation Act and to repeal the Canada Oil and Gas Act

Canadian Ownership and Control Determination Act
An Act respecting Canadian ownership and control determination

Cape Breton Development Corporation Act
An Act to establish the Cape Breton Development Corporation

Cape Breton Development Corporation Divestiture Authorization and Dissolution Act
An Act to authorize the divestiture of the assets of, and to dissolve, the Cape Breton Development Corporation, to amend the Cape Breton Development Corporation Act and to make consequential amendments to other Acts

Cooperative Energy Act
An Act to establish the Cooperative Energy Corporation and the Cooperative Energy Development Corporation

Department of Natural Resources Act
An Act to establish the Department of Natural Resources and to amend related Acts

Energy Administration Act
An Act to provide for charges, compensation and pricing in respect of certain energy sources and for the administration and control of other matters respecting energy sources in Canada

Energy Efficiency Act
An Act respecting the energy efficiency of energy-using products and the use of alternative energy sources

Energy Monitoring Act
An Act respecting energy monitoring

Energy Supplies Emergency Act
An Act to provide a means to conserve the supplies of energy within Canada during periods of national emergency caused by shortages or market disturbances affecting the national security and welfare and the economic stability of Canada

Explosives Act
An Act respecting the manufacture, testing, sale, storage, transportation and importation of explosives and the use of fireworks

Export and Import of Rough Diamonds Act
An Act providing for controls on the export, import or transit across Canada of rough diamonds and for a certification scheme for the export of rough diamonds in order to meet Canada's obligations under the Kimberley Process

Forestry Act
An Act respecting forestry development and research

Hibernia Development Project Act
An Act respecting the Hibernia Development Project and to amend certain Acts in relation thereto

International Boundary Commission Act
An Act respecting the International Boundary Commission

National Energy Board Act
An Act to establish a National Energy Board

Northern Pipeline Act
An Act to establish the Northern Pipeline Agency, to facilitate the planning and construction of a pipeline for the transmission of natural gas from Alaska and Northern Canada and to give effect to the Agreement between Canada and the United States of America on principles applicable to a Northern natural gas pipeline

Nuclear Energy Act
An Act relating to the development and utilization of nuclear energy

Nuclear Fuel Waste Act
An Act respecting the long-term management of nuclear fuel waste

Nuclear Liability Act
An Act respecting civil liability for nuclear damage

Nuclear Safety and Control Act
An Act to establish the Canadian Nuclear Safety Commission and to make consequential amendments to other Acts

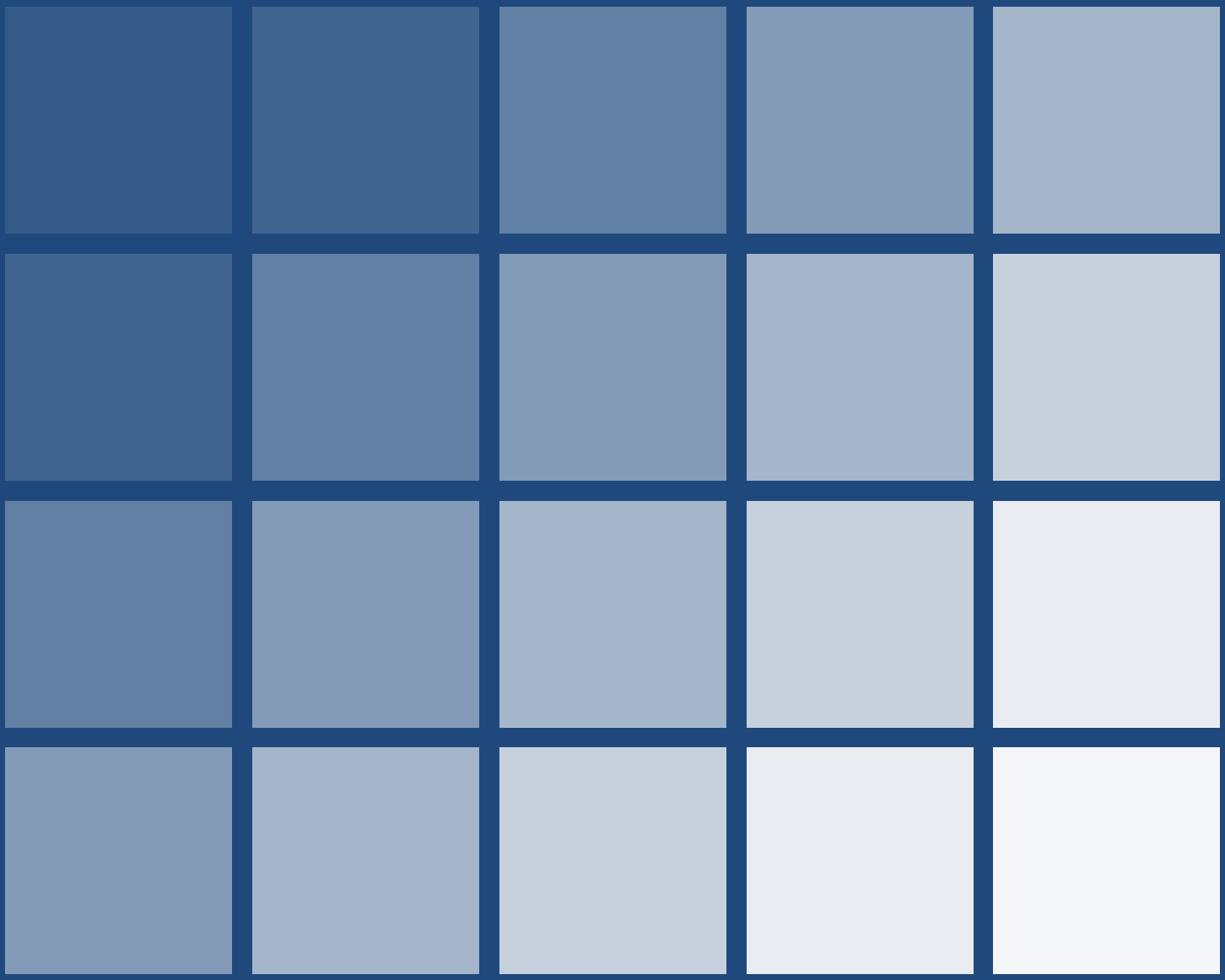
Oil Substitution and Conservation Act
An Act respecting oil conservation and the substitution for oil of other energy sources

Resources and Technical Surveys Act
An Act respecting resources and technical surveys

Not in force

Greenhouse Gas Technology Investment Fund Act
An Act to establish the Greenhouse Gas Technology Investment Fund for the reduction of greenhouse gas emissions and the removal of greenhouse gases from the atmosphere

Date Modified: 2010-10-14



Hickling Arthurs Low Corporation

150 Isabella Street
Suite 1300
Ottawa, ON
K1S 1V7

Phone: 613.237.2220
Fax: 613.237.7347
Email: hal@hal.ca