



Natural Resources  
Canada

Ressources naturelles  
Canada



## CANADIAN GEOSPATIAL DATA INFRASTRUCTURE INFORMATION PRODUCT 36e

### **Geospatial Data Preservation Primet**

GeoConnections  
Hickling Arthurs Low Corporation

2013

©Her Majesty the Queen in Right of Canada, as represented by the Minister of Natural Resources Canada, 2013

Canada

---

## ACKNOWLEDGEMENTS

---

GeoConnections would like to acknowledge the contributions made to the Primer by Hickling Arthurs Low Corporation – Tracey P. Lauriault and Ed Kennedy (researchers, writers and editors), Yvette Hackett (subject matter specialist), and Marcel Fortin (external reviewer). The staff at GeoConnections who provided management, input and direction for the project were Cindy Mitchell and Simon Riopel, in addition Jenna Findlay and Nadia Eckardt provided assistance in assembling this document.

# Table of Contents

<b>1. Preamble.....</b>	<b>1</b>
<b>2. Introduction .....</b>	<b>3</b>
2.1 Data Archives and Preservation.....	3
2.1.1 Basic Terminology.....	4
2.1.2 Trusted Digital Repositories .....	5
2.1.3 The Importance of Metadata.....	6
2.1.4 The Impacts of Technological Change .....	7
2.2 Geospatial Data Creator Guidelines.....	8
2.3 Geospatial Data Preserver Guidelines.....	10
<b>3. Legislation and Policy Affecting Archiving and Preservation .....</b>	<b>13</b>
3.1 Archival and Preservation Responsibilities .....	13
3.2 Federal Acts and Regulations .....	15
3.3 Federal Policies and Directives.....	17
<b>4. Archiving and Preservation Frameworks.....</b>	<b>18</b>
4.1 Introduction.....	18
4.2 Reference Model for an Open Archival Information System.....	19
4.3 European Long Term Data Preservation Common Guidelines.....	21
4.4 Trustworthy Repositories Audit & Certification: Criteria and Checklist .....	24
<b>5. Geospatial Data Preservation Examples .....</b>	<b>26</b>
5.1 Case Study: Earth Observation Data Management System (EODMS) .....	26
5.1.1 Introduction.....	26
5.1.2 Operational Model in Use – Implementation to Date .....	27
5.1.3 Challenges Encountered.....	32
5.1.4 Lessons Learned.....	33
5.2 Profile: Ontario Geographic Information Archive (GIA) .....	34
5.2.1 Introduction.....	34
5.2.2 Operational Model in Use.....	34
5.2.3 Good Practices .....	35
5.3 Profile: Integrated Science Data Management (ISDM).....	37

5.3.1	Introduction.....	37
5.3.2	Operational Model in Use.....	37
5.3.3	Good Practices.....	38
5.4	Profile: International Polar Year (IPY) Data Preservation.....	39
5.4.1	Introduction.....	39
5.4.2	Operational Model in Use.....	39
5.4.3	Challenges Encountered.....	41
5.4.4	Lessons Learned.....	42
<b>6.</b>	<b>Establishing a Geospatial Data Preservation System.....</b>	<b>43</b>
6.1	Introduction.....	43
6.2	Establishing the System’s Scope and Objectives.....	44
6.3	Defining the System’s User Community.....	47
6.4	Acquiring and Managing Resources.....	49
6.5	Preservation Planning.....	50
6.6	Developing Policies and Procedures.....	52
6.7	Appraising Records for Preservation Value.....	54
6.8	Acquiring and Ingesting Records.....	55
6.9	Preserving Records.....	56
6.10	Describing the Archival Metadata.....	58
6.11	Managing and Maintaining Records.....	59
6.12	Providing Access to Records.....	60
<b>7.</b>	<b>Challenges and Solutions.....</b>	<b>62</b>
<b>8.</b>	<b>Conclusions.....</b>	<b>66</b>
<b>A.</b>	<b>References.....</b>	<b>68</b>
<b>B.</b>	<b>Glossary of Terms.....</b>	<b>76</b>
<b>C.</b>	<b>Acts and Regulations Specific to the Preservation and Management of Government Information.....</b>	<b>82</b>
<b>D.</b>	<b>Information Management Policies and Directives of the Treasury Board of Canada Secretariat (TBS).....</b>	<b>84</b>
<b>E.</b>	<b>Gearchiving Business Planning Guidebook Highlights.....</b>	<b>87</b>

# 1. Preamble

**This primer is one in a series of Operational Policy documents being developed by GeoConnections. It is intended to inform Canadian Geospatial Data Infrastructure ([CGDI](#)) stakeholders about the nature and scope of digital geospatial data archiving and preservation and the realities, challenges and good practices of related operational policies.**

Burgeoning growth of online geospatial applications and the deluge of data, combined with the growing complexity of archiving and preserving digital data, has revealed a significant gap in the operational policy coverage for the Canadian geospatial data infrastructure (CGDI).

Currently there is no commonly accepted guidance for CGDI stakeholders wishing or mandated to preserve their geospatial data assets for long-term access and use. More specifically, there is little or no guidance available to inform operational policy decisions on how to manage, preserve and provide access to a digital geospatial data collection. The preservation of geospatial data over a period of time is especially important when datasets are required to inform modeling applications such as climate change impact predictions, flood forecasts and land use management. Furthermore, data custodians may have both a legal and moral responsibility to implement effective archiving and preservation programs.

The *GeoConnections* program is a national initiative led by Natural Resources Canada. GeoConnections supports the integration and use of the Canadian Geospatial Data Infrastructure (CGDI).

The *CGDI* is an on-line resource that improves the sharing, access and use of Canadian geospatial information – information tied to geographic locations in Canada. It helps decision makers from all levels of government, the private sector, non-government organizations and academia make better decisions on social, economic and environmental priorities.

Based on research and analysis of the Canadian legislative framework and current international practices in digital data archiving and preservation, this primer provides guidance on the factors to be considered and the steps to be taken in planning and implementing a data archiving and preservation program. It describes an approach to establishing a geospatial data archives based on good practices from the literature and Canadian case studies.

This primer will provide CGDI stakeholders with information on how to incorporate archiving and preservation considerations into an effective data management process that covers the entire life cycle (DCC, 2013) (LAC, 2006) of their geospatial data assets (i.e., creation and receipt, distribution, use, maintenance, and disposition). It is intended to inform CGDI stakeholders on the importance of long term data preservation, and provide them with the information and tools required to make policy decisions for creating an archives and preserving digital geospatial data.

This primer also discusses legal topics, current at the time of publication, for general informational purposes only. It builds on the GeoConnections *Research and Analysis Report: Geospatial Data Archiving and Preservation* (HAL, 2011). Material found in this primer may not apply to all jurisdictions. GeoConnections is not responsible for the use of any materials or contents of the primer. The contents of this primer do not constitute legal advice and should not be relied upon as such.

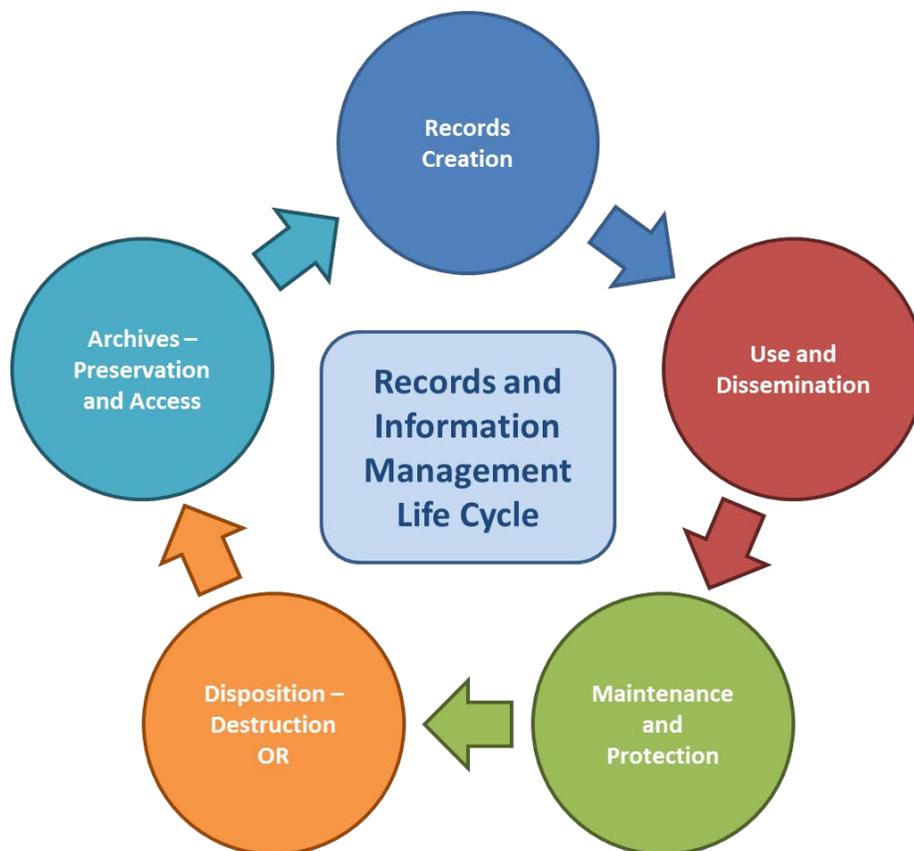
## 2. Introduction

The purpose of this chapter is to introduce the reader to basic digital data preservation and archiving concepts, processes and terminology. It is important to recognize that some terms may not have the same meanings in different communities of practice, and some of these differences are referenced. In addition, specific guidance for geospatial data creators and preservers is briefly discussed.

### 2.1 Data Archives and Preservation

Data (or records, the term most often used by the archiving community) preservation is a normal part of the information management life cycle, as illustrated in Figure 1. For this reason, it is important for data creators to think about the possibility of its preservation for long-term access and use at the beginning of this cycle.

Figure 1: Records and Information Management Life Cycle



Source: <http://slcoarchives.wordpress.com/2012/04/13/managing-records-now-for-the-future/>

Preservation of data may be for the short-, medium- or [long-term](#) and this is determined during the organization's data [appraisal](#) process, which may be documented in mandate statements, acquisition policies and agreements as determined by creators, preservers and users. While there may continue to be a requirement to preserve hard copies of data in the form of maps or charts, such preservation activities typically begin long after creation and rarely involve the data creators. With digital resources, however, there is a need to actively manage the resource at each stage of its life cycle, to recognize the inter-dependencies between each stage and to commence preservation activities as early as practicable (RLG and OCLC, 2002) (Lauriault, Craig, Pulsifer, & Taylor, 2008). Because much of the supporting information necessary to preserve archived information is more easily available or only available at the time when the original information is produced, results are best when these organizations participate in the preservation effort. This primer focuses on the considerations that organizations need to take into account in the design and development of [digital](#) data archiving and preservation policies and processes.

### 2.1.1 Basic Terminology

At the outset, it is important to recognize that archiving and preservation terms are often used interchangeably, for example: **archives**, **preservation**, **back-up system** and **storage**. For the sake of consistency, this primer will refer to the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2 Project *Terminology Database* (InterPARES 2, 2013) and the terminology used in the reviewed preservation frameworks (see Chapter 4). A glossary of key terms is provided in Appendix B.

Firstly, the “agency or institution responsible for the preservation and communication of records [e.g., data and metadata] selected for permanent preservation” is known as an **archives** (InterPARES 2, 2013). **Preservation** on the other hand is the series of managed activities necessary to ensure continued [access](#) to digital materials for as long as necessary, beyond the limits of media failure or technological change (Digital Preservation Coalition, 2008). More specifically, preservation is the “whole of the principles, policies, rules and strategies aimed at prolonging the existence of an object [dataset, database, software] by maintaining it in a condition suitable for use, either in its original format or in a more persistent format, while leaving intact the object's intellectual [form](#)” (InterPARES 2, 2013). A **back-up system** is the technology used to make a copy of a data file for the purpose of system recovery, while **storage** is the placement of that data into a storage system on a digital medium (e.g., storage tape). Information technology (IT) administrators sometimes consider their back-up and storage systems to be an archives, even though these are not permanent and tapes are often overwritten. A geospatial data archives will include a back-up system and data storage as part of its operational infrastructure and these will form part of its [records preservation system](#).

The archival community refers to the creation of **archival records**, which are “documents [data and metadata] made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference” (InterPARES 2, 2013). In some instances, organizations create and set aside records which remain an active part of their ongoing business processes. For example, earth observation (EO) raw data are collected from satellite

receiving stations and ingested into a preservation system as a normal part of the record life cycle process. These data remain active and are preserved because legislation often compels EO data creators (e.g., Remote Sensing Space Systems Act (S.C. 2005, c. 45)) to do so, there is a business case to preserve them and it is expected that these data will fulfill future requests.

The primary actors in the data archiving process, according to the *Open Archival Information System (OAIS) Reference Model* (CCSDS, 2012), discussed in Chapter 4, are [producers](#), [consumers](#) and [management](#). Producers (i.e., record creators) may be the individual creators of the geospatial dataset or the legal entity responsible for its creation. Producers may have their own internal preservation system or be external organizations contributing data to the archives by mandate or voluntarily. Management oversees the process, but is not involved in the day-to-day operations of the archives, which are normally carried out by an administrative functional entity. Consumers are archives users, stakeholders or a [designated community](#). A designated community may be the Canadian public, or a specific and distinct group such as geomatics professionals, earth scientists, or oceanographers each having specific needs, requiring different functionality and support. This primer will focus on geomatics data creators and professionals in general but will introduce, in the case studies and profiles below, some distinct designated communities.

## 2.1.2 Trusted Digital Repositories

A digital geospatial data archives will need to demonstrate that it has a [trusted preservation system](#) to be considered trustworthy by its designated communities (InterPARES 2, 2002) (MacNeil, 2000). An archives is considered trustworthy when it can demonstrate that the preserved digital data in its collection, will be [accurate](#), [reliable](#) and [authentic](#) and that the business unit doing the preservation can demonstrate that it “has no reason to alter the preserved records or allow others to alter them and is capable of implementing all of the requirements for the preservation of authentic copies” (InterPARES 2, 2013). While these three concepts are known to scientists and geomatics practitioners, they are understood differently by archivists (Roeder, Eppard, Underwood, & Lauriault, 2008). It is important to recognize that, for archivists these terms refer to the record’s attributes and not the scientific or methodological aspects that affect the quality of the data that form the record. A **trusted digital repository** (TDR) is a **digital archives** “whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future” (RLG and OCLC, 2002, p. 5).

A multidisciplinary and multisectoral group of international archival audit and certification experts and preservers developed a certification process. The *Trustworthy Repositories Audit and Certification: Criteria Checklist* (OCLC and CRL, 2007) was created to assist the preservation community, especially those responsible for implementing digital archives. The *TRAC Checklist* is a necessary tool for managers and administrators to assess their current practices, to identify gaps and develop solutions. In 2012 the *TRAC Checklist* became the *ISO<sup>1</sup> 16363:2012 Space Data and Information Transfer Systems: Audit and certification of trustworthy digital repositories* recommended practice (ISO, 2012). While the *Checklist* is designed for the purpose of auditing preservation systems, is long, and is generalized to all digital archives, it is an indispensable tool for geospatial data producing organizations who wish to develop a geospatial data archives or who wish to self-assess their systems. It will be revisited in Section 4.3 below.

#### GOOD PRACTICE

*The Trusted Digital Repositories: Attributes and Responsibilities Report (2002)*, produced by the Research Library Group (RLG) and the Online Computer Library Centre (OCLC) in collaboration with archivists, ascribes the following attributes to a TDR:

- accepts responsibility for the long-term maintenance of digital resources on behalf of its depositors (i.e., creators or producers) and for the benefit of current and future users, consumers or designated communities;
- has an organizational system that supports not only long-term viability of the preservation system, but also the digital information for which it has responsibility;
- demonstrates fiscal responsibility and sustainability;
- designs its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it;
- establishes methodologies for system evaluation that meet community expectations of trustworthiness;
- is depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly; and
- has policies, practices, and performance that can be audited and measured.

### 2.1.3 The Importance of Metadata

In the *Geospatial Data Archiving and Preservation* report (HAL, 2011), geospatial data portals were recognized as being more than discovery tools. Portals are also access points to collections of geospatial data that have been [appraised](#) as having business or scientific value by the portal's host institution, data contributors and their user communities. In addition, these data are discoverable via extensive geospatial [metadata](#) and in some cases geospatial data portals have policies in place to manage these data for the long-term. Those portals that adhere to open specifications, interoperability standards and the *ISO 19115 Geographic Information – Metadata* standard and adopt open source software stand a greater chance of withstanding the test of time as compared with those that use proprietary systems (Roeder J. , Eppard, Underwood, & Lauriault, 2008, pp. 44, 45). Finally, as expected, it was discovered that geospatial data are **complex**, **dynamic** and **interactive**, and are in multiple formats, held in specialized systems, and accessed and disseminated in distributed systems according to discipline specific practices.

<sup>1</sup> International Organization for Standardization

However, since geospatial data are collected according to well-defined and normalized scientific models and prescribed methodologies, they should be well described by creator metadata.

### 2.1.4 The Impacts of Technological Change

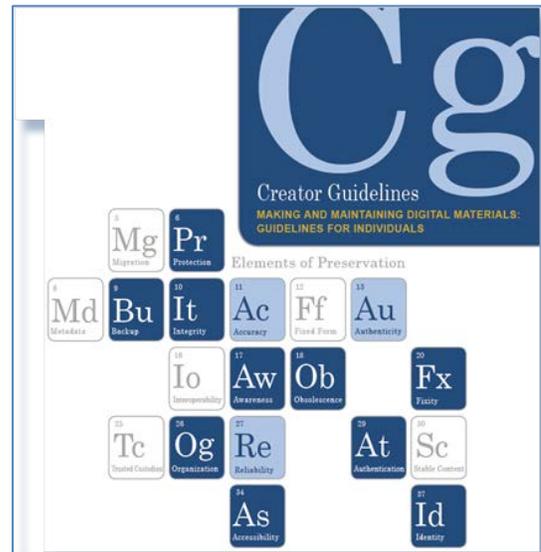
**Technological change** is a particularly important consideration in data preservation, especially if data are to be preserved and remain accessible for the long term. No matter how well an archives maintains its current holdings, it will eventually need to migrate much of its content to different media (which may or may not involve changing the bit sequences) and, or to a different hardware or software environment to keep them accessible. Overcoming technological obsolescence of hardware and software may be accomplished by various means, including [emulation](#) and [migration](#) techniques. Information [reformatting](#) and [refreshing](#) may also be required to move data between data storage media. Denise Bleakly's paper, *Long-Term Spatial Data Preservation and Archiving: What are the Issues?*, provides an overview of these technological issues (Bleakly, 2002). The frameworks examined in Chapter 4 address technological change for geospatial data creators and preservers, and provide guidance on how to plan accordingly.

In addition, geospatial data formats are complex, frequently change and become obsolete and, as a result, format registries have been created to enable the long-term preservation and access to data. Registries enhance format migration efforts, such as keeping emulation environments portable (Erwin & Sweetking-Singer, 2009). The Library of Congress hosts the [Sustainability of Digital Formats](#) registry, which includes geospatial data (Library of Congress, 2013). This registry is a useful tool for geospatial data creators and preservers, and creators should register their file formats into this registry, while file format descriptions should form part of archival metadata description (Hoebelheinrich & Munn, *Assessing the Utility of Current Format Registry Efforts for Geospatial Formats*, 2009).

Even though geospatial data are unique, there remain some general guidelines that apply to all record creators and preservers (InterPARES 2, 2007), which are discussed in the following sections.

## 2.2 Geospatial Data Creator Guidelines

The InterPARES 2 Project *Creator Guidelines* (InterPARES 2, 2007) provide recommendations that will be familiar to geomatics professionals. Software choices should be backward compatible (i.e., it can work with [input](#) generated by an older product) and interoperable across time and space, be broadly adopted and adhere to standards. This is especially true if creators wish to have their data read in the future and want the information created from their data to be understood and visualized as they intended. To inform migration strategies, all specialized software should be fully documented including software customizations. Notes in the software code are particularly helpful in this regard. The construction of the geospatial system, including structure and functions, hardware and software, operating system and how all these operate with each other, should be documented in basic specifications, as these will inform upgrades. Finally, widely accepted, non-proprietary, platform-independent and uncompressed formats, with access to specifications and documented versions and encoding, should be used. The *Creator Guidelines* also recommend the logical grouping of records, and the identification of retention strategies for groupings of records at the point of creation, as this is more expedient.



Because geospatial data are often rendered into interactive forms, accessed via web services or data portals and may be dynamically generated with near real time sensors, it is important to understand the concepts of [fixed form](#) and [bounded variability](#) (InterPARES 2, 2007). In order to ensure that a record's appearance is the same each time it is retrieved, the content of the record (e.g., the data to create the map and the algorithm used to render it) needs to be fixed. Furthermore, the documentary form of the record (e.g., specifications and the software used to create and view the interactive map) should be immutable to ensure that its presentation remains the same across time. Bounded variability means that fixed rules need to be established for the selection of content and form, which allows for a stable range of variability in the interactive map or model. This ensures that a reliable and authentic record of the interactive map is accessible in the way the creators intended. The Treasury Board of Canada (TBS) endorsement of the *ISO 19128:2005 Geographic information Web map server interface* (ISO, 2005) for all Government of Canada (GoC) geographic information (TBS, 2012) promotes a common open standards approach to the production of maps which will simplify the work of preservers.

**GOOD PRACTICE**

A more detailed analysis of metadata elements for the long-term preservation of geospatial data sets is available in the *Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)* publications, created under the auspices of the US Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP) (GeoMAPP, 2011).

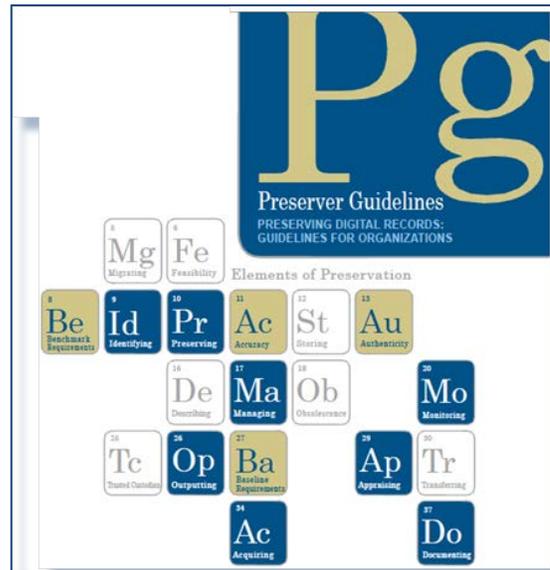
Geomatics professionals are well versed in **metadata**, or **creation documentation** as it is referred to in the *OAIS Reference Model*. For archivists, metadata help maintain the record's **identity** which is the quality that one record can be identified and distinguished from another while **integrity** is the quality of it being unaltered and complete (InterPARES 2, 2013). **Archival identity** metadata elements are very similar to most required geomatics metadata elements but some additional elements may be required, such as file format descriptions, component inter-linkages and preservation environment, to name a few.

The *Creator Guidelines* also refer to **authentication**, which is a declaration at a point in time of a record's [authenticity](#). This is done by inserting an element or adding a statement to the record by an authoritative person who has the right to do so. [Technology-independent authentication](#) is one method, and many of the recommendations in the *Guidelines* help with determining if a data set or a grouping of data can be presumed authentic, which is drawn from the known facts about how these were created. In addition, administrative policies and practices which are technologically independent or neutral, such as protecting data from tampering by controlling access to where the data are stored by way of restricting physical access, are recommended. Technology-dependent authentication can also be accomplished with transmission technologies (e.g., cryptography and digital signatures), but these are subject to technological obsolescence. Restricting access to the data with passwords and other protective and security measures and the development of access permission protocols are other approaches. Whatever methods are employed, it is important to be able to demonstrate that records cannot be tampered with.

Finally, the *Creator Guidelines* discuss the need for security measures, regular back-up of operational data and protection against hardware and software obsolescence. In addition, it is recommended that data creators develop a preservation strategy, consider long-term preservation issues and identify a [trusted custodian](#).

## 2.3 Geospatial Data Preserver Guidelines

*Preserver Guidelines* (InterPARES 2, 2007) were designed to provide concrete advice to those responsible for the long-term preservation of digital records and are based on a **chain of preservation** framework, which is “a system of controls that extends over the entire life cycle of records in order to ensure their identity and integrity over time” (InterPARES 2, 2013). This framework includes the policies, strategies, and methodologies needed to manage digital records. The scope and the objectives of the geospatial preservation program should include deciding which geospatial data are to be preserved, how these are to be made accessible, to which designated community and to meet what specific needs and technical requirements. The *Preserver Guidelines* also point to a *Policy Framework* document (InterPARES 2, 2008) which includes policy principles for creators and preservers, and it recommends using the *OAIS Reference Model* (CCSDS, 2012) and the *TRAC Checklist* (OCLC and CRL, 2007) for developing the functional aspects of the archives, which will be discussed in Chapter 4.



### GOOD PRACTICE

A good example of a guideline for developing a business plan for a geospatial data archiving initiative is the *Geoarchiving Business Planning Guidebook* (GeoMAPP, 2011a). The *Guidebook* provides a detailed approach to describing how data archiving objectives will be achieved, as well as how to describe the necessary justification for the initiative. A summary of the *Guidebook's* contents is provided in Appendix F.

A geospatial data archives requires technological, human and financial resources and it is imperative that these be secured and sustainable through the development of a solid business plan. This includes a clear communication strategy to convince potential funders, and the ability to leverage new resources once the preservation program is in place. Funders may be more receptive to an incremental resource acquisition strategy and collaborations for pooling resources. The Geospatial Multistate Archive and Preservation Partnership (GeoMAPP) *Geoarchiving Business Cost-Benefit Analysis Guidance Document* can also help with justifying archiving and preservation

initiatives to funders (GeoMAPP, 2012).

Once the geospatial archives is established, it should have a record creator advisory function to ensure that preservation is part of the record creation process and it should adhere to the InterPARES 2 *Requirements for Assessing and Maintaining the Authenticity of Electronic Records* (InterPARES 2, 2002), which provides benchmarks for a good preservation environment. The *Preserver Guidelines* recommend that preservers establish controls over “records transfer, maintenance, and reproduction, including the procedures and system(s) used to: transfer records to their own organization or program within the organization; maintain them; and reproduce them in a way” that satisfies the InterPARES 2 *Requirements* and guarantees a record’s **identity** and **integrity**. In addition, the implementation of clear maintenance strategies is key to a preservation strategy (see text box).

### GOOD PRACTICE

The InterPARES 2 *Preserver Guidelines* recommend the following maintenance strategies for preserved data (InterPARES 2, 2007):

- Clear allocation of responsibilities;
- Provision of appropriate technical infrastructure;
- Implementation of a plan for system maintenance, support and replacement;
- Implementation of a plan for the transfer of records to new storage media on a regular basis;
- Adherence to appropriate storage and handling conditions for storage media;
- Redundancy and regular backup of the digital entities;
- Establishment of system security; and
- Disaster planning.

**Appraising** records is part of all retention plans and the *Preserver Guidelines* recommend that preservers work with records creators on appraisal, along with transfer methods, early in the geospatial data creation process. Furthermore, preservers may occasionally find it useful to participate in the design of record creation and maintenance systems in order to build in preservation wherever possible. Appraisal also includes identifying the owners of the geospatial data in order to assess preservation ramifications, which is complex when dealing with distributed data or data accessed via a geospatial data portal. Assessing the **authenticity** of the geospatial data is essential and this can be captured in an appraisal report based on the [benchmark authenticity requirements](#). The geospatial data that have been identified for retention should also be monitored to ensure that: these are not deleted by accident, software upgrades do not change

### GOOD PRACTICE

The US [National Geospatial Digital Archive](#) (NGDA) project shares its collection development policies, data provider agreements and a Procedure Manual, which describe transfer plans, practices and procedures (NGDA, 2009). See <http://www.ngda.org/policies.html>

their attributes, organizational change does not affect earlier retention decisions, and record management practices are adhered to in the creator’s environment. Furthermore, [digital components](#) need to be identified and implicit relationships need to be made explicit in the metadata and components (e.g., formats, file containers, ESRI Shapefiles) before transfer, to ensure that once the data are extracted from the system that created them they can be re-created in a manner the creator intended. Appraisal also includes determining the feasibility of preservation by carefully investigating and then assessing technical preservation requirements and their associated costs over the selected time frame, be it short-, medium- or long-term.

Once geospatial data have been appraised they need to be acquired for preservation and this requires moving the data from the creator's custody to that of the preserver, whether that role exists inside or outside the creator's organization. Requirements include: a transfer plan where both parties agree to a physical and a [logical format](#), an examination of the systems within which the records exist, and an agreed process to ensure their safe transfer into a new system. This will also involve monitoring and enforcing transfer procedures and testing the process. Wherever possible, it is recommended that the oldest available logical format be kept. Also, avoiding ingesting duplicates by ensuring that records are transferred only once is important, as is documenting all processes such as virus checks, validating [checksums](#) and confirming the identity of the record.

The *Preserver Guidelines* suggest that, once the geospatial data are [accessioned](#) into a [trusted preservation system](#), rules and procedures should be established for the ongoing production of authentic copies, as preservation systems become obsolete and technologies need to be upgraded. The archival description of the geospatial data is helpful in this case and this includes how data were collected, appraised and processed, along with access, and intellectual property and privacy rights. Understanding the [digital rights management](#) aspects of the geospatial data record is also important, as there may be legal ramifications in circumventing proprietary environments to extract data for preservation. It is also important to test the selected preservation strategy to ensure that it is effective and to maintain proper storage.

Making the preserved geospatial data [accessible](#) is a final consideration for data preservers. The *Preserver Guidelines* recommend that preservers provide documentation about the data reproduction, transfer and monitoring processes, which enable users to assess the authenticity of the record and to decide if it is coming from a trusted source. Also, the preserver should provide the technological means for users to access the geospatial data (e.g., raw data via a portal, within a new mapping environment, or any other visualization tool). Access technology choices and methods should be based on the skills and requirements of the archives' designated communities.

# 3. Legislation and Policy Affecting Archiving and Preservation

The purpose of this chapter is to encourage the reader to consider the legislative mandates, including regulation, policies and directives that should compel them to carefully examine their preservation practices. While the discussion in this chapter is limited to the federal legislative framework, similar obligations within provincial legislation, regulation, policies and directives may also need to be taken into consideration.

## 3.1 Archival and Preservation Responsibilities

Governments are responsible for the management and preservation of the data and information they produce. The Government of Ontario, for example, must adhere to its *Archives and Recordkeeping Act*, and the Archives of Ontario provides a series of guidelines, retention schedules, and fact sheets to assist government offices with their record keeping (Ontario Ministry of Government Services, 2011). One of the fact sheets matches [Retention Periods](#), [Archival Access and the FOI Act](#). The *Northwest Territories (NWT) Archives Act* mandates how Territories records are kept and a NWT [Archives policy](#) guides the process (Government of NWT, 1993). The NWT provides a list of agencies and departments which have specific Records Retention and Disposition Schedules which also dictate record acquisition decisions (Prince of Wales Northern Heritage College, 2013).



Government of Canada (GoC) organizations are also impacted by overarching legislation and regulation that may explicitly dictate the management, preservation or deposit of some of the records they produce. Primary among these is the *Library and Archives of Canada Act (2004)* which forbids the destruction of records by all GoC institutions without first obtaining written permission. Furthermore, some legislation and regulation explicitly reference geospatial data and their associated software which may warrant record management or preservation actions. Finally, a number of directives and policies make recordkeeping and preservation an obligation. The primary responsibility for information management policy in the GoC lies with the Treasury Board of Canada Secretariat (TBS). In other cases, data are preserved because they: may be deemed critically important to the day-to-day business of a GoC organization (e.g., Canada Centre for Remote Sensing images discussed in Section 5.1); may be used to inform government decisions (e.g., the Land Information Ontario process of determining which data are official as discussed in Section 5.2); may have significant scientific merit (e.g., Department of Fisheries

data discussed in section 5.3), or may be identified by stakeholders as important, as is the case with the International Polar Year data discussed in Section 5.4.

The 2011 GeoConnections Report *Geospatial Data Archiving and Preservation* noted that a geospatial database, a data set or a map may be a [record](#) (InterPARES 2, 2013), to be set aside for future reference and that preservation decisions would be contingent upon the record creating organization's legislative, regulatory, policy and information management requirements. A record creating institution such as the Department of Fisheries and Oceans (DFO), for instance, may select data sets for preservation based on how these are acted upon, or set aside for action. For example, fish count data used to establish fish quotas over time may be preserved because they have scientific merit or because they informed a key GoC policy decision. Alternatively, the *Charts and Nautical Publications Regulations* (JC, 2007), pursuant to the *Arctic Waters Pollution Prevention Act* (JC, 2010), refer to the electronic chart display and information system (ECDIS) and its associated electronic navigational chart database (ENC), and in this case the software and the data may be considered worthy of preservation.

In addition, when considering the disposition of data, questions such as “Does the information in the database protect the rights of citizens? or the interests of the GoC?” should also be considered (LAC, 2009). Guidelines to help creators decide what information resources are of business value are outlined in the Library and Archives of Canada (LAC) Recordkeeping (RK) Toolkit (LAC, 2012). Decisions on when and how often to capture snapshots of these records and when to [accession](#) these into the archives must be made between record creators, their stakeholders and preservers and these decisions form part of an organization's preservation policies. Also, as explained in Section 2.2, 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. The DFO data and software just mentioned could potentially remain [active](#) throughout their life cycle and also be preserved. There are many life-cycle models to choose from and the *Review of Data Management Lifecycles Models* produced by the University of Bath in the UK explains eight models in tangible and concise fashion (Ball, 2012).

The legislation, regulation, directives and policy briefly discussed in the following sections should be taken into account when forming an archives' mission, mandate and objectives. Provinces and territories may have similar legislation and regulation, and cities and municipalities may need to follow those and any local resolutions, policies and directives.

## 3.2 Federal Acts and Regulations

Government of Canada geospatial data producing organizations are mandated to adhere to the obligations of several overarching Acts and Regulations. The obligations, limitations and challenges related to the following legislation are summarized in Appendix C and are more fully discussed in the *Research and Analysis Report: Geospatial Data Archiving and Preservation* (HAL, 2011):

- Library and Archives of Canada Act (LAC Act) – This act stipulates that: “the documentary heritage of Canada be preserved for the benefit of present and future generations”; 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 that it serve as the memory of the GoC. 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 documentary heritage known to Canadians and to anyone with an interest in Canada and facilitating access to it” (JC, 2004).
- Copyright Act – This act stipulates that: record creators can assert control over their works; works containing geospatial data, created by the GoC employees, belong to Her Majesty; and these works can be licensed (JC, 1985). It is important to note that Canada copyright does not subsist in data itself, but may subsist in the *original* selection or arrangement of data. In order for a compilation of data to be copyright protected, there must be an author who has created something "original". The Supreme Court of Canada in its 2004 *CCH* decision stated that *originality* requires that a work “must be the product of an author's exercise of skill and judgement” that “must not be so trivial that it could be characterized as a purely mechanical exercise” ([2004] I S.C.R. 339).
- As of today, there are no universal licences for geospatial or any other data being used within the GoC, but there are guidelines such as the *Dissemination of Government Geographic Data in Canada: Guide to Best Practices* (GeoConnections, 2008). The anticipated Open Government Directive is expected to produce an “open data” licence, referred to as the Open Government Licence, that will offer a common licence for GoC information and data, including geospatial data, in the near future. Other jurisdictions that also have Crown Copyright are adopting international interoperable licenses such as Creative Commons and Open Data Commons licenses (Scassa, 2011).
- Access to Information Act – The purpose of this act is “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” (JC, 1985).
- Privacy Act and Privacy Regulations – The spirit of this act is to “protect the privacy of individuals with respect to personal information about themselves held by a government

institution and provide individuals with a right of access to that information” (JC, 1985). The *Privacy Act* also stipulates how data should be maintained, and the quality of the data along with disposition rules. In addition, personal information is only to be used for the purposes for which it was initially collected and strict rules apply on how personal information can and cannot be disclosed.

- Personal Information Protection and Electronic Documents Act – The purpose of this act, which applies to the private sector in Canada, 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” (JC, 2000).
- Canada Evidence Act – This act provides clear statements on the provision of documentary evidence and, among other things, how to assess the authenticity, integrity and the certification of the digital records being called into evidence (JC, 1985).

Legislation that references geospatial records were reviewed in detail in the *Geospatial Data Archiving and Preservation* report (HAL, 2011) and a brief summary of recommendations can be found in Appendix E of that report. In most cases the recommendation is that data, maps and software be created and managed according to the LAC Act provisions, keeping in mind overarching Acts such as the Copyright Act and the TBS policies and directives. In some cases specific records need to be preserved in the event they get called into evidence (e.g. the Canada-Newfoundland Atlantic Accord Implementation Act).

### 3.3 Federal Policies and Directives

---

The Treasury Board of Canada Secretariat (TBS) has produced policies and directives for GC record creators in general and for GoC geospatial data producers specifically. These are listed below while Appendix D provides a summary of objectives and obligations for geospatial data creators. TBS policies and directives affect how information resources are managed during the course of a record's life cycle and if adhered to, will facilitate the geospatial data preservation process. The TBS Geospatial Data Standard refers to description and dissemination; GoC geospatial data producers must adopt *ISO19115:2005 Geographic Information – Metadata and ISO 19128 Geographic information – Web map server interface*. The other policies and directives are general good practices for all types of records including data, metadata and technologies such as software, including:

- Policy Framework for Information and Technology
- Policy on Information Management (IM Policy)
- Policy on Management of Information Technology
- Directive on Information Management Roles and Responsibilities
- Directive on Recordkeeping
- Standard for Electronic Documents and Records Management (EDRM) Solutions
- Standard on Metadata
- Standard on Geospatial Data

Additionally, LAC, in keeping with its mandate and in accordance with the TBS Directive on Recordkeeping, has produced a number of information products, tools, guidelines and methodologies to assist record creators, which are available on its website (LAC, 2011).

# 4. Archiving and Preservation Frameworks

## 4.1 Introduction

---

The purpose of this chapter is to introduce and briefly profile the following two important frameworks and one checklist that provide guidance to geospatial data creators and preservers for their preservation planning processes:

- the *Reference Model for an Open Archival Information System (OAIS)* (CCSDS, 2012), developed by the [Consultative Committee for Space Data Systems \(CCSDS\)](#);
- the European *Long Term Preservation of Earth Observation Space Data: European LTDP Common Guidelines* (LTDP Working Group, 2012), developed by the [Long Term Data Preservation \(LTDP\) Working Group of the Ground Segment Coordination Body \(GSCB\)](#); and
- the *Trustworthy Repositories Audit & Certification (TRAC) Audit and Certification: Criteria and Checklist* (OCLC and CRL, 2007), developed by the [Center for Research Libraries](#) and the [Online Computer Library Center, Inc.](#)

The *OAIS Reference Model* and the *LTDP Framework* are used to guide the functional development of a preservation system and, along with the *TRAC Checklist*, refer to the same preservation activities, at differing levels of detail and in a different order due to their adopted organizing principles and their focus. The *OAIS Reference Model* discusses the functional components of a preservation system and does not consider the creator's environment. The *LTDP Framework* is an OAIS preservation system tailored to EO mission data environments. The *TRAC Checklist* was designed as an OAIS model compliance certification and audit tool, and is also used to self-assess and guide preservation planning. Since terminology varies between these three key documents, wherever possible the reader will be referred back to InterPARES 2 archival terminology and concepts.

## 4.2 Reference Model for an Open Archival Information System

Recognizing that the rapid obsolescence of digital technologies was creating an increasing risk of being unable to restore, render or interpret preserved information, the CCSDS developed the *OAIS Reference Model* to provide a common framework from which to view archival challenges. There is a particular focus on digital information, both as the primary forms of information held and as supporting information for both digitally and physically preserved materials. While the model accommodates information that is inherently non-digital (e.g., a physical map), the modeling and preservation of such information is not addressed in detail. The model applies to archives that need to accommodate



steady input streams of information (e.g., sensor webs) as well as those that experience primarily irregular inputs. The model has been adopted by the International Organization for Standardization (ISO) as the standard [ISO 14721:2012 Space data and information transfer systems -- Open archival information system \(OAIS\) -- Reference model](#).

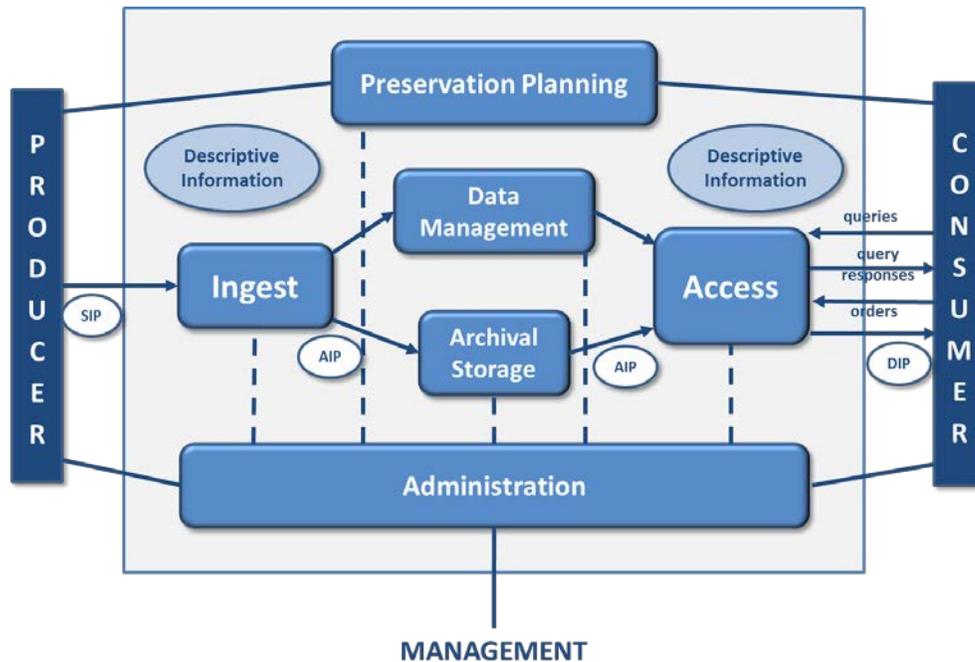
An Open Archival Information System (OAIS) is defined as an archives “consisting of an organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a designated community” (CCSDS, 2012, pp. 1-13). An ‘archives’ in the OAIS context is analogous to a [records preservation system](#) as defined by InterPARES 2. The *OAIS Reference Model* accommodates the highly distributed nature of digital information holdings and the need for local implementations of effective policies and procedures supporting information preservation. This allows for a wide variety of organizational arrangements, including traditional archives and distributed digital archives, which is particularly suitable for collaborative geospatial data producing environments. The information flows in the *OAIS Reference Model* are between three primary roles that interact with the preservation system:

- *Producer* – contributes the information (files, digital data objects) to be preserved (i.e., the data creator).
- *Management* – sets the archives’ overall policies and manages them within the preservation system environment (i.e., the preserver).
- *Consumer* – interacts with preservation system services to search, access, and acquire preserved information of interest (i.e., the user or the designated community).

The model encompasses the following six functional entities as illustrated in Figure 2:

- *Ingest*: The interface between the creator and the preserver, this entity receives the data from the producer as a [submission information package \(SIP\)](#), which may be data, a database, their associated software and metadata, and any other descriptive information related to the record that might be recorded in an appraisal report. During ingest, an [archival information packages \(AIP\)](#) is created, containing the converted SIP and archival description or [preservation description information \(PDI\)](#).
- *Archival Storage*: This entity receives AIPs from Ingest for permanent storage and retrieval and handles the information storage processes for the preservation system.
- *Data Management*: This entity manages the preservation system's records and any administrative data about managing the preservation system.
- *Administration*: This entity contains the services and functions to control the entire operations of the other entities in the preservation system. This includes negotiating submission agreements with producers, auditing submissions, managing hardware and software configurations, and monitoring operations.
- *Preservation Planning*: This entity ensures that the integrity and the identity of the records is maintained through transformations such as migration and updates. It ensures that the records are accessible and understandable across time and space – that they are authentic. It also monitors the consumer's technology requirements. This is where migration plans, software prototypes and test plans are developed for migration purposes.
- *Access*: This entity enables access to the AIPs by retrieving them from the system, receives and responds to requests, and creates [dissemination information packages \(DIPs\)](#) and delivers them to consumers.

While **Management** does not play an operational role in the preservation system, it provides the system with its charter and scope and can play a very important supportive role. For example, management can implement policies that require all funded activities within its sphere of influence to submit data products to the preservation system and also adhere to standards and procedures. Management also provides funding resources, evaluates the system's performance and may participate in conflict resolution involving Producers, Consumers and internal Administration if necessary.

Figure 2: OAIS Functional Model<sup>2</sup>

Source: CCSDS (2012. 4-1)

### 4.3 European Long Term Data Preservation Common Guidelines

In 2006, the European Space Agency (ESA) initiated coordination action to develop a common approach to the preservation of EO space data. The goal was to ensure the long-term preservation of all European (including Canadian) EO space data and to facilitate their accessibility and usability through the implementation of a cooperative and harmonized collective approach among EO space data owners. A consolidated European *LTDP Guidelines* document has been produced that addresses eight main LTDP “themes” defining for each the



“guiding principle” and a set of “key guidelines” that should be applied to guarantee the preservation of EO space data in the long term.

<sup>2</sup> The lines connecting entities in the model identify communication paths over which information flows in both directions. The lines to Administration are dashed only to reduce diagram clutter.

The *LTDP Guidelines* is fully compatible with TRAC (see following section). The TRAC metrics are an audit and certification method that tests the trustworthiness of all types of digital repositories. The LTDP guidelines are, on the other hand, a set of practical recommendations specifically addressing Earth Observation data archives and covering the TRAC metrics in most cases. The *LTDP Guidelines* are an example of how a specific geospatial data producing community (i.e., creators) adapted the *OAIS Reference Model* to their context.<sup>3</sup> In the EO context, preservation is often a key operational component implemented as part of a record's life cycle management since preservation is mandatory for most EO data creators. The creator in this instance is most often also the preserver. The EODMS case study is an example of this (see Section 5.1), where CCRS is both the creator and preserver of the data. In these guidelines, like in the OAIS, the term “archives” is used instead of “preservation system”. The eight themes are:

1. Preserved Data Content Definition and Appraisal
2. Archive Operations and Organization
3. Archive Security
4. Data Ingestion
5. Archive Maintenance
6. Data Access and Interoperability
7. Data Exploitation and Reprocessing
8. Data Purge Prevention

Each of these themes is briefly described below.

The **Preserved Data Content Definition and Appraisal** defines a consistent and complete set of data to enable its current and possible future utilization. This includes the data, processing software, mission documentation, description and archival metadata and within these, elements related to EO data specific Quality Indicators (QA4EO, 2013). This also includes supporting documentation regarding intellectual property, access restrictions, specifications and standards, which is documentation that would form part of an appraisal report. This is analogous to an SIP and AIP in the *OAIS Reference Model*, and archival records as defined by InterPARES 2.

**Archive Operations** consist of all daily technological and administrative activities that are carried out to run and monitor the preservation system (e.g., execution and control of the applications, system monitoring, anomaly reporting, error recovery, and activity reporting and statistics). One of the guidelines here is the adoption of the OAIS reference model. The preservation system is situated within an **organization** structured to be the archive, with mandates, policies, laws, and sustainable resources to meet the goals and perform the tasks and processes of long-term preservation.

---

<sup>3</sup> The OAIS-ISO 14721 standard was used in the definition of the structure of the LTDP Common Guidelines document.

**Archive Security** encompasses all the activities dedicated to the implementation of security measures for data access and storage in order to guarantee confidentiality, integrity and availability of the archived data. This ensures the data are authentic, and provides users with the information they need to trust that the records in the system are what they purport to be, that the chain of custody is secure, and that data have not been tampered with. This covers technical and non-technical authenticity requirements including physical, information and staff security.

**Data Ingestion** contains the services and functions that, according to the OAIS standard: accept [Submission Information Packages \(SIPs\)](#) from Producers (e.g., creator data and metadata); prepare [Archival Information Packages \(AIPs\)](#) for storage (i.e., reformatted files, archival description, etc.); and ensure that the AIPs and their supporting [Descriptive Information](#) (i.e. archival metadata) are stored within the preservation system.

**Archive Maintenance** consists of all the activities aimed at guaranteeing the integrity of the archived data. Data integrity assures that the archived data are complete and unaltered through loss, tampering or data corruption. Archive maintenance is based on the storage of equipment and storage media in secured and environmentally controlled rooms and a set of defined activities to be performed on a routine basis (e.g., migration to new systems and media in accordance with the technology and consumer market evolution, data compacting and data format/packaging conversion).

**Data Access** corresponds to the services and functions which make the archival information holdings and related services visible to consumers. **Interoperability** is related to the possibility of accessing data in a common and standardized way despite the intrinsic differences between the data sets and the systems being used to access them. This includes retrieval and delivery of the data in the form of DIPs, and since EO data are heterogeneous, interoperability and the harmonization of data access are very important.

The **Data Exploitation and Reprocessing** theme contains all the exploitation activities related to data processing and reprocessing, regeneration or enhancement of the catalogues (e.g., with data mining), integration of new services (e.g., through service work-flow orchestration) and quality assessment of the products and services. This is to guarantee the reusability of these data over time.

#### GOOD PRACTICE

*The LTDP EO Preserved Data Set Content LTDP/PDSC document (LTDP, 2012) is a preservation checklist used by EO data creators and producers to guide mission specific preservation decisions. It is an excellent example for geospatial data producers who are preparing appraisal reports and data inventories. The document also includes an annex which maps PDSC content to the OAIS Reference Model.*

The **Data Purge Prevention** theme defines a set of procedures to be applied with the objective of preventing or minimizing EO space data loss and ensuring resources are applied towards EO data preservation and access activities through a management approval process. This theme is of particular importance when EO space data holders and archives owners can no longer preserve the data.

#### 4.4 Trustworthy Repositories Audit & Certification: Criteria and Checklist

The *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist* (OCLC and CRL, 2007) provides those responsible for digital repositories with an objective tool to measure the trustworthiness of their preservation system. It has since become the ISO 16363:2012 *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories* standard (ISO, 2012).



The *TRAC Checklist* is divided into three sections: Organizational Infrastructure; Digital Object (i.e., archival record or Preserved Data Content Definition) Management; and Technologies, Technical Infrastructure, and Security. Each of these is accompanied by a description of overall expectations and is divided into a number of elements with accompanying criteria and sub-criteria represented as a series of questions. Questions cover methods to measure adequacy (e.g., based on indicators such as documentary evidence, the degree of transparency, whether or not an archives can meet its stated objectives), and measurability (e.g., based on indicators such as a degree of trustworthiness). In addition, there is an expectation that preservation systems will adhere to a number of ISO quality, security and process standards and the *OAIS Reference Model* standard ISO 14721. The *TRAC Checklist* facilitates the objective evaluation of OAIS compliance. The contents of the three sections are:

- A. **Organizational Infrastructure** – includes the following **elements**: governance, organizational structure, mandate or purpose, scope, roles and responsibilities, policy framework, funding system, financial issues and assets, contracts, licences and liabilities and transparency. These elements are measured against the following **criteria**: governance and organizational viability, structure and staffing, procedural accountability and policy framework, financial sustainability and contracts, licence and liabilities.

- B. **Digital Object Management** – includes organizational and technical elements related to repository functions, processes and procedures needed to ingest, manage and provide access to digital objects (i.e., records, Preserved Data Set Content) for the long term. Requirements are as follows and are based on the preservation system’s functionality according to the OAIS functional entities discussed in Section 4.1:
- **Ingest:** The acquisition of digital objects.
  - **Ingest Creation of AIPs:** The transformation of digital objects (i.e., SIPs) into AIPs.
  - **Preservation planning:** Current, sound and documented preservation strategies with mechanisms to keep them up-to-date.
  - **Archival Storage & Preservation/Maintenance of AIPs:** Minimal conditions for performing long-term preservation of AIPs.
  - **Information Management:** The preservation system’s ability to produce and disseminate accurate and authentic versions of digital objects.
  - **Access Management:** Provision of access to users.
- C. **Technologies, Technical Infrastructure, & Security** – does not dictate which hardware and software to use but instead describes best practices for data management and security, which are:
- **General System Infrastructure**
  - **Appropriate technologies:** Building on system requirements and the needs of designated communities.
  - **Security:** Includes IT systems and physical protection from fire, flood, and the actions of people, technical infrastructure risk management and security risk management.

The *TRAC Checklist* is particularly useful as it prompts evaluators to look for concrete documents, processes and components which are not specifically listed in the *OAIS Reference Model* or the *LTDP Framework*. The *Checklist* considers the preservers’ environment.

The first certified Trustworthy Digital Repository (TDR) in Canada is The Ontario Council of University Libraries (OCUL) Scholars Portal. OCUL’s mandate is to provide “a robust long-term preservation environment for all the materials in its collection, working from concepts set out in the Open Archival Information System (OAIS) reference model and codified in the Trustworthy Repositories Audit & Certification (TRAC) checklist and ISO 16363 standard (*Audit and certification of trustworthy digital repositories*)”. The full audit report, policies and plans describing the workflow and the mandate, and responses to the audit checklist are available from the OCUL website (OCUL, 2011), where its *Preservation Action Plan – Journals* can also be located. Geospatial data are not currently being ingested into the Scholars Portal.

# 5. Geospatial Data Preservation Examples

This chapter highlights four examples of digital data preservation initiatives that are good sources of good practices in the field.

## 5.1 Case Study: Earth Observation Data Management System (EODMS)<sup>4</sup>

### 5.1.1 Introduction

The core mandate of the Canada Centre for Remote Sensing (CCRS), Natural Resources Canada (NRCan) is to provide access to satellite technology to monitor Canada's land and borders. CCRS is also the GoC's centre of expertise for remote sensing and geodesy. In addition, CCRS must adhere to the Remote Sensing Space Systems Act (JC, 2007) and the Remote Sensing Space Systems Regulations (JC, 2007), which explicitly state that raw data will be preserved and that records will be managed. Through partnerships with government stakeholders, strong links to academia and the private sector, including international collaborations, CCRS ensures that satellite data are available to serve the needs of the Canadian Government and Canadians. Since 1971, CCRS has accumulated an archive of approximately 800 terabytes of EO data originating from various satellites and airborne sensors (CCRS, 2013).

CCRS's EO imagery and derived products support the GoC's priorities, including economic development of the North, safety, security, sovereignty, and environmental monitoring. Products are delivered to Government stakeholders and Canadians through the Canadian Earth Observation Catalog (CEOCAT), The National Air Photo Library (NAPL) and the National Earth Observation Data Framework (NEODF). The NEODF is a pilot EO preservation system that makes EO data available in a timely manner to federal government organizations. CCRS ingests, preserves and provides access to remote sensing data acquired from two existing Canadian satellite receiving stations, the Remote Sensing National Master Standing Offer for Commercial Satellite Imagery (NMSO-CSI), and other international and private sector EO data providers.

Organizationally, the CCRS Data Acquisition Division is responsible for: acquisition of EO data in Canada; maintenance of the EO data preservation system; development of advanced ground systems for efficient data reception, preservation and distribution of EO data; development of user-friendly EO processing systems; and provision of informatics and computer services to the CCRS. The Earth Observation Data Service (EODS) data reception, dissemination and archives

---

<sup>4</sup> Much of the content of this case study was derived from an in-depth interview with key members of the CCRS Data Acquisition Division that are involved in the CEODAS Project.

access services are provided through CCRS's two satellite receiving stations at Prince Albert, SK and Gatineau, QC and the soon to be online Inuvik Satellite Station Facility (ISSF). The ground segment infrastructure of EODS is capable of directly receiving data of the North American territory. Global data can also be received from satellites equipped with on-board recorders such as RADARSAT. The ground stations receive EO data from several satellite sensors and maintain archival data dating back to 1972.

Natural Resources Canada developed a revitalization plan to improve Canada's EO satellite capacity and to access EO data, which was supported in the 2012 Federal Budget. Accordingly, CCRS is procuring an Earth Observation Data Management System (EODMS), which is a data management, preservation and access system. CCRS also received resources from the GeoConnections program to develop a Canadian EO Data Access and Services (CEODAS) project, which defines the EO data preservation strategy (policy framework and guidelines) and LTDP operation flows for EODMS operations. The EODMS is part of the implementation of CEODAS and it will become a core function of CCRS and an EO preservation system for the GoC.

### 5.1.2 Operational Model in Use – Implementation to Date

In 2010, the CCRS completed the centralization of the EO data in a new preservation facility in Ottawa. The previous preservation system model was a standard satellite mission and sensor driven archives and users were primarily EO experts. The new model is moving toward a service oriented architecture (SOA). A request for proposal (RFP) was launched in January 2013 to create the Earth Observation Data Management System (EODMS), which will be the preservation system for the CCRS archives. The EODMS will adhere to the OAIS: *ISO 14721:2012 Space data and information transfer systems – Open archival information system (OAIS) – Reference model* and *LTDP Common Guidelines*. Once established, the EODMS will aim for compliance with the recommended practices of *ISO 16363:2012 Space data and information transfer systems – Audit and certification of trustworthy digital repositories*.

The two step implementation of the CEODAS is as follows (CCRS, 2013):

- develop a policy framework and guidelines; and
- implement the framework and guidelines, by introducing the EODMS preservation system and also deliver a series of operational procedures and documentations.

CCRS will be capable of managing increased EO data collections within the Federal Committee on Geomatics and Earth Observation (FCGEO) framework and its current EO data collection activities, which are unique in Canada. Providing improved access to those data will facilitate the creation of EO-derived products and knowledge, as well as the integration of EO data with other scientific data. The EODMS will also be able to accommodate the increasing volume of data generated in the future, such as from upcoming RADARSAT Constellation Mission, Landsat Continuation Mission and Sentinel Missions.

## Proposed EODMS

The EODMS will replace CCRS's existing catalog systems and improve functionality for NRCAN, authorized users and the general public to access imagery data (raw data, products, EO derived products and aerial photography (including LiDAR) from CCRS's national multi-mission EO data holdings.

The EODMS will ingest and manage data sets, and include data preservation and cataloguing, processing, end-user licence management and the packaging of products for dissemination employing various Open Geospatial Consortium (OGC) standards. The system will enable services and bilingual access portals that are managed via complex user profiles. It will be modular, scalable and flexible to support current and future needs and will have fully documented API's so that NRCAN or third-party developers are able to customize, extend, enhance, or add functionality to the preservation system. The EODMS will most likely use commercial off-the-shelf (COTS) and open source software to minimize the cost of maintenance, customization and life cycle replacement of the preservation system and its applications. There may also be some proprietary software used outside the core solution.

The EODMS will migrate or integrate existing platform technologies such as PostgreSQL Database and schemas, Client's Accounts Catalog and FTP servers, FTP Server Configuration and the Booth Street Archive. In addition, the initiative will include a training program for administrators, supported by an Administrator & Operator Training Manual, Operators and Administrator User Training Manuals, COTS Vendor Manuals and EODMS (Solution) – Help and FAQ documentation.

The EODMS will ingest the following types of data sets: raw data from Canadian data reception facilities or processors; image products from commercial satellite image providers; photo, LiDAR, and hyperspectral data sets from airborne providers; EO-derived long-term satellite data records from internal sources; and other EO-derived products from internal and other government departments. These are submitted with metadata, attributes and licences as described in the LTDP/Preserved Data Set Content document. The following is a selection of some of the EODMS's functionality, which is illustrated in Figure 3:

- Ingest multiple mission SIPs and have the ability to view SIP content in ISO 19115 format, with data validation and verification schemes. The image formats that will be ingested are: GeoTIFF, NITF, JPEG2000, CEOS, TIFF, and JPEG. The system will also ingest EO raster or vector derived data sets such as Esri Shapefile and KML/KMZ, with metadata and generic raster domain with metadata.
- View, manage, update, delete, and create AIPs and DIPs through a configurable data flow chain with suitable reporting, audits, logging and failure and security alerts.
- Set up various user accounts with access rules related to the dataset's attributes and will respect licensing rules and restrictions associated with the data sets.

- Support discovery of multi-mission data sets through a catalogue and through a Catalog Service – Web (CSW) interface.
- Ensure data integrity and consistency across the entire system and prevent the corruption, orphaning or loss of information.
- Support three working environments:
  - *Development environment* – used to develop new code and interfaces.
  - *Test environment* – used to apply patches, upgrades, and new features before being deployed to the production environment. The test environment will be a duplicate of the production environment as it will be used to run simulations of any changes to the system.
  - *Production environment* – used for normal daily operations.
- Provide the status on all subsystems of the operations management and administration functional entities and generate client reports.
- Manage end-user licence agreements (EULAs) specific to an individual data set.
- Create DIP profiles for a package creation subsystem, which define data sets and their components.
- Report and log user interactions with the system and conduct security checks.
- The access subsystem client will export areas of interest (AOIs) from a mapping interface or a client’s profile as an AOI file (e.g., Esri Shapefile, KML/KMZ, or GML) and users will be able to “clip” certain dataset collections or sub-collections by defining a clipping area or a geographic extent within the web mapping interface.



## EODMS Connections

The CEODAS preservation strategy is in alignment with CCRS’s mandate, as per the *Department of Natural Resources Act* to “...participate in the development and application of codes and standards for technical, geophysical and geodetic surveys...” and to “promote the development and use of remote sensing technology” (1994, c.41, art. 6, para d). Also, CEODAS and EODMS contribute to the CCRS Space EO Data Stewardship pillar set by the Director General in 2011. The CEODAS project also aligns with the following GoC policies: Information Management, Operational Security Standard, Management of Information Technology Security (MITS), Web Usability, Accessibility and Common Look and Feel, Geospatial Standard, Standard on Metadata, Information Technology; and Web Interoperability. The EODMS will benefit the CGDI as it will develop a state of the art catalogue and portal for data discovery based on CGDI-endorsed standards and will be accessible to CGDI stakeholders. In addition, CCRS must adhere to Mission specific Memoranda of Understanding (MOU), agreements with private sector data providers, the National Master Standing Offers (NMSOs), the Remote Sensing Legislation and Regulations as discussed earlier, and other overarching laws discussed in Chapter 3.

The CEODAS will also support the mandate of the Canadian Space Agency (CSA), which participates in the development and operation of numerous EO space missions and the development of EO data products and services. The CSA will collaborate with the CCRS to meet its EO data service delivery objectives. The Department of National Defence (DND), through its Polar Epsilon Program, has access to RADARSAT-2 data which are preserved at CCRS. Also, DND’s Mapping and Charting Establishment is consolidating some of its access services and aims to interoperate with CCRS’ metadata holdings and to streamline their EO ordering process. CCRS also collaborates with the German Aerospace Centre (DLR), US National Oceanic and Atmospheric Administration (NOAA) and the European Space Agency (ESA) on the development of the EODMS. DLR has a long-standing relationship with CCRS, NOAA has been leading and funding an interoperability catalog initiative and CCRS is working on a functional prototype for it, and ESA is instrumental in the development of the OAIS and has offered to share tools and subject matter expertise.

## EODMS Designated Users

CCRS’s current EO users include: 14 federal departments and agencies (e.g., Canadian Ice Service, Canadian Forest Service, Parks Canada, Department of Fisheries and Oceans, Environment Canada, Agriculture and Agri-Food Canada), academia, and national and international agencies. Some of their international partners include the United States Geological Survey (USGS), the National Aeronautics and Space Administration (NASA), ESA, NOAA, the Brazil Space Agency (INPE) and the China National Satellite Meteorological Center. The EODMS will provide access according to the requirements of these and other users. In January 2013, CCRS conducted a *CCRS EO data long-term preservation user requirement survey*. The results were intended to identify EODMS’s designated communities and inform system

configurations, especially the **Access** functional entity of the *OAIS Reference Model*. Users will gain single portal access to global EO data sets instead of having to access multiple mission or theme based portals, while adhering to interoperability standards will increase the ability to integrate data more readily into decision-making tools.

### 5.1.3 Challenges Encountered

The current CCRS preservation system is limited in its ability to provide easy access to preserved EO data, the data volume is increasing and its technology is becoming obsolete. The transition from a traditional mission archives and the moving of its collection to a new location necessitated the creation of a new and standardized digital geospatial data preservation system. It is important to keep in mind that CCRS has been able to preserve and make available over 800 terabytes of EO mission data for over 40 years despite significant changes in technology, storage media and computerization during that period. The efforts of CCRS EO scientists, IT professionals and management represent leading preservation practices, as no other GoC geospatial data producing organization can make such claims. It is therefore not surprising that the current system has some of the following identified shortcomings which are to be addressed with the new EODMS and in the CEODAS preservation strategy:

- Due to the practice of using non-standard metadata from EO data producers, CCRS does not currently have a standardized data ingest function, there currently exists no overall ingest framework for EO data, and many ad hoc loaders were created over time to ingest data from different types of commercial and Canadian satellite receiving stations.
- The current data management system and catalogue operate in an aging architecture, there are many lines of code from many organizations in one system and the expertise to segment service delivery is limited.
- There is no published CCRS preservation planning policy or data preservation plan in place, but the CEODAS is a start on a preservation strategy.
- The current storage system is limited.
- There are too many catalogues and discovery metadata are not standardized according to the TBS standard (i.e., ISO 19115).
- Roles and responsibilities are unclear and the policy implementation path is undefined. There is limited training of personnel, procedures and policies are unspecified, and there is a lack of documentation.
- There are limited financial resources available to support the long-term preservation of EO data. EODMS is being incrementally funded and a cost model is under development. Costs up to and including 2015 have been identified.
- Mission-specific MOUs are not digitized or accessible as they are paper based, nor are EULAs. These files are retained by multiple mission leads and are in different mission specific catalogues. Requirements and restrictions are adhered to in the system, but these are not documented nor are they easily available to users. Terms and conditions, access privileges and retention expectations are articulated in these documents.

- Some “processors” that turn raw EO data into “products” are not under CCRS’s control, therefore access to algorithms and software for the purpose of preservation are uncertain.
- Licences are not necessarily interoperable, as they are tied to specific missions.
- Policies for the patriation of Canadian satellite mission data from the international satellite receiving stations need to be re-examined as they are inconsistent. MOUs are generally the instruments used between nations regarding the collection and patriation of mission specific data.
- The originators of the data – often private sector entities – are not mandated to preserve the data in their holdings or the data they process, and do not provide access to EO image processing algorithms and software.
- Retention policies are mission-specific; however CCRS has actively been engaged in preserving all of the data from its missions.

Most of these issues have been identified in the CEODAS preservation plan and are to be addressed with the new EODMS. Once EODMS is in place, however, a risk analysis and emergency preparedness plan should be developed. Also absent from the CEODAS plan is a third-party assessment of the implemented plan and the preservation system. In addition, there is currently no explicit agreement between Shared Services Canada and EODMS. This is essential as the preservation system is a separately managed subsystem within SSC and may need a different administrative environment.

#### 5.1.4 Lessons Learned

This case study primarily described a preservation strategy that is in progress, a preservation system that is not yet built and shared insight collected about an existing EO archives that will soon be replaced. Irrespective, there are some important lessons learned from CCRS’s experience, such as:

- User requirements, the volume of data, and shifts in technology among other things, motivated the transition to a new system. This fueled the organization to closely examine its existing system, to identify its strengths and limitations and to address these by developing a preservation strategy and developing the specifications of a more formal preservation system.
- CCRS created the CEODAS strategy and EODMS specifications by adopting international best practices, including the *OAIS Reference Model*, *LTDP Guidelines*, and *TRAC Checklist*.
- CCRS has participated in the global EO community and has developed close relationships with a number of EO data creators, and is now able to leverage those relationships into knowledge and skills transfer collaborations to help it build its archives and to inform ongoing international endeavours.

## 5.2 Profile: Ontario Geographic Information Archive (GIA)

---

### 5.2.1 Introduction

The Geographic Information Branch (GIB) of the Ministry of Natural Resources (MNR) in the Government of Ontario implemented a Geographic Information Archive (GIA) in 2009 to preserve geospatial data accessioned from the Land Information Ontario (LIO) Warehouse. The MNR LIO Warehouse is used to disseminate geospatial data to MNR staff, the Ontario Public Service (OPS) and external clients. It is also the official repository for geospatial data both created and used by MNR as reference material to track the change of natural resources over time, and to support decision-making, information gathering, program activities, legislation and policy, and scientific research. The objectives of the GIA are to: preserve vital geospatial data records to comply with legislative and policy requirements; support MNR's Information Management Strategy; preserve data of scientific and historical value for long term analysis; and better defend long-term decision-making around the protection of valuable resources.

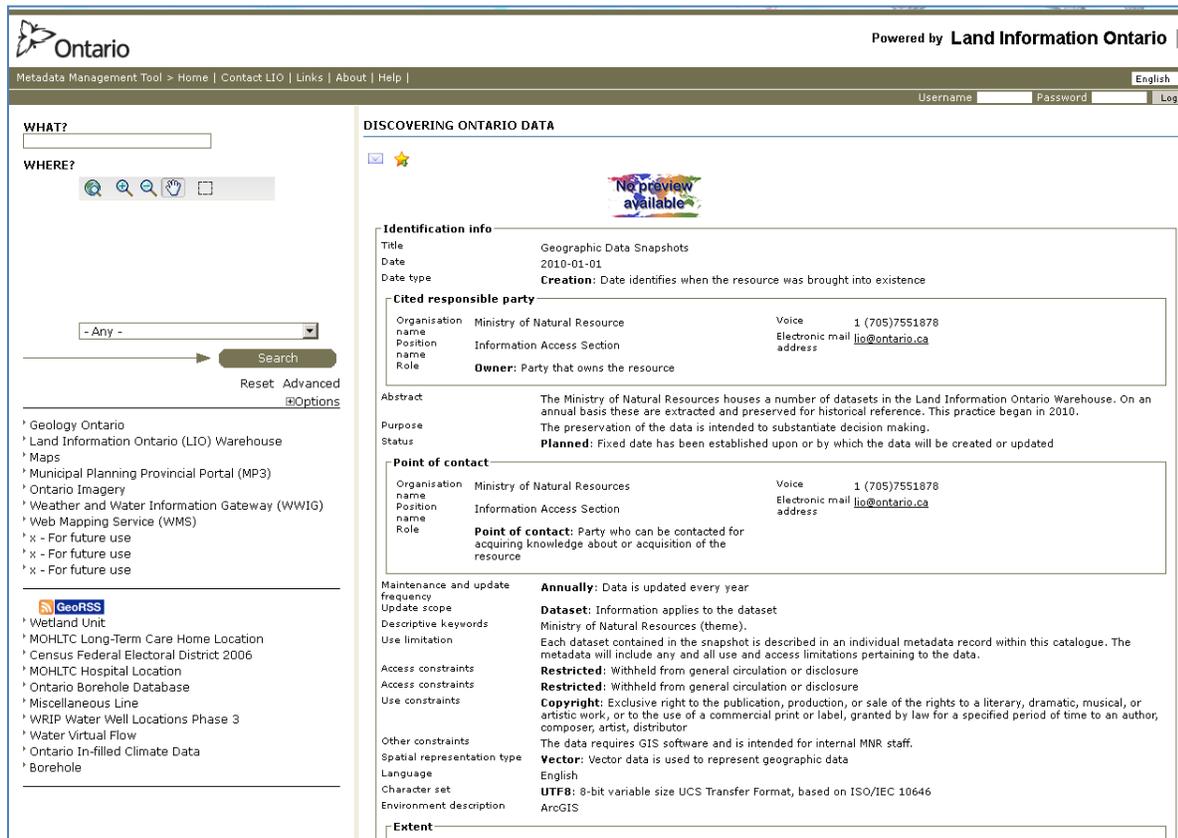
The GIB conducted an environmental scan to research archiving policies and practices in other jurisdictions, leverage current research and identified successes, and better define MNR's preservation challenges. MNR staff identified the Ministry's need for the long-term preservation of retired geospatial data from the LIO warehouse, some of which needed to be kept in perpetuity to meet business requirements.

### 5.2.2 Operational Model in Use

The GIA is a customized, near-line storage solution with data are stored on a series of hard drives. Two back-up copies are created and the three copies are all stored in different locations. There are two classes of records in the GIA – snapshots of the LIO Warehouse taken yearly and retired records, each with different preservation approaches. Snapshots of all data with their metadata, Standard NRVIS Interface Format (SNIF) packages (used to disseminate data sets to users ordering data from the LIO Warehouse) and technical reports stored in the LIO Warehouse have been taken yearly since 2009 and loaded into the GIA. SNIF packages contain header information, which provides metadata about the package content including the data source, the file with the spatial data projection information, the spatial geometry of the data, and some non-spatial (attribute) data elements of the data set which are associated with the spatial objects in the SNIF package. This data is extracted from many different database tables and is provided as custom text files in the SNIF package.

In accordance with Records Schedule MNR-4401-01, an annual snapshot is taken of all geospatial data stored or disseminated through the LIO Warehouse along with associated documentation such as metadata, data management models and technical bulletins (see example in Figure 4). The retention strategy adopted is to keep annual snapshots for 10 years, after which only every 5th year will be preserved for 200 years. After 200 years, the records are transferred to the Archives of Ontario.

Figure 4: Screen Capture of the Description of a LIO Warehouse Preserved Snapshot



### 5.2.3 Good Practices

In collaboration with the GIB the MNR Information Access Section prepared a *Guideline for Retiring and Retaining Geospatial Data Stored in the Land Information Ontario Warehouse* (OMNR, 2010). The purpose of this guideline is to ensure that all geospatial data moved from the LIO Warehouse to the GIA support the MNR’s mandate. Also, the plan provides guidance on how to retire a record class from the warehouse.

The *Retiring and Retaining Guidelines* are quite comprehensive, clearly articulate roles and responsibilities, and provide a list of issues to consider for each class of record. In this preservation system the roles include Information Owners, the Information Access Section, the Land Resources Cluster and users. For example, only an Information Owner has the right to retire records from the LIO Warehouse, and data cannot be removed without completing a retirement process and obtaining the agreement of the Manager of Information Access Section. When removing a record, the Information Owner needs to take into account appropriate communications, what impacts there might be on data users, and whether the data product is official and must be preserved or transitory and can be deleted. A series of appraisal questions are provided to help determine the business value of the record (see text box).

In addition, Metadata Resources and Training tools are provided with education modules, descriptions of adopted standards (e.g., ISO 19115) and implementation guides (OMNR, 2011).

### GOOD PRACTICE

The OMNR *Retiring and Retaining Guidelines* employ the following questions to help users determine the type of record:

- Is the geospatial data record to be retired official?
- Do the data involve or reflect any legal right of the Government?
- Will the data be needed to defend the Government against charges of data fraud or misrepresentation?
- Could the data be useful to other geospatial data users or the broader geospatial data community?
- Will other users require access to the data?
- Have the geospatial data been made available to other users through data sharing agreements or a clearinghouse?
- Can secondary users understand or interpret the data without technical expertise from the producer?
- Are the data difficult or expensive to replicate?
- Are there significant costs or consequences to the program if the data are lost?
- Can the data be usefully integrated with newer data resulting from improved methods of data collection?
- Does the estimated research value of the data exceed the costs to maintain them for secondary use by researchers?
- Will the data be useful for analyzing geographic distributions over time?
- Do the data support the study of geophysical changes over time?

## 5.3 Profile: Integrated Science Data Management (ISDM)

### 5.3.1 Introduction

Within Fisheries and Oceans Canada (DFO), the Integrated Science Data Management Service is responsible for: managing and preserving physical, chemical and biological ocean data collected by DFO, or acquired through national and international programs conducted in ocean areas adjacent to Canada; and disseminating data, data products, and services to the marine community (DFO, 2012). This information is used in real time for weather forecasting, ship routing, and prediction of weather windows for conducting weather sensitive offshore operations. The data management program also 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. The data maintenance commitment extends across all kinds of scientific data for which DFO is responsible, as evidenced by the department's *Management Policy for Scientific Data*, which includes several references to the requirement for data archiving and preservation (DFO, 2001).



### 5.3.2 Operational Model in Use

As the Responsible National Oceanographic Data Centre for Drifting Buoys (RNODC), ISDM has managed a continuously updated repository of data collected from around the globe since 1978 (DFO, 2013). The RNODC is a national data centre assisting the World Data Centres. In order to preserve drifting buoy data, it is deposited with the ISDM at the earliest possible time after capture. These initial files are then replaced by higher quality versions of the data as they become available. ISDM preserves all of the information associated with the data, including all of the data quality flags and origination information (DFO, 2012).

DFO scientific data are managed as part of an integrated system accessible through regional, zonal and national data centres. ISDM functions as a national data centre for departmental data with preservation functions shared as appropriate with existing regional data centres. ISDM provides co-ordination among the centres as appropriate, to ensure that all data are properly managed (DFO, 2001).

All DFO science project proposals and plans must demonstrate the existence of a comprehensive data management plan, which must include strategies and schedules for the transfer of the data to the responsible data centre. The project budget must clearly indicate the allocation of resources for data management and how these resources will be used. DFO Science and Oceans managers are responsible for ensuring that data collectors under their control submit their data, as well as data collected under contract, to the appropriate data centre in a timely fashion. Data

encompassed by this policy include the data sets identified below, and any other scientific data that may be created or otherwise acquired by DFO:

- Freshwater and marine habitat data;
- Meteorological data;
- Fisheries data;
- Biological oceanographic data;
- Hydrological data (e.g. Flow volumes of streams and rivers);
- Experimental Lakes Area (ELA) data;
- Freshwater biological data;
- Marine chemistry data;
- Fish health data;
- Biological data (from catch sampling, trawl and acoustic surveys, sentinel fisheries and industry surveys, science logbooks, etc.);
- Field and lab data in support to stocks' assessment process;
- Contaminants data;
- Physical oceanographic data; and
- Data collected by the Canadian Hydrographic Service, subject to CHS agreements and operational practices.

### 5.3.3 Good Practices

The DFO data management policy ensures that data are quickly copied into a “managed” environment where they are properly backed up and secured from accidental or circumstantial loss, and where the supporting metadata are linked with the data to preserve the long-term usefulness of the preserved data sets.

Each month, a summary of the drifting buoy data received in real time is published on the [web](#) along with global and regional maps of drifting buoy tracks for the previous month.





The IPYDIS goals and intentions are being carried forward through a voluntary international network of data centres and portals. In Canada, the challenge of preserving IPY data has been taken up by the Canadian Polar Data Network (CPDN), which was established in 2010 with funding from the Canadian International Polar Year Program Office, Aboriginal Affairs & Northern Development Canada (CPDN, 2012). The current CPDN partners are:

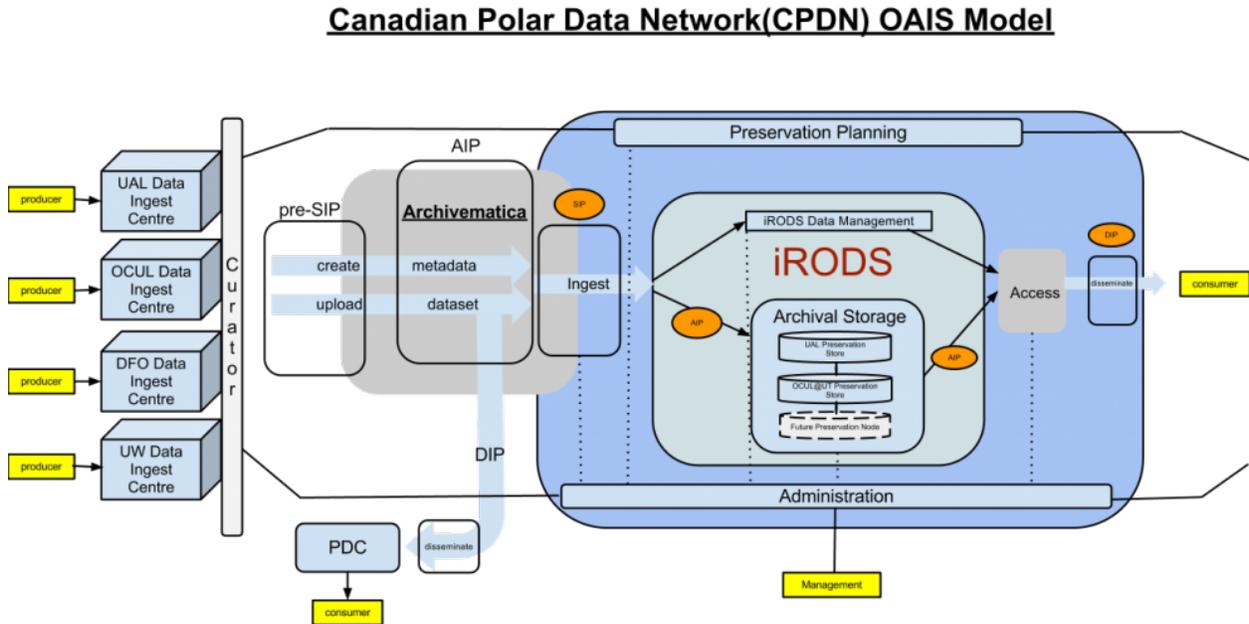
- [NRC Canada Institute for Scientific and Technical Information \(CISTI\)](#)
- [Department of Fisheries & Oceans, Science Sector \(DFO\)](#)
- [Polar Data Catalogue, University of Waterloo \(UW\)](#)
- [Scholars Portal, Ontario Council of University Libraries \(OCUL\)](#)
- [University of Alberta Libraries \(UAL\)](#)

The preservation purpose of the CPDN is to provide “a secure network housing the infrastructure needed to provide long-term preservation of digital research data,... [which] requires an archival information system that constantly verifies data integrity and upgrades to new standards over time.” The operational archival system is based on the OAIS Reference Model, as illustrated in Figure 5.

**GOOD PRACTICE**

IPY Data Policy (Parsons et al, 2011)  
 ...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.

Figure 5: Illustration of CPDN Operational Archival System



Source: Canadian Polar Data Network ([http://polardatanetwork.ca/?page\\_id=101](http://polardatanetwork.ca/?page_id=101))

### 5.4.3 Challenges Encountered

In 2011, an assessment was conducted of how well IPY performed against specific objectives within the elements of the data strategy (Parsons et al, 2011). Table 1 summarizes performance related to the two preservation objectives of the strategy.

Table 1: Assessment of Performance against Preservation Strategic Objectives

Strategic Objective	Performance
All raw IPY data should be preserved and well stewarded in long-term archives following the ISO standard Open Archives Information System Reference Model	Many disciplines did not have long-term archives. Long-term, archival standards are still evolving and adherence to good practices is highly variable across projects and disciplines. Beyond ongoing government commitment in some disciplines, no clear and sustainable business models have emerged to support long-term data stewardship.
Data should be accompanied by complete documentation to enable preservation and stewardship.	Most documentation is <i>ad hoc</i> and largely geared towards discovery. Some guidelines on documentation have been developed on a disciplinary or project basis, but some issues, such as describing detailed and ongoing provenance, have not been resolved in the general archiving community.

At the time of the assessment it appeared that only 30 of 124 IPY projects had adequately considered long-term preservation. Since more than 75% of IPY projects collected data without clear plans or resources for archiving, it has been a challenge to identify all the IPY data collected, let alone ensure they find their way to secure archives (e.g., obtaining metadata remains a large challenge). Many IPY investigators were unclear about their data preservation responsibilities or where they should submit their data. In many disciplines, long-term archives do not exist, and there is no comprehensive data preservation strategy reaching across disciplines and nations. The primary root causes of these challenges are:

- the difficulties that most scientists face in finding the time to prepare data for preservation; and
- the lack of sustained resources for data centres to preserve IPY data and ensure coordination across these centres.

#### 5.4.4 Lessons Learned

The key lessons learned concerning archiving and preservation of IPY data can be summarized as follows (Parsons et al, 2011):

- Although many IPY project managers may have assumed that the ICSU World Data Centers (WDCs) would be the natural home for much IPY data, the WDCs as a whole have not been a central or leading force for IPY data management. It is hoped that the new [ICSU World Data System \(WDS\)](#)<sup>5</sup> will better serve polar science in the long run by growing a true data network.
- Scientists need incentives to share and describe their data and to adhere to relevant data strategies and policies. Funders have to allow data preservation as an acceptable expense when granting funds, as well as the enforcement suggestions above.
- Long-term data preservation needs to be a consideration throughout the entire scientific process, and this requires a major shift in some of the institutions of science. For example, Universities need to include data management instruction as a core requirement of advanced degrees. They should consider data publication and stewardship equally with journal publication in conferring degrees, advancement and tenure.
- New business models are required that can provide sustained support for preserving dynamic and evolving scientific data. IPY experience suggests that data preservation is most successful when nations commit program resources to data management and coordination, and provide explicit repositories for preservation.
- Nations should fund archives to fill disciplinary gaps and require archives to work together on standards and interoperability as a contingency of their funding of interdisciplinary research.

<sup>5</sup> WDS is striving to ensure the long-term stewardship and provision of quality-assessed data and data services to the international science community.

# 6. Establishing a Geospatial Data Preservation System

## 6.1 Introduction

This chapter provides organizations considering the development of a digital geospatial data preservation system with guidance on planning and implementing such a system, the steps of which are illustrated in Figure 6. The *InterPARES 2 Preserver Guidelines* inspired the logical structure of the chapter, while aspects of the *OAIS Reference Model*, *LTDP Guidelines* and the *InterPARES 2 Creator Guidelines* are also present. Reference is made to the functionality, processes and documentation from the *TRAC Checklist* that are required for a preservation system to be certified as a Trustworthy Digital Repository (TDR). While there is overlap between the references, each treats the development of preservation systems from a different perspective, and readers are encouraged to select ideas that best match their preservation context. In addition, the reader will be pointed to useful tools and resources and to good practices derived from the case study and profiles. While the information provided is by no means comprehensive, it will familiarize the reader with the overall scope of preservation system creation.

Figure 6: The Process of Establishing a Geospatial Data Preservation System



## 6.2 Establishing the System's Scope and Objectives

Planning any records preservation initiative starts with scoping. The scoping exercise most often takes the form of a business plan, and includes: identifying business drivers for the system; developing programmatic preservation goals that tie in with legislation, regulation, policies and directives; identifying the minimum level and type of reuse which the archives will need to maintain for its user community; conducting an inventory of technology and assessing its suitability; identifying standards; establishing a budget; and creating an evaluation process. The GeoMAPP [Geoarchiving Business Planning Toolkit](#), which includes the [Geoarchiving Business Planning Guidebook](#), a [Geoarchiving Self-Assessment Template](#), and a [System Inventory Template](#), are invaluable tools for this part of the planning process.

The objectives of the preservation system should reflect legislative requirements, applicable policies and directives, business needs, and designated

community requirements. Objectives should concisely describe the system's mandate, how it will be achieved, and what the guaranteed level of service and mode of engagement supported by

### KEY POINTS

- Development of a business plan will help define scope
- A key consideration is identifying the optimum system configuration
- Appraise what data are to be archived and assess current staffing, administrative conditions, software, networking, storage and standards

### GOOD PRACTICES

The *OAIS Reference Model* includes scope determination as a managerial function and suggests that scope determines the breadth of both the producer and the consumer groups that are served by the archive.

The *LTDP EO Preserved Data Set Content* report provides insight into basic preservation principles and a few items to consider when scoping for the preservation of mission data, such as retention time, accession schedules, what should be preserved and how to do so.

The *LTDP Guidelines* provide staffing recommendations but emphasize that operations are governed through an organizational structure that oversees planning and operations.

The *TRAC Checklist* provides criteria for administrative responsibility, organizational viability, and procedural accountability. For example, the archive is expected to have: a mission statement that demonstrates its commitment to long-term preservation; designated staff with the necessary skills and training for ongoing development; clearly articulated roles, skills, and job descriptions; an organizational chart; and an ongoing development plan.

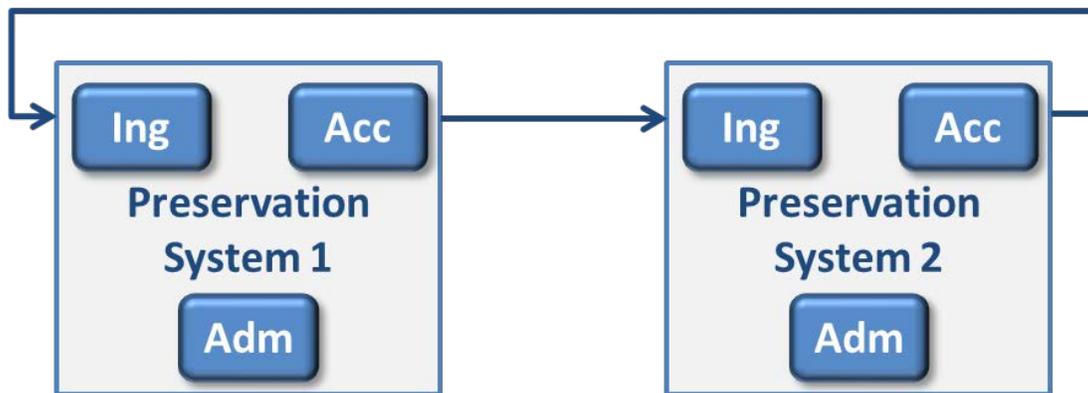
the preservation system will be. A key objective will be to ensure the preservation of metadata and data products of known quality at all levels required by users, or the capability to generate them on request through proper processing.

The scoping exercise should also consider collaboration among organizations, as this will drive organizational, technical, and business arrangements and processes. The [OAIS Reference Model](#) describes the following models, with the first three having successively higher degrees of interaction:

**Independent preservation systems** are local or thematic and may be located at one site or be physically distributed over many sites. They may choose to design DIPs and discovery mechanisms based on formal or de-facto standards, which could facilitate voluntary cooperation with other preservation systems that implement the same standards.

**Cooperating preservation systems** have agreements in place among two or more archives. The simplest form of cooperation is when one system acts as the consumer of material from another, in which case it must support the DIP and SIP formats of the producing system. No common access, submission or dissemination standards are assumed, but cooperating groups must support at least one common SIP and DIP format to enable inter-system requests. Figure 7 illustrates the concept of cooperating preservation systems where ingest and access entities share SIP and DIP formats.

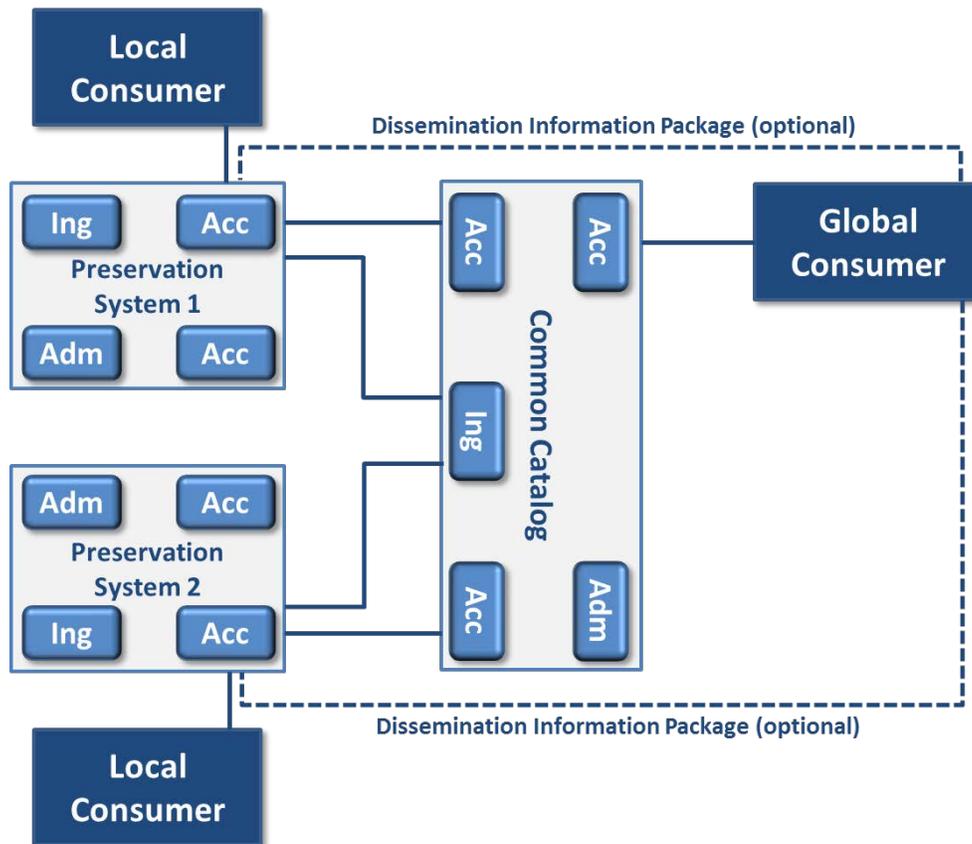
**Figure 7: Cooperating Preservation Systems with Mutual Exchange Agreement**



Source: CCSDS (2012)

**Federated preservation systems** are conceptually consumer-oriented, and serve both a local community (i.e., original designated community) and a global community (i.e., an extended designated community) that has interests in the holdings of several preservation systems. The global community may influence archives to provide access to their holdings by means of one or more common information [finding aids](#). Figure 8 illustrates the functional architecture for two preservation systems that have similar designated communities and have decided to federate to allow consumers to locate AIPs of interest from either in a single search session.

Figure 8: Federated Preservation Systems Employing a Common Catalog

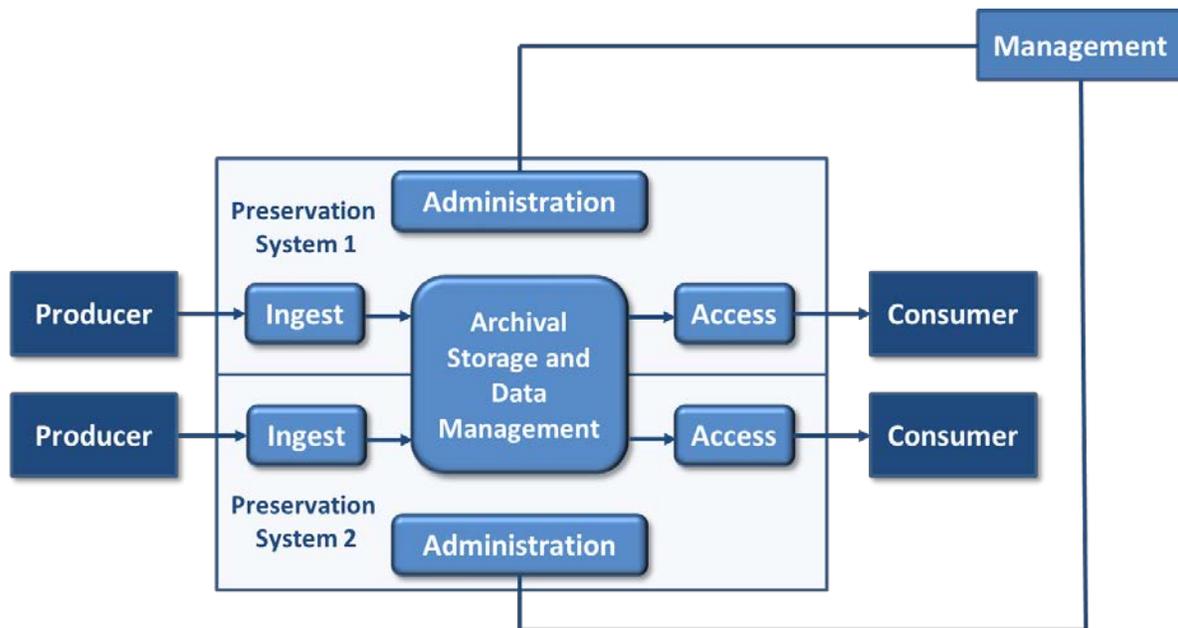


Source: CCSDS (2012)

The common catalog is the external (global) binding element that serves as a common access point for information in both systems. DIPs containing the finding aids from each preservation system are ingested into the common catalog, which may limit its activity to being a finding aid or include the common dissemination of products from either or both, as shown by the dashed lines in the figure.

**Preservation systems with shared resources** are more integrative, whereby management from multiple systems has entered into agreements to share or integrate functional areas. This association is fundamentally different from the previous examples, in that the internal architecture of the preservation system must be taken into account. Figure 9 illustrates the sharing of a common storage function, consisting of an Archival Storage entity and a Data Management entity, between two systems. Each system can serve totally independent communities but, for the common storage element to work, standards are needed at the internal Ingest-storage and Access-storage interfaces.

Figure 9: Preservation Systems with Shared Resources



Source: CCSDS (2012)

The current EO data preservation situation in Europe, for example, is fragmented with distributed and independent centralized systems, using different technologies and different AIPs and DIPs. One of the objectives of the [LTDP Guidelines](#) (and its accompanying [LTDP/PDSC document](#)) is to enable greater standardization to facilitate collaboration among European and Canadian EO data holders and their systems.

### 6.3 Defining the System's User Community

Identifying who are the users, consumers or designated communities of a preservation system will determine the scope, objectives and configuration of the system, the type of data the system will ingest, and the kind of access functionality required to meet user needs. The overall user community can be composed of different designated communities, which generally have different skills, knowledge bases and resources. Designated communities can include scientists, researchers, businesses, value-added resellers, and the general public and can be further differentiated on the basis of respective application domains and areas of interest (e.g., ocean, atmosphere, infrastructure, land administration, etc.).

In addition, new users having different objectives for the use of the data and completely different skills and knowledge base than the ones identified today may want to access the preservation system in the future. Therefore, the definition of a “designated community” should be generic and large enough so that the identified content to be preserved in the long term for that community will allow other users, not considered at the time preservation was initiated, to make use of the data in the future.

It is a good practice to define the user community and preservation objectives together in order to meaningfully contain the scope of what is to be preserved in the long term. It is also a good practice to engage users in the planning process, since their involvement will help to achieve buy-in for the process itself and contribute to business plan that best meets their needs. The CCRS CEODAS program, for example, conducted a survey to gain a better understanding of its EO data users and the results informed access functionality. The North Carolina Geospatial Data Archiving Project also conducted a survey of local creator agencies to determine the frequency of geospatial data capture and practices for their archives (NC CGIA, 2006).

### KEY POINTS

- Needs of all current designated user communities must be identified
- Definition should be broad enough to meet different needs of new future users
- Engaging users in system planning will increase chances of success

In the [TRAC Checklist](#), user needs drive how documentation is described and understood by that community, and clear operational definitions need to be tied to user requirements and to how humans will understand the data. While documentation strategies and standards must satisfy professional requirements, they must also be relevant to designated communities. Also, preservation system technologies need to be appropriate to the services provided to users. Finally, the [LTDP Guidelines](#) theme for access and interoperability recommends that users be able to readily access the EO data in standard formats, and software and services be interoperable to ensure homogeneous access to EO data accessed from heterogeneous systems in the long term.

## 6.4 Acquiring and Managing Resources

Substantial resources are required to fund technological capabilities and the human resources required to operate a geospatial data preservation system. Organizations can acquire new or reallocate existing resources or leverage collaborative relationships with others. Some of the collaborative distributed models discussed in Section 6.2 represent examples of leveraging

### GOOD PRACTICE

Evidence of acceptable financial management practices under the *TRAC Checklist* includes:

- operating plans, budgets, financial reports, and audits are in place;
- business plans are reviewed annually;
- financial practices and procedures are transparent and audited by third parties;
- risks, investments and expenditures are regularly analyzed and reported; and
- a commitment exists to bridge funding gaps.

resources. Resources also need to be sustainable and a solid communication plan accompanying the business plan is helpful to convince funding sources, as is a flexible funding strategy. The CCRS CEODAS program, for example, is incrementally developing its infrastructure as funds become available.

Financial sustainability is of significant concern in the [TRAC Checklist](#), which recommends that an archives develop a sustainable business plan that includes financial implications related to preservation system development and production activities, and indicates the level of financial support from contributing agencies, subscribers and other parties. In addition, it should include how the future costs of migration, capital improvements and enhancements will be covered.

Contingency plans should also be formulated to deal with catastrophic failures and to ensure the preservation system is insulated from political uncertainties if it is government-sponsored. The [GeoArchiving Comprehensive Cost-Benefit Analysis Guidance](#) tool developed as part of the GeoMAPP *GeoArchiving Business Planning Toolkit* can help with planning the acquisition of resources process (GeoMAPP, 2011).

## 6.5 Preservation Planning

The next step in establishing a preservation system is planning the records preservation process. Preservation activities always involve a mix of research into evolving technologies and the development of new preservation strategies, as well as the day-to-day activities required to maintain existing holdings. Each framework divides these two types of responsibilities somewhat differently.

A preservation system must demonstrate that it has sound, documented preservation policies, practices and procedures in place for it to be trusted. In the [TRAC Checklist](#) (which includes a guide for the development of *Preservation Planning & Strategies*), this activity involves the identification of the most appropriate actions to guarantee the preservation and future usability of records in a preservation system. Preservation planning is an [OAIS Reference Model](#) entity that provides the services and functions for monitoring the environment of the preservation system. The result of this activity will be a “preservation strategy and approach” document describing: all the AIPs (or in the case of the *LTDP Guidelines*, the PDSC Inventory elements); the elements on which they are dependent, or which are necessary to understand and use them; the associated preservation actions identified for each of them; and the preservation state of each element.

In the [OAIS Reference Model](#), policies and procedures are considered as inputs into a preservation system and preservation planning includes the following functions:

- *Monitor Designated Community* – This involves tracking changes in consumers’ and producers’ service requirements and available product technologies (e.g., data formats, media choices, preferences for software packages, new computing platforms, or mechanisms for communicating with the Archive). It provides reports, requirements alerts and emerging standards to the *Develop Preservation Strategies and Standards* function and sends preservation requirements to the *Develop Packaging Designs and Migration Plans* function.
- *Monitor Technology* – This includes tracking emerging digital technologies, information standards and computing platforms to identify technologies which could cause obsolescence in the Archive’s computing environment and prevent access to some of the system’s holdings. It sends reports, external data standards, prototype results and technology alerts to the *Develop Preservation Strategies and Standards* function and sends prototype results to the *Develop Package Designs and Migration Plans* function.
- *Develop Preservation Strategies and Standards* – This involves development and recommendation of strategies and standards, and assessment of risks, to enable the preservation system to make informed trade-offs. Based on information received from the *Monitor Designated Communities* and *Monitor Technology* functions and from Administration, this function identifies changes that would require migration of some current preservation system holdings or new submissions (e.g., updating AIPs with additional or revised Representation Information). Periodic risk analysis reports and recommendations on system evolution and on AIP updates are sent to Administration.

- *Develop Packaging Designs and Migration Plans:* New information package designs and detailed migration plans and prototypes are developed to implement Administration policies and directives, and advice is provided on their application to specific Archive holdings and submissions. This function applies format, metadata and documentation standards to preservation requirements and provides AIP and SIP template designs to Administration. Migration goals may require transformations of the Content Information to avoid loss of access due to technology obsolescence. This work may involve the development of new AIP designs, prototype software, test plans, community review plans and implementation plans for phasing in the new AIPs, requiring consultation with other preservation system functional areas and the designated community. Preservation Planning develops, validates and supplies the migration packages and Administration schedules and performs the migration.

The [TRAC Checklist](#) contains the following requirements for preservation planning systems, which are to be articulated in an archives' documented policies and procedures:

- Make relevant decisions about file formats.
- Have a comprehensive automated and, or manual workflow to accession digital objects including transfer protocols, clear creator and preserver roles and responsibilities, and explicit evidence of file conversion that occurs when AIPs are generated from SIPs.
- Anticipate and, or apply preservation actions pertaining to AIPs (e.g., testing and applying preservation plans, creating action logs, etc.).
- Access preservation system storage policies, procedures and practices to ensure the effective use of reliable storage, and be responsive to technological change.
- Have an independent means to verify preservation system content based on secure traceable digital objects (e.g., an auditable acquisition register and an inventory that cannot be altered)<sup>6</sup>.

#### KEY POINTS

- Preservation plans are necessary to guarantee the preservation and future usability of records in the system
- Preservation plans must be well documented in the system's policies and procedures
- The impacts of technology change are particularly important and must be monitored

The archives must have formal technology watches in place as well as mechanisms for monitoring and notification when metadata and formats are approaching obsolescence, or software and hardware changes are needed, and be able to change plans as a result of its monitoring activities. A technology watch procedure, hardware inventories and designated community profiles are the evidence that these requirements are being met. The archives must

<sup>6</sup> The creation of digital object identifiers (DOIs) for datasets is one of the strategies to create persistent identifiers. The National Research Council of Canada launched *DataCite Canada* in 2012 to help data creators with this process and it is also Canada's DOI allocation agent. This is part of a larger international initiative by national-scale libraries and research organizations to make data more accessible on the Internet (NRC, 2010)

also demonstrate that its preservation plan is effective and this can be done by having users test the systems over time.

In addition, an archives must demonstrate that it has a well-supported operating system and other core infrastructural software, that its back-up system is sufficient, and that copies of records are managed and synchronized. Also, bit corruption loss must be detectable and any errors or incidents need to be reported to the systems administration. Finally, change management processes must be documented, the archives must have a process to test the effects of change and react to updates, and a risk analysis process must be in place.

The InterPARES 2 [Preserver Guidelines](#) includes a list of preservation strategies (InterPARES 2, 2007) and the audit and certification documentation available on the [Scholars Portal](#) website is another valuable resource.

## 6.6 Developing Policies and Procedures

---

As part of the preservation planning process, data preservers need to develop policies and procedures (i.e., explicit instructions and rules to guide decisions and actions) to control records transfer, maintenance and reproduction. The [Preserver Guidelines](#) suggest that procedures be designed to satisfy the following three *Baseline Requirements* (InterPARES 2, 2002) to produce authentic records and ensure their identity and integrity (InterPARES 2, 2007):

- Control over records transfer, maintenance and reproduction, whereby the system and the procedures ensure that transfers between creator and preserver have adequate controls to guarantee the records' integrity and identity.
- Documentation of records reproduction activity, including: the date of reproduction; the relationship between records acquired and the copies produced by the preserver; the impact of the reproduction process on form, content, accessibility and use; and recording and communication of the fact that elements of a record are not fully reproducible.
- Documentation about the changes to the record over time.

The *Preserver Guidelines* also recommend that: security measures be clearly articulated in the policies and procedures and that these address methods to ensure the chain of custody of a record is maintained; security controls and procedures be documented and adhered to; and the records remain unchanged. Procedures should also clearly describe how creators transfer their data to the preserver and how preservers describe acquisition (transfer – ingest), processing, maintenance and the provision of access to the geospatial data records.

In the [OAIS Reference Model](#), a preservation system has the responsibility for developing policies and procedures related to:

- Negotiating and accepting appropriate information from creators;
- Obtaining control of this information to a level needed to ensure preservation;
- Determining designated communities;
- Ensuring that preserved records are understandable to the designated community;
- Protecting the records from security breaches and recovering from disasters;
- Ensuring that records are preserved against all reasonable contingencies, including the demise of the archives; and
- Making the records available to the designated communities.

To be certified as a TDR, an archives must clearly document requirements, decisions, developments, and actions to ensure long-term access to the records in its care. This documentation assures users, creators and management that the preservation system is meeting its requirements. Certification as a TDR is considered the clearest indicator that sound standards based practice is facilitated by procedural accountability. The [TRAC Checklist](#) includes elements such as written policies, procedures, protocols, rules, manuals, handbooks and workflows, specification review cycles, and update and review mechanisms. Furthermore, the preservation system must meet the needs of its designated community, and maintain written policies regarding legal permissions to preserve digital content over time, (e.g., to preclude DRM circumvention). Some of the documents pertaining to this are deposit agreements, legislation, policies and service agreements. User feedback processes and the documenting of the history of changes in operations and procedures are also necessary. Policies should include a commitment to transparency, whereby documentation about the archives' operations and management is made available to all stakeholders.

#### GOOD PRACTICES

The *GeoMAPP Best Practices for Geospatial Data Transfer for Digital Preservation* provides a step-by-step approach to the data transfer process and includes templates and a checklist (GeoMAPP, 2011).

The Scholars Portal has made available its policies and plans describing its preservation mandate and workflows, which are very useful templates for those developing preservation systems (OCUL, 2012).

The *LTDP EO Data Set Content* (LTDP, 2012) provides a procedural checklist for the preservation of EO data that is an excellent resource for data preservers.

## 6.7 Appraising Records for Preservation Value

A final planning step in establishing a geospatial data preservation system is appraising the records that have been identified for potential preservation to determine if there is sufficient value in preserving them. Appraisal is based on such considerations as the archives' objectives, its legislative mandate, its user community's needs and requirements, the host organizations' business needs and the intrinsic value of the data to users. The appraisal process includes examining the provenance and content of a record, its reliability, and its condition and the costs to preserve it. Enlisting the support of an institution's librarians and government archivists, if possible, can assist with the creation of an appraisal process. As stated earlier, they should also be a part of designing the preservation system. The GeoMAPP [Introduction to Appraisal Mentoring](#) provides a concise overview of the process.

### KEY POINTS

- Appraisal involves assessing a record's content, authenticity, provenance and reliability, and the cost of its preservation
- Appraisal collaboration between records creators and preservers is beneficial

The InterPARES 2 [Preserver Guidelines](#) recommend establishing transfer methods and identifying preservation strategies and appraisal methods collaboratively with records creators. This may include locating multiple owners of the geospatial data, and understanding intellectual property rights, contracts and liabilities. The appraisal process also consists of assessing the authenticity of the records, requiring a close examination of the chain of custody and record-keeping practices. An appraisal report should document the controls in place to guarantee a record's identity and integrity and meet the [Benchmark Requirements for Supporting the Presumption of Authenticity of Electronic Records](#). The [Creator Guidelines](#) provide guidance on how to manage records and assess those of business value that might be transferred to an archives at a later time.

## 6.8 Acquiring and Ingesting Records

Once records have been appraised, the next step is for creators and preservers to develop a shared plan for records transfer and initiate the process of acquiring selected records. Work on the plan can be started while assessing technical feasibility during the appraisal process. Acquisition procedures need to be standardized and enforced, which includes according to the InterPARES 2 [Preserver Guidelines](#): establishing, monitoring and implementing procedures to register record transfer; verifying the authority of the transfer; ensuring that the records being transferred correspond to the records designated for transfer; verifying the authenticity of the records transferred; and accessioning the records.

In the [OAIS Reference Model](#), acquisition is part of the **Ingest** functional entity and the transfer of appraised geospatial data is covered by negotiated submission agreements between the preserver and creator. Submission agreements identify the SIPs to be submitted, the length of time over which these submissions will be made to the preservation system, and the data model to be used for submissions. The data model specifies the logical components of the SIP (e.g., the Content Data Objects, Representation Information, PDI, Packaging Information, and Descriptive Information) and how (and whether) they are represented in each data submission session.

The purposes of the Ingest functional entity are to:

- *Receive Submission Information Packages (SIPs)* – Digital SIPs may be delivered from Producers (or from internal elements under Administration control) via electronic transfer, loaded from media submitted to the preservation system, or simply mounted (e.g., CD-ROM) on the system’s file system for access. Evidence for authenticity is provided by the Producer as part of the PDI in the submission, and this evidence is maintained, updated, and, or incremented by the preservation system over time.
- *Perform Quality Assurance on SIPs* – The successful transfer of the SIP to the temporary storage area is validated. For digital submissions, validation mechanisms might include [Cyclic Redundancy Checks \(CRCs\)](#) or [checksums](#) associated with each data file, or the use of system log files to record and identify any file transfer or media read/write errors.
- *Generate Archival Information Packages (AIPs)* – SIPs are transformed into one or more AIPs that conform to the preservation system’s data formatting standards and documentation standards. This may involve file format conversions, gathering adequate Representation Information, data representation conversions or reorganization of the Content Information in the SIPs. A request is sent to **Data Management** to obtain reports of information needed to produce the Descriptive Information that completes the AIP.
- *Extract Descriptive Information from AIPs* – Descriptive Information is extracted from the AIPs and Descriptive Information collected from other sources to provide to the *Coordinate Updates* function and to **Data Management**. This includes metadata to support searching and retrieving AIPs (e.g., who, what, when, where, and why), and could also include special browse products (e.g., thumbnails, images) to be used for data searching.

- *Coordinate Updates to Archival Storage and Data Management* – AIPs are transferred to **Archival Storage** and the Descriptive Information is transferred to **Data Management**. After the transfer is completed and verified, **Archival Storage** provides the storage identification information for the AIP, which is incorporated into the Descriptive Information for the AIP and transferred to the **Data Management** entity along with a database update request.

In the [TRAC Checklist](#), ingest elements related to the creation of AIPs are similar to the OAIS processes. The types of evidentiary materials related to records ingestion into archives that are used in a certification process are: mission statements, submission/deposit agreements, workflow and policy and procedures documents, processing procedures and documentation of the properties to be preserved. In addition, evidence is required that appropriate technological measures are in place to ingest the records, logs of procedures and authentication are kept, and acquisition registers are maintained.

The GeoMAPP publication, [Best Practices for Archival Processing for Geospatial Datasets](#), provides excellent guidance on a workflow for archival organizations' processing of geospatial data sets.

#### KEY POINTS

- Acquisition procedures should be developed collaboratively by creators and preservers, standardized and enforced
- Submission agreements are a good way to ensure long-term sustainability of the acquisition process
- Processes for ingestion of records into the preservation system should also be well-documented

## 6.9 Preserving Records

After the geospatial data have been acquired and ingested they need to be preserved. The process of preserving accessioned records was touched upon in the previous discussions of preservation planning, policies and procedures, appraisal and acquisition. However, some specific matters pertaining to records storage remain important to discuss. The OAIS **Archival Storage** functional entity is designed to handle the records storage processes for the preservation system. In this context, the term “media” is used to designate one or more local or remote mechanisms for storing digitally encoded information. In this model, data are received, they are managed in a storage hierarchy, media are replaced (i.e., AIP packaging information may be refreshed), errors are checked, disaster management mechanisms are constructed, and processes are in place to provide the data to consumers.

The purposes of the Archival Storage functional entity are to:

- *Receive AIPs from Ingest* – A storage request and an AIP are received from **Ingest** and the AIP is moved to permanent storage within the preservation system. An indication of the anticipated frequency of utilization of the Data Objects making up the AIP will facilitate

the selection of the most appropriate storage devices or media for storing. The media type will be selected, the devices or volumes prepared, and the physical transfer to the **Archival Storage** volumes performed. Upon completion of the transfer, this function sends a storage confirmation message to **Ingest**, including the storage identification of the AIPs.

- *Manage the Storage Hierarchy* – The AIPs are positioned on the appropriate media, conforming to any required special levels of service for the AIP or any special security measures, and the appropriate level of protection for the AIP is ensured. This function also ensures that AIPs are not corrupted during transfers, and provides **Administration** with statistics on media on-hand, available storage capacity in the various tiers of the storage hierarchy, and preservation system usage.
- *Refresh Storage Media* – The capability to reproduce the AIPs over time, without altering the Content Information and PDI. If media-dependent attributes (e.g., tape block sizes, CD-ROM volume information) have been included as part of the Content Information, a way must be found to preserve this information when migrating to higher capacity media with different storage architectures (refer to the [OAIS Reference Model](#) for a detailed examination data migration issues).
- *Conduct Error Checking* – This function provides statistically acceptable assurance that no AIP components are corrupted in Archival Storage or during any internal **Archival Storage** data transfer. Procedures should be developed for random verification of the integrity of Data Objects using CRCs or some other error checking mechanism.
- *Provide Disaster Recovery Capabilities* – Mechanisms are provided for duplicating the digital contents of the preserved collection and, for example, storing the duplicate in a physically separate facility, in accordance with disaster recovery policies specified by **Administration**. This function is normally accomplished by copying the Archive contents to some form of removable storage media (e.g., digital linear tape, CD-ROM), but may also be performed via hardware transport or network data transfers.
- *Provide AIPs to Access* – This function provides AIP(s) on the requested media type or transfers them to a temporary storage area upon receipt of an AIP request from **Access**. This function also sends a notice of data transfer to **Access** upon completion of an order.
- *Manage System Configuration* – The functionality of the entire Archive system is monitored continuously and changes to the configuration are systematically controlled. This function audits system operations, performance and usage and receives reports on system information from **Data Management** and reports on operational statistics from **Archival Storage**. It summarizes those reports and periodically provides Archive performance information and Archive holding inventory reports to **Preservation Planning**.

## 6.10 Describing the Archival Metadata

Geospatial data are described by creators in their metadata. Archival metadata describe the ingested data and its metadata (part of the SIP) and includes information about the records and

### GOOD PRACTICE

The GeoMAPP report on *Archival Metadata Elements for the Preservation of Geospatial Datasets* (GeoMAPP, 2011) based on OAIS description metadata is a useful guide. It includes a table which crosswalks OAIS to FGDC and Dublin Core metadata elements. The table was informed by:

- the CEDARS project preservation metadata elements (Stone & Day, 1999);
- the National Geospatial Digital Archive (NGDA) Report by Susan Hoebelheinrich (Hoebelheinrich, 2009), which provides recommendations for complex geospatial data sets;
- the Center for International Earth Science Information Network (CIESIN) Geospatial Electronic Records metadata model (CIESEN, 2005); and
- the PREMIS data dictionary for preservation metadata (Library of Congress, 2013).

their contexts (part of the AIP), which is collected during the appraisal, processing, ingest and preservation phases by preservers. The [TRAC Checklist](#) includes data description in its Information Management element and the following supporting documentation is required when examining a preservation system for certification: descriptive metadata, persistent identifiers associated with the AIP, depositor agreements, metadata linkages to the AIP, logs documenting referential integrity, and process flow documentation.

**Archival identity** in the InterPARES 2 [Creator Guidelines](#) is established in metadata elements that are very similar to most geospatial data elements. Archival metadata also include information about: wrappers and

encoding information, the archiving filing date, whether or not a digital signature is present, authentication indicators such as the corroboration or the means to validate the data set, an attestation of the record which is analogous to the signing of the data set when it was issued, and draft version numbers. Information about a dataset's integrity is also required and this includes elements such as: the names of handling persons or office, the names of responsible authorities, indications of annotations to the data and if any technical changes were made such as version upgrades, wrappers and encoding. In addition, elements regarding restriction codes, access privileges, the importance of the record, planned disposition (i.e., when records can be transferred to the archives) are recommended elements.

The InterPARES 2 [Baseline Requirements](#) includes a documentation requirement entitled *Documentation of Reproduction Process and its Effects*, which recommends including the following:

- The date the record was reproduced and the name of the responsible person or business unit;
- The relationship between the record acquired from the creator and the copies produced by the preserver;
- The impact of the reproduction process on the record’s form, content, accessibility and use; and
- In cases where a copy of a record is known not to fully and faithfully reproduce the element expressing its identity and integrity, document such information and make this documentation readily accessible to users.

## 6.11 Managing and Maintaining Records

---

Once records have been ingested and preserved, the next step is to establish processes to ensure their proper management and maintenance. The InterPARES 2 [Preserver Guidelines](#) emphasize records maintenance and preservation strategies, many of which were discussed in the preservation planning section. The minimum necessary requirements to protect and maintain the accessibility of authentic copies of digital records include:

- *Clear allocation of responsibilities* – a person or office having “unambiguous” responsibility to manage record storage and protection, with skilled personnel dedicated to this activity.
- *Provision of appropriate technical infrastructure* – physical and administrative resources to enable record-keeping such as buildings, computer hardware, computer networks, and auxiliary staff.
- *System maintenance, support and replacement* – an implementation plan for maintaining, updating and replacing hardware and software.
- *Transfer of data to new storage media on a regular basis* – systematic procedures for copying data from one storage medium to another to avoid media decay.
- *Adherence to appropriate conditions for storage media* – avoidance of media decay by adhering to appropriate environmental conditions such as the avoidance of moisture, dust, and heat.
- *Redundancy and geographic location* – duplication of digital entities and the storing of multiple copies on different storage media to protect against failure, and storage in different locations to protect against disasters and poor environmental conditions.
- *System security* – restriction of access to records to authorized users and processes, including restriction of physical access to computers and storage, and to records on computers.

- *Disaster planning* – disaster recovery plans in place with detailed procedures to restore a damaged system and to guide the recovery of the preservation system following a disaster.

The [TRAC Checklist](#) includes maintenance strategies that cover many of the items discussed above and also includes: backup functionality, copy synchronization, the detection of bit corruption and records loss, risk assessments to address threats and denial of service attacks, and implementation of controls to address security needs. There is also the expectation that preservation systems will adhere to [ISO 17799 Information technology – Security techniques – Code of practice for information security management](#) (ISO, 2005) and that there will be disaster recovery plans, proof of offsite copies, service continuity plans, organizational charts, and logs of recorders.

#### KEY POINTS

- Effective management and maintenance strategies are critical for ensuring ongoing accessibility of authentic copies of digital records
- Adherence to the *ISO 17999 security management code of practice for information management security* will help to ensure that preservation systems properly maintain their records

## 6.12 Providing Access to Records

The final step in establishing a geospatial data preservation system is ensuring that users can access the preserved records. How preserved data are made available to users is dependent on

user requirements and the archives mission and policies. A Web portal is often used to provide access to records, which can be searched via descriptive metadata. Mechanisms need to be designed to make it easy for users to: discover a dataset(s) of interest and how the records were created; learn about licences, use rights and potentially the cost to access that record; and place orders. Data delivery mechanisms also need to be clearly described.

#### KEY POINTS

- Open and transparent policies and procedures on records access will help users to discover and access the right data for their needs
- An Order Agreement, which may be a simple online form, is typically used as a mechanism for users to communicate with preservers

Access policies need to be public, so that authorization rules and authentication requirements that may be included in deposit agreements are transparent. Users need the functionality to track what they accessed back to the original copy. The [TRAC Checklist](#) includes a useful guide to [Understanding Digital Repositories and Access Functionality](#)

to help with developing the access component of the preservation system. Supporting evidence of acceptable preservation system access functions in the *TRAC Checklist* include: access logs, request and denial logs, system design documents, DIP production logs, and process walkthroughs.

In the [OAIS Reference Model](#), the **Access** functional entity provides the interface between the preservation system and the user. The user establishes an Order Agreement with the preservation system, which identifies the AIP(s) of interest, how those AIPs are to be transformed and mapped into (DIPs) and how those DIPs will be packaged in a data dissemination session. The Order Agreement may also specify other needed information such as delivery details (e.g., name, mailing address, etc.), rights information (e.g., usage restrictions, authorized users, or licence fees) and pricing, if applicable. The agreement may be a formal document or the process of developing an Order Agreement may be no more than the completion of an online form to specify the AIPs of interest.

The *Archival Information Update* function in **Administration** also submits dissemination requests to obtain the DIPs needed to perform its update functions. This function will determine if resources are available to perform a request, assure that the user is authorized to access and receive the requested items, and notify the users that a request has been accepted or rejected. It will then transfer the request to **Data Management** or to the *Generate DIP* function for execution as follows:

- *Generate DIP* – This function accepts a dissemination request, retrieves the AIP from **Archival Storage**, moves a copy of the data to a temporary storage area for further processing, and transmits a report request to **Data Management** to obtain Descriptive Information needed for the DIP. If special processing is required, the function accesses Data Objects in temporary storage and applies the requested processes (e.g., inserting digital rights management information and filtering the personal data to ensure consistency with the user rights). It places the completed DIP response in the temporary storage area and notifies the *Coordinate Access Activities* function that the DIP is ready for delivery.
- *Deliver Response* – For on-line delivery, this function accepts a response from the *Coordinate Access Activities* function and prepares it for on-line distribution in real time via communication links. It identifies the intended recipient, determines the transmission procedure requested, places the response in the temporary storage area to be transmitted, and supports the on-line transmission of the response. For off-line delivery, it retrieves the response from the *Coordinate Access Activities* function, and prepares and ships the response. When the response has been shipped, a notice of shipped order is returned to the *Coordinate Access Activities* function and billing information is submitted to **Administration**.

The [LTDP Guidelines](#) provide direction on creating the mechanisms to: search and discover data and mission specific documentation, maintain searchable metadata, browse an image catalogue, and employ visualization tools relevant to designated communities. The *Guidelines* also provide a list of OGC access interface standards.

## 7. Challenges and Solutions

There are numerous challenges encountered during the preservation of geospatial data. The *InterPARES 2 Guidelines*, the *OAIS Reference Model*, the *LTDP Guidelines* and the *TRAC Checklist* were created to respond to the most common organizational, digital object and technological preservation challenges. The recommendations of the *Research and Analysis Report: Geospatial Data Archiving and Preservation* (HAL, 2011) also still apply. The GeoMAPP tools and the LTDP/Preserved Data Content Definition documents are more specific and geared to help the preservers of geospatial data, while the TBS directives and policies were created to guide the data management actions of all federal government record creators. Table 2 describes some of the more common preservation challenges, good practices and lessons learned to meet these challenges as revealed by the case studies, profiles, and the document and literature research conducted during the creation of this primer.

**Table 2: Federated Archive Challenges and Solutions**

Common Challenge	Solutions
Requirements based on the type of digital geospatial data.	<p>A thorough assessment of the types of data to be ingested is conducted during the appraisal phase. At that time the digital components of each record set are defined and the elements required to ensure that they can be accessed and used are identified. The file format registry is a useful tool for this process.</p> <p>All of these elements, including the data, need to be preserved and linked to each other at submission and maintained during preservation and access.</p> <p>In the LIO case, metadata, Standard NRVIS Interface Format (SNIF) packages and technical reports are loaded into the GIA with the data.</p> <p>CCRS includes EO mission specific elements as described in the LTDP/Preserved Data Set Content document.</p> <p>[Note: In the case of CCRS, private sector entities provide derived EO products and CCRS does not have access to these software and algorithms used to generate the products. New agreements should be in place to ensure these are transferred to enable preservation.]</p>

Common Challenge	Solutions
Dynamic digital geospatial data changes.	<p>When data are set aside for preservation they must be fixed. Bounded variability means that the content of the record (e.g., the data to create the map and the algorithm used to render it) needs to be fixed. Furthermore, the documentary form of the map (e.g., specifications and the software used to create and view the interactive map) should be immutable to ensure that its presentation remains the same across time.</p> <p>Fixed rules need to be established for the selection of content and form, which allows for a stable range of variability in the interactive map or model.</p> <p>When data are not yet fixed, they should still be subject to stringent data management controls, as in the case of DFO's floating buoy data.</p>
Determining frequency of archiving digital geospatial data.	<p>The decisions that drive this are based on user requirements and the business practices of the creator and the preserver. Organizations will have to determine this in collaboration with users, by following the instructions of data creators and by examining the business value of the records in question.</p> <p>In the case of CCRS, all the EO mission data are preserved, and in some cases distinctions are made between essential and non-essential data.</p> <p>LIO chose to capture a yearly snapshot of the Warehouse and retired records deemed official.</p> <p>The North Carolina NGDIA survey is a useful tool in this context.</p>
Rapid technology changes.	<p>The frameworks, guidelines and tools address this issue in the preservation planning, maintenance and access phases of the preservation process. The <i>OAIS Reference Model</i> and the <i>TRAC Checklist</i> in particular discuss status monitoring, technological transition plans, and risk management. The <i>InterPARES Creator and Preserver Guidelines</i> provide useful recommendations regarding manuals and the keeping of software copies.</p> <p>The CCRS EODMS system for instance is being designed to include a development, test and production environment to enable the transition of new technologies without disrupting the system. Software choices are important and the more open and standardized they are the easier it will be to manage them. Long-term sustainable funding is also required.</p> <p>The <i>InterPARES 2 Creator Guidelines</i> also suggest that creators take this into consideration.</p>
Distributed digital data archiving.	<p>The key recommendation is the identification of all data owners at the time of preservation and the development of agreed upon preservation strategies among all of them. The concept of bounded variability applies here.</p>

Common Challenge	Solutions
<p>Privacy, confidentiality and use restriction considerations.</p>	<p>Legislation will guide the decision-making regarding privacy and confidentiality while use restrictions may be addressed in MOUs, licences and agreements. This will be identified during the appraisal phase and this documentation forms part of the submission package. The preservation system must carry forward all use rights and ensure these are addressed at access.</p> <p>For CCRS there are mission specific restrictions, for LIO these are addressed in the Warehouse and in IPY this is associated with the EULAs.</p>
<p>Acquiring the necessary resources.</p>	<p>Sustainable financial resources are required to maintain a preservation system, which includes skilled human resources, technology, and physical space. Clear cost analysis and business plans are required and the GeoMAPP tools are helpful in this regard. The <i>TRAC Checklist</i> emphasizes revising plans on a yearly basis.</p> <p>The IPY project suffered as a result of a lack of sustainable resources and the CCRS project is being developed incrementally and will have to secure long-term dedicated funding in the future.</p>
<p>Shortages of archival and preservation skills and experience.</p>	<p>Ongoing education and training of personnel is a requirement in the <i>OAIS Reference Model</i> and the <i>TRAC Checklist</i>. Knowledge transfer is also required.</p> <p>CCRS will be implementing a training program.</p> <p>The IPY identified this as an issue as scientists did not have the requisite skills to manage their records and recommended that data management be a key educational component for all researchers.</p> <p>The ISDM relies on a distributed network of curators under centralized control.</p> <p>Also, many government institutions have access to skilled librarians and archivists, and they should be invited to help design preservation systems and to develop procedures and policies.</p>
<p>When to start the preservation process.</p>	<p>Preservation activities begin before records are created and continue throughout a data set's life cycle.</p> <p>The IPY project recommends the creation of a data management plan combined with incentives for researchers to properly keep their records and to find suitable archives for records deposit.</p> <p>GIA is part of the LIO Warehouse record keeping process, which carries through to preservation.</p> <p>CCRS includes preservation as part of their normal business practices in collecting EO mission data.</p> <p>TBS directives and policies also mandate that records be kept and managed throughout their life-cycle.</p>

Common Challenge	Solutions
Where to store the geospatial data.	<p>At the time of writing, there are few digital data archives.</p> <p>Once the CCRS EODMS is online this will provide a centralized Government of Canada EO archive with the potential to grow and ingest other data sets.</p> <p>The ISDM provides some records preservation and it uses the World Data Centre for some of its data.</p> <p>Beyond these, there are few digital repositories that can ingest geospatial data and these need to be built if there is an expectation that data will be preserved.</p>
Risk preparedness and security.	<p>The <i>TRAC Checklist</i> advises the creation of an emergency preparedness plan in the event of human (e.g., fires, chemical spills) or natural (e.g., floods, earthquakes) disasters or political upheavals. This includes having offsite storage but also being able to demonstrate that the operating and the back-up systems can resume operation after catastrophic events. This also means being able to access the buildings where the preservation system resides, and all necessary documentation to get started again.</p> <p>Adherence to <i>ISO/IEC 17799:2005 Information technology – Security techniques – Code of practice for information security management</i> is a recommended practice.</p>
Roles and responsibilities.	<p>These need to be explicitly and unambiguously defined in policies and procedures documents. This means having a clear chain of command, a governance structure and an organizational chart that outlines responsible institutions, as well as clear and explicit roles and terms of references.</p>

## 8. Conclusions

Establishing archives and developing the associated preservation systems is a complex undertaking, requiring a thorough understanding of basic archival and preservation concepts, a solid business plan and the long-term commitment of the archives host and data contributing organizations. This primer seeks to introduce readers to those concepts and provide a high level overview of the basic steps involved in creating a digital geospatial data preservation system. The ideas in this document have been informed by a review of the literature on digital data archives and preservation and by good practices and lessons learned from some early adopters of digital data preservation approaches and methods.

While preserving data may be optional for some data creators, government organizations are often mandated to do so. Readers are cautioned to take into consideration Government of Canada legislation (and in particular the *Library and Archives of Canada Act*), regulations, policies and directives that impact records preservation when evaluating whether or not to establish a geospatial data archives. The primer highlights important obligations, limitations and challenges associated with key GoC legislation and policy that will influence data preservation decisions.

The primer draws heavily on the guidance provided by five particularly important documents: the *OAIS Reference Model*, *LTDP Guidelines*, *InterPARES 2 Creator and Preserver Guidelines*, and *TRAC Checklist*. These documents follow many of the same approaches, methods and models, use similar or identical terminology and are widely accepted within the archival community as international best practices. Another excellent set of guidance documents that is referenced frequently, generated by the Geospatial Multistate Archive and Preservation Project (GeoMAPP), is particularly relevant for those wishing to preserve digital geospatial records.

At the time of publication, there was limited experience with digital geospatial data preservation in Canada. A number of factors have contributed to the paucity of digital geospatial archives examples, including: resource constraints, limited user demand, lack of digital data preservation knowledge and skills, and uncertainties about the optimum frequency of accessioning data to a preservation system. However, four examples of digital data preservation efforts were examined (CCRS's planned Earth Observation Data Management System, Ontario MNR's Geographic Information Archive, DFO's Integrated Science Data Management system and the data preservation efforts associated with the International Polar Year), their experiences are described and their good practices are highlighted.

Many geospatial data creators are actively considering data preservation initiatives, and this primer has been developed to provide guidance in those efforts. Readers are encouraged to share their good preservation practices and lessons learned with GeoConnections, so that the primer can be improved, to help the geospatial information community increase its effectiveness in addressing this challenging area of practice.

## A. References

- Ball, A. (2012). *Review of Data Management Lifecycle Models*. Retrieved from OPUS: University of Bath Online Publication Store: <http://opus.bath.ac.uk/28587/>
- Bleakly, D. (2002). *Long-Term Spatial Data Preservation and Archiving: What are the issues?* Retrieved February 2, 2013 from: <http://prod.sandia.gov/techlib/access-control.cgi/2002/020107.pdf>
- CCRS. (2013). Canadian Earth Observation Data Access and Services (CEODAS) Version 3.0. Ottawa: Natural Resources Canada.
- CCSDS. (2012). *Reference Model for an Open Archival Information System (OAIS)*. Retrieved January 2, 2012, from Consultative Committee for Space Data Systems Reference Models: <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- CIESEN. (2005). *Data Model for Managing and Preserving Geospatial Electronic Records*. Retrieved February 11, 2011 from Centre for International Earth Science Information Network: Geospatial Electronic Records: [http://www.ciesin.columbia.edu/ger/DataModelV1\\_20050620.pdf](http://www.ciesin.columbia.edu/ger/DataModelV1_20050620.pdf)
- CPDN. (2012). *CPDN Purpose*. Retrieved January 4, 2013, from Canadian Polar Data Network: [http://polardatanetwork.ca/?page\\_id=59](http://polardatanetwork.ca/?page_id=59)
- DCC. (2013). *Digital Curation Centre Lifecycle Model*. Retrieved 2013, from DCC: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
- DPC. (2008). *Preservation Management of Digital Materials: The Handbook*. Retrieved January 2, 2013, from Digital Preservation Coalition: <http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>
- Department of Justice. (1985). *Access to Information Act (R.S.C., 1985, c. A-1)*. Retrieved March 11, 2013, from Canada Justice Laws Website: <http://laws-lois.justice.gc.ca/eng/acts/A-1/FullText.html>
- Department of Justice. (2004). *An Act to establish the Library and Archives of Canada, to amend the Copyright Act and to amend certain Acts in consequence (S.C. 2004, c. 11)*. Retrieved March 11, 2013, from Canada Justice Laws Website: <http://laws-lois.justice.gc.ca/eng/acts/L-7.7/FullText.html>

- Department of Justice. (1985). *Canada Evidence Act (R.S.C., 1985, c. C-5)*. Retrieved March 11, 2013, from Canada Justice Laws Website: <http://laws-lois.justice.gc.ca/eng/acts/C-5/FullText.html>
- Department of Justice. (1985). *Copyright Act (R.S.C., 1985, c. C-42)*. Retrieved March 11, 2013, from Canada Justice Laws Website: <http://laws-lois.justice.gc.ca/eng/acts/C-42/FullText.html>
- Department of Justice. (2000). *Personal Information Protection and Electronic Documents Act (S.C. 2000, c. 5)*. Retrieved March 11, 2013, from Canada Justice Laws Website: <http://laws-lois.justice.gc.ca/eng/acts/P-8.6/FullText.html>
- Department of Justice. (1985). *Privacy Act (R.S.C., 1985, c. P-21)*. Retrieved March 11, 2013, from Canada Justice Laws Website: <http://laws-lois.justice.gc.ca/eng/acts/P-21/FullText.html>
- DFO. (2001). *Management Policy for Scientific Data*. Retrieved January 3, 2013, from Department of Fisheries and Oceans: <http://www.dfo-mpo.gc.ca/science/data-donnees/policy-politique-eng.htm>
- DFO. (2012). *Integrated Science Data Management*. Retrieved January 4, 2013, from Department of Fisheries and Oceans: <http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/index-eng.html>
- DFO. (2012). *Responsible National Oceanographic Data Centre*. Retrieved January 23, 2013, from Fisheries and Oceans Canada: <http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/rnodc-cnrdc/index-eng.htm#concept>
- DFO. (2013). *Drifting Buoys*. Retrieved January 23, 2013, from Fisheries and Oceans Canada: <http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/drib-bder/index-eng.htm>
- Erwin, T., & Sweetking-Singer, J. (2009). *The National Geospatial Digital Archive: A Collaborative Project to Archive Geospatial Data*. Retrieved March 11, 2013, from Journal of Map And Geography Libraries: <http://dx.doi.org/10.1080/15420350903432440>
- GeoConnections. (2008). *The dissemination of government geographic data in Canada: guide to best practices*. Retrieved March 11, 2013, from GeoConnections Publications: [http://geoscan.ess.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/geoscanfastlink\\_e.web&search1=R=288853](http://geoscan.ess.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/geoscanfastlink_e.web&search1=R=288853)
- GeoConnections. (2013). *Operational Policies*. Retrieved February 13, 2013 from GeoConnections Publications: <http://geoconnections.nrcan.gc.ca/18>
- GeoMAPP. (2009). *System Inventory Template*. Retrieved January 21, 2013, from Geospatial Multistate Archive and Preservation Partnership Publications and Tools: [http://www.geomapp.net/docs/GeoMAPP\\_System\\_Inventory\\_Template.pdf](http://www.geomapp.net/docs/GeoMAPP_System_Inventory_Template.pdf)

- GeoMAPP. (2010). *Geoarchiving Self-Assessment*. Retrieved January 21, 2013, from Geospatial Multistate Archive and Preservation Partnership Publications and Tools: [http://www.geomapp.net/publications\\_categories.htm#assess](http://www.geomapp.net/publications_categories.htm#assess)
- GeoMAPP. (2011). *Archival Metadata Elements for the Preservation of Geospatial Datasets*. Retrieved February 2, 2013 from Geospatial Multistate Archive and Preservation Partnership (GeoMapp) Publications and Tools: [http://www.geomapp.net/docs/GIS\\_OAIS\\_Archival\\_Metadata\\_v1.0\\_final\\_20110921.pdf](http://www.geomapp.net/docs/GIS_OAIS_Archival_Metadata_v1.0_final_20110921.pdf)
- GeoMAPP. (2011). *Best Practices for Geospatial Data Transfer for Digital Preservation*. Retrieved from Geospatial Multistate Archive and Preservation Partnership: Publications and Tools: [http://www.geomapp.net/docs/Geo\\_Data\\_Transfer\\_BestPractices\\_v1.0\\_final\\_20111201.pdf](http://www.geomapp.net/docs/Geo_Data_Transfer_BestPractices_v1.0_final_20111201.pdf)
- GeoMAPP. (2011). *Geoarchiving Business Planning Toolkit*. Retrieved from GeoMAPP: Publications and Tools: Geospatial Archival Business Planning : [www.geomapp.net/docs/00\\_Geoarchiving\\_Business\\_Toolkit\\_20111231.zip](http://www.geomapp.net/docs/00_Geoarchiving_Business_Toolkit_20111231.zip)
- GeoMAPP. (2011). *New Partner Mentoring: Introduction to Archival Appraisal*. Retrieved from Geospatial Multistate Archive and Preservation Partnership: [http://www.geomapp.net/docs/Appraisal\\_mentoring\\_presentation\\_final\\_20110610.pdf](http://www.geomapp.net/docs/Appraisal_mentoring_presentation_final_20110610.pdf)
- GeoMAPP. (2011). *Publications and Tools*. Retrieved February 18, 2013, from Geospatial Multistate Archive and Preservation Partnership: [http://www.geomapp.net/publications\\_categories.htm](http://www.geomapp.net/publications_categories.htm)
- GeoMAPP. (2011a). *Geoarchiving Business Planning Guidebook*. Retrieved January 18, 2013, from Geospatial Multistate Archive and Preservation Partnership Publications and Tools: [http://www.geomapp.net/publications\\_categories.htm](http://www.geomapp.net/publications_categories.htm)
- GeoMAPP. (2011b). *Best Practices for Archival Processing for Geospatial Datasets*. Retrieved January 18, 2013, from Geospatial Multistate Archive and Preservation Partnership Publications and Tools: [http://www.geomapp.net/publications\\_categories.htm](http://www.geomapp.net/publications_categories.htm)
- GeoMAPP. (2011c). *Geospatial Multistate Archive and Preservation Partnership Publications and Tools*. Retrieved January 21, 2013, from [http://www.geomapp.net/docs/GIS\\_Archival\\_Processing\\_Process\\_v1.0\\_final\\_20111102.pdf](http://www.geomapp.net/docs/GIS_Archival_Processing_Process_v1.0_final_20111102.pdf)
- GeoMAPP. (2011d). *GeoMAPP Key Findings and Best Practices*. Retrieved January 21, 2013, from Geospatial Multistate Archive and Preservation Partnership Project Documents: [http://www.geomapp.net/docs/GeoMAPP\\_ProjectFindings\\_BestPractices20111231.pdf](http://www.geomapp.net/docs/GeoMAPP_ProjectFindings_BestPractices20111231.pdf)
- GeoMAPP. (2012). *Geoarchiving Business Cost-Benefit Analysis Guidance*. Retrieved January 21, 2013, from Geospatial Multistate Archive and Preservation Partnership Publications and Tools: [http://www.geomapp.net/docs/03\\_Geoarchiving\\_Cost-Benefit\\_Analysis\\_Guidance\\_20111231.pdf](http://www.geomapp.net/docs/03_Geoarchiving_Cost-Benefit_Analysis_Guidance_20111231.pdf)

- Government of NWT. (1993). *NWT Archives Policy*. Retrieved March 11, 2013, from Government of the North West Territories:  
<http://www.pwnhc.ca/programs/downloads/NWTArchivesPolicy.pdf>
- HAL. (2011). Hickling, Arthur and Low (HAL) *Research and Analysis Report: Geospatial Data Archiving and Preservation*. Produced under contract for GeoConnections Science & Technology Policy Research and Analysis Resource Team, Natural Resources Canada. Ottawa, Ontario, Canada: Natural Resources Canada: GeoConnections.
- Hoebelheinrich, N. (2009). *An Investigation Into Metadata for Long-Lived Geospatial Data Formats*. Retrieved February 11, 2013 from Library of Congress: National Geospatial Digital Archive Project:  
[www.digitalpreservation.gov/meetings/documents/ndiipp08/session7\\_hoebelheinrich\\_paper.doc](http://www.digitalpreservation.gov/meetings/documents/ndiipp08/session7_hoebelheinrich_paper.doc)
- Hoebelheinrich, N., & Munn, N. K. (2009). *Assessing the Utility of Current Format Registry Efforts for Geospatial Formats*. Retrieved March 11, 2013, from The National Geospatial Digital Archive, Archiving 2009: Final Program and Proceedings. The National Geospatial Digital Archive: [http://www.ngda.org/docs/Pub\\_Hoebelheinrich\\_Arch09\\_09.pdf](http://www.ngda.org/docs/Pub_Hoebelheinrich_Arch09_09.pdf)
- InterPARES 2. (2002). *Requirements for Assessing and Maintaining the Authenticity of Electronic Records*. Retrieved February 18, 2013, from InterPARES 2 Project:  
[www.interpares.org/book/interpares\\_book\\_k\\_app02.pdf](http://www.interpares.org/book/interpares_book_k_app02.pdf)
- InterPARES 2. (2007). *Creator Guidelines: Making and Maintaining Digital Materials: Guidelines for Individuals*. Retrieved February 10, 2013 from InterPARES 2:  
[http://www.interpares.org/public\\_documents/ip2%28pub%29creator\\_guidelines\\_booklet.pdf](http://www.interpares.org/public_documents/ip2%28pub%29creator_guidelines_booklet.pdf)
- InterPARES 2. (2007). *Preserver Guidelines - Preserving Digital Records: Guidelines for Organizations*. Retrieved February 10, 2013 from InterPARES 2 Project Products:  
[http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2%28pub%29preserver\\_guidelines\\_booklet.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2%28pub%29preserver_guidelines_booklet.pdf)
- InterPARES 2. (2008). *Policy Framework - A framework of principles for the development of policies, strategies and standards for the long-term preservation of digital records*. Retrieved February 18, 2013, from InterPARES 2 Project:  
[www.interpares.org/ip2/display\\_file.cfm?doc=ip2\(pub\)policy\\_framework\\_document.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)policy_framework_document.pdf)
- InterPARES 2. (2013). *Dictionary definitions*. Retrieved January 24, 2013 from InterPARES 2 Terminology Database: [http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm)
- IPY. (2010). *About IPY*. Retrieved January 3, 2013, from International Polar Year:  
<http://ipy.arcticportal.org/about-ipy>

- IPY Data Management Sub-committee. (2008). *International Polar Year 2007-2008 Data Policy*. Retrieved January 4, 2013, from International Polar Year:  
[http://classic.ipy.org/Subcommittees/final\\_ipy\\_data\\_policy.pdf](http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf)
- ISO. (2003). *ISO 19115:2003 Geographic information -- Metadata*. Retrieved February 11, 2013 from International Standards Organization:  
[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=26020](http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020)
- ISO. (2005). *ISO/IEC 17799:2005 Information technology -- Security techniques -- Code of practice for information security management*. Retrieved from International Standards Organization:  
[http://www.iso.org/iso/catalogue\\_detail?csnumber=39612](http://www.iso.org/iso/catalogue_detail?csnumber=39612)
- ISO. (2005). *ISO 19128:2005 Geographic information - Web map server interface*. Retrieved March 10, 2013 from International Standards Organization:  
[http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=32546](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=32546)
- ISO. (2012). *ISO 16363:2012, Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*. Retrieved February 10, 2013 from International Standards Organization:  
[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510)
- JC. (2007). *Department of Justice, Justice Laws Website*. Retrieved February 18, 2013 from Charts and Nautical Publications Regulations, 1995 (SOR/95-149):  
<http://laws-lois.justice.gc.ca/eng/regulations/SOR-95-149/>
- JC. (2007). *Justice Laws Website*. Retrieved February 18, 2013, from Remote Sensing Space Systems Act, S.C. 2005, c. 45: <http://laws-lois.justice.gc.ca/eng/acts/R-5.4/FullText.html>
- JC. (2007). *Justice Laws Website*. Retrieved February 18, 2103, from Remote Sensing Space Systems Regulations, SOR/2007-66:  
<http://laws-lois.justice.gc.ca/eng/regulations/SOR-2007-66/FullText.html>
- JC. (2010). *Department of Justice, Justice Laws Website*. Retrieved February 18, 2013 from Arctic Waters Pollution Prevention Act (R.S., 1985, c. A-12):  
<http://laws-lois.justice.gc.ca/eng/acts/A-12/>
- LAC. (2006). *Library and Archives Canada, Information Management*. Retrieved February 18, 2013, from Records and Information Life Cycle Management:  
<http://www.collectionscanada.gc.ca/gouvernement/produits-services/007002-2012-e.html>
- LAC. (2009). *Setting Retention Specifications for Information Resources of Business Value*. Ottawa, Ontario, Canada: LAC.
- LAC. (2011). *Library and Archives Canada*. (L.A. Canada, Producer) Retrieved February 18, 2013, from Information Management (IM) - Library and Archives Canada:  
<http://www.collectionscanada.gc.ca/government/produits-services/007002-6000-e.html>

- LAC. (2012). *Library and Archives Canada*. Retrieved February 18, 2013, from Disposition and Recordkeeping Program: <http://www.bac-lac.gc.ca/eng/disposition-recordkeeping-program/Pages/disposition-recordkeeping-program.aspx>
- Lauriault, T. P., Craig, B., Pulsifer, P. L., & Taylor, D. R. (2008). Today's Data are Part of Tomorrow's Research: Tomorrow's Research: Archival Issues in the Sciences. *Archivaria*, 123-179.
- Library of Congress. (2013). *PREMIS Data Dictionary for Preservation Metadata*. Retrieved February 11, 2013 from Library of Congress: Standards: <http://www.loc.gov/standards/premis/>
- Library of Congress. (2013). *Sustainability of Digital Formats Planning for Library of Congress Collections*. Retrieved March 11, 2013, from Library of Congress Digital Preservation: <http://www.digitalpreservation.gov/formats/index.shtml>
- LTDP. (2012). *Long Term Data Preservation Earth Observation Preserved Data Set Content LTDP/PDSC*. Retrieved from GCSB Introduction: [http://earth.esa.int/gscb/ldp/LTDP\\_PDSC\\_4.0.pdf](http://earth.esa.int/gscb/ldp/LTDP_PDSC_4.0.pdf)
- LTDP Working Group. (2012). *Long Term Preservation of Earth Observation Space Data: European LTDP Common Guidelines*. Retrieved January 2, 2013, from Ground Segment Coordination Body (GSCB): [http://earth.esa.int/gscb/ldp/EuropeanLTDPCommonGuidelines\\_Issue2.0.pdf](http://earth.esa.int/gscb/ldp/EuropeanLTDPCommonGuidelines_Issue2.0.pdf)
- MacNeil, H. (2000). Providing Grounds for Trust: Developing Conceptual Requirements for the Long-Term Preservation of Authentic Term Preservation of Authentic. *Archivaria*, 52-78.
- NC CGIA. (2006). *Survey Frequency of Geospatial Data Capture*. Retrieved March 11, 2013, from Library of Congress North Carolina Geospatial Data Archiving Project: <http://www.digitalpreservation.gov/partners/ncgdap.html>
- NGDA. (2009). *Collection Development Policies and Contracts*. Retrieved March 11, 2013, from US National Geospatial Digital Archive: <http://www.ngda.org/policies.html>
- NRC. (2010). *DataCite Canada*. Retrieved March 11, 2013, from National Research Council of Canada: <http://cisti-icist.nrc-cnrc.gc.ca/eng/services/cisti/datacite-canada/about.html>
- NRCan. (2010). *Natural Resources Canada*. Retrieved February 18, 2013, from Acts and Regulations: <http://www.nrcan.gc.ca/acts-regulations/332>
- NSIDC. (2013). *IPY at NSIDC*. Retrieved January 4, 2013, from National Snow & Ice Data Centre: <http://nsidc.org/ipy/>

- OCLC and CRL. (2007). *Trustworthy Repositories Audit & Certification: Criteria and Checklist, Version 1.0*. Retrieved February 18, 2013 from Centre for Research Libraries, Metrics for Repository Assessment: [www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf)
- OCUL. (2011). *Trusted Digital Repository*. Retrieved from Ontario Council of University Libraries: <http://www.ocul.on.ca/node/97>
- OCUL. (2012). *Trusted Digital Repositories Document Checklists*. Retrieved from Scholars Portal: <http://spotdocs.scholarsportal.info/display/OAIS/Document+Checklist>
- OMNR. (2010). Ontario Ministry of Natural Resources. *Guideline for Retiring and Retaining Geospatial Data Stored in the Land Information Ontario Warehouse*. Province of Ontario.
- OMNR. (2011). *Metadata Resources & Training*. Retrieved March 11, 2013, from Ontario Ministry of Natural Resources, Land Information Ontario: <http://www.mnr.gov.on.ca/en/Business/LIO/2ColumnSubPage/266883.html>
- Ontario Ministry of Government Services. (2011). *Promoting Excellence in Recordkeeping*. Retrieved March 11, 2013, from Government of Ontario: <http://www.archives.gov.on.ca/en/recordkeeping/index.aspx>
- Parsons, M., de Bruin, T., Tomlinson, S., Campbell, H., Godoy, O., & LeClert, J. (2011). *Chapter 3.11 The State of Polar Data: The IPY Experience, PART THREE: IPY Observing Systems, Their Legacy and Data Management*. Retrieved January 4, 2013, from International Council for Science: <http://www.icsu.org/publications/reports-and-reviews/ipy-summary/ipy-jc-summary-part3.pdf>
- Prince of Wales Northern Heritage College. (2013). *Records Retention and Disposition Schedules by Department*. Retrieved March 11, 2013 from Prince of Wales Northern Heritage College: [http://www.pwnhc.ca/programs/archives/record\\_schedules/Departments/record\\_schedules.asp](http://www.pwnhc.ca/programs/archives/record_schedules/Departments/record_schedules.asp)
- QA4EO. (2013). *Documentation*. Retrieved February 18, 2013, from Quality Assurance Framework for Earth Observation: <http://qa4eo.org/documentation.html>
- RLG and OCLC. (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Retrieved February 18, 2013 from Attributes of Trusted Digital Repositories: <http://www.oclc.org/research/activities/trustedrep.html>
- RLG and OCLC. (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. RLG-OCLC.

- Roeder, J., Eppard, P., Underwood, B., & Lauriault, T. P. (2008). *Part 3: Authenticity, Reliability and Accuracy of Digital Records in the Artistic, Scientific and Government Sectors, Domain 2 Task Force Report*. Retrieved February 18, 2013, from International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records Book:  
[http://www.interpares.org/display\\_file.cfm?doc=ip2\\_book\\_part\\_3\\_domain2\\_task\\_force.pdf](http://www.interpares.org/display_file.cfm?doc=ip2_book_part_3_domain2_task_force.pdf)
- SAA. (2013). *Glossary of Archival Records and Terminology*. Retrieved February 10, 2013 from The Society of American Archivists: [www2.archivists.org/glossary/terms/a](http://www2.archivists.org/glossary/terms/a)
- Scassa, T. (2011). *Final Report: Review of IP Law and Instruments (Copyright, Licensing) in the Context of Geospatial Data*, . Ottawa: Hickling Arthurs Low Corporation for Natural Resources Canada.
- Supreme Court of Canada. (2004). *CCH Canadian Ltd. v. Law Society of Upper Canada ([2004] 1 S.C.R. 339)*. Retrieved March 11, 2013, from Supreme Court of Canada Website: <http://scc.lexum.org/decisia-scc-csc/scc-csc/scc-csc/en/item/2125/index.do>
- Stone, A., & Day, M. (1999). *Cedars Preservation Metadata Elements: Cedars Project Document AIW02*. Retrieved February 11, 2013 from Cedars Access Issues Working Group: <http://www.ukoln.ac.uk/metadata/cedars/papers/aiw02/>
- TBS. (2012). *Standard on Geospatial Data*. Retrieved February 11, 2013 from Treasury Board Secretariat of Canada Policy Suite: <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16553&section=text>

## B. Glossary of Terms

### Glossary Sources:

- Digital Curation Centre (DCC) Glossary: <http://www.dcc.ac.uk/digital-curation/glossary>
- Digital Preservation Coalition (DPC): <http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>
- Geospatial Multistate Archiving and Preservation Project (GeoMAPP)
- InterPARES 2 Project (IP2) Terminology Database: [http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm)
- Minnesota State Archives (MSA): <http://www.mnhs.org/preserve/records/electronicrecords/erglossary.html#c>
- Open Archival Information System (OAIS): <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- Quality Assurance Framework for Earth Observation (QA4EO): <http://qa4eo.org>

Acronym	Term	Definition
	Access	Continued, on-going usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes for which the digital material was created and, or acquired. (DPC)
	Accessibility	The availability and usability of information. (IP2)
	Accession	To take legal and physical custody of a body of records and to document it in a register. (IP2)
	Accuracy	The degree to which data, information, documents or records are precise, correct, truthful, free of error or distortion, or pertinent to the matter. (IP2)
	Active Record	A record needed by the creator for the purpose of carrying out the action for which it was created or for frequent reference. (IP2)
	Appraisal	The process of assessing the value of records for the purpose of determining the length and conditions of their preservation. (IP2)
AIP	Archival Information Package	An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS. (OAIS)
	Archives	An agency or institution responsible for the preservation and communication of records selected for permanent preservation. (IP2)

Acronym	Term	Definition
	Associated Description	The information describing the content of an Information Package from the point of view of the software program or document that allows Consumers to locate, analyze, order or retrieve information from an Archives. (OAIS)
	Authenticity	The digital material is what it purports to be; refers to the trustworthiness of the electronic record as a record or to the fact that whatever is being cited is the same as it was when it was first created unless the accompanying metadata indicates any changes. (IP2)
	Benchmark Authenticity Requirements	The conditions that serve as a basis for the preserver's assessment of the authenticity of a creator's digital records during appraisal. (IP2)
	Bounded Variability	The changes to the form and/or content of a digital record that are limited and controlled by fixed rules, so that the same query, request or interaction always generates the same result. (IP2)
	Checksum	A count of the number of bits in a transmission unit that is included with the unit so that the receiver can check to see whether the same number of bits arrived. If the counts match, it is assumed that the complete transmission was received. (MSA)
	Consumer	The role played by those persons or client systems that interact with OAIS services to find preserved information of interest and to access that information in detail. (OAIS)
	Content Information	A set of information that is the original target of preservation or that includes part or all of that information. It is an Information Object composed of its Content Data Object and its Representation Information. (OAIS)
CRC	Cyclic Redundancy Check	An error-detecting code commonly used in digital networks and storage devices to detect accidental changes to raw data. A short check value is attached to data entering these systems, based on the remainder of a polynomial division of their contents; on retrieval the calculation is repeated, and corrective action can be taken against presumed data corruption if the check values do not match. (OAIS)
	Data Object	A digital object composed of a set of bit sequences. (OAIS)
	Descriptive Information	The set of information, consisting primarily of Package Descriptions, which is provided to Data Management to support the finding, ordering, and retrieving of OAIS information holdings by Consumers. (OAIS)

Acronym	Term	Definition
	Designated User Community	An identified group of potential consumers who should be able to understand a particular set of information. The designated community may be composed of multiple user communities. (OAIS)
	Digital Components	A digital object that is part of one or more digital documents, and the metadata necessary to order, structure or manifest its content and form, requiring a given preservation action. (IP2)
	Digital Rights Management	A collection of systems used to protect the copyrights of electronic media. (TechTerms.com)
DIP	Dissemination Information Package	An Information Package, derived from one or more AIPs, and sent by Archive to the Consumer in response to a request to the OAIS. (OAIS)
	Emulation	A means of overcoming technological obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers. (DPC)
	Finding Aid	A type of Access aid that allows a user to search for and identify Archival Information Packages of interest. (OAIS)
	Fixed Form	The quality of a record that ensures its content remains complete and unaltered. (IP2)
	Form	Rules of representation that determine the appearance of an entity and convey its meaning. (IP2)
	Geoarchiving	The identification of significant geospatial data and its preservation for future use. (GeoMAPP)
	Information Object	A Data Object together with its Representation Information. (OAIS)
	Information Package	A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information. (OAIS)
	Knowledge Base	A set of information, incorporated by a person or system, that allows that person or system to understand received information (e.g., a person whose Knowledge Base includes an understanding of English will be able to read and understand an English text). (DCC)

Acronym	Term	Definition
	Logical Format	The organized arrangement of data on a digital medium that ensures file and data control structures are recognizable and recoverable by the host computer operating system. Two common logical formats for files and directories are ISO 9660/13490 for CDs, and Universal Disk Format (UDF) for DVDs. (IP2)
	Long-Term Preservation	Continued access to digital materials, or at least to the information contained in them, indefinitely. (DPC)
	Management	The role played by those who set overall archiving policy as one component in a broader policy domain, for example as part of a larger organization. (OAIS)
	Medium-Term Preservation	Continued access to digital materials beyond changes in technology for a defined period of time but not indefinitely. (DPC)
	Metadata	Information describing significant aspects of a resource that are required to successfully manage and preserve digital materials over time and that will assist in ensuring essential contextual, historical, and technical information are preserved along with the digital resource. (DPC)
	Migration	A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. (DPC)
OAIS	Open Archival Information System	An Archive, consisting of an organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. (OAIS)
	Order Agreement	An agreement between the Archive and the Consumer in which the physical details of the delivery, such as media type and format of data, are specified. (DCC)
	Package Description	The information intended for use by Access Aids. (OAIS)
	Packaging Information	The information that is used to bind and identify the components of an Information Package. For example, it may be the ISO 9660 volume and directory information used on a CD-ROM to provide the content of several files containing Content Information and Preservation Description Information. (DCC)
PDI	Preservation Description Information	The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, Context, and Access Rights Information. (OAIS)
	Producer	The role played by those persons or client systems that provide the information to be preserved. (OAIS)

Acronym	Term	Definition
QI	Quality Indicator	A measure that provides sufficient information to allow all users to readily evaluate a data products' suitability for their particular application (i.e. its "fitness for purpose"). A QI should be based on a quantifiable assessment of evidence demonstrating the level of traceability to internationally community agreed (where possible SI) reference standards. (QA4EO)
	Record	A document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference. (IP2)
	Records Preservation System	A set of rules governing the permanent intellectual and physical maintenance of acquired records and the tools and mechanisms used to implement these rules. (IP2)
	Reformatting	Copying information content from one storage medium to a different storage medium (media reformatting) or converting from one file format to a different file format (file re-formatting). (DPC)
	Refreshing	Copying information content from one storage media to the same storage media. (DPC)
	Reliability	The trustworthiness of a record as a statement of fact. It exists when a record can stand for the fact it is about, and is established by examining the completeness of the record's form and the amount of control exercised on the process of its creation. (IP2)
	Representation Information	The information that maps a Data Object into more meaningful concepts. An example is JPEG software, which embodies an understanding of the JPEG standard and maps the bits into pixels that can then be rendered as an image for human viewing. (OAIS)
	Short-Term Preservation	Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and, or until it becomes inaccessible because of changes in technology. (DPC)
	Submission Agreement	The agreement reached between an Archive and the Producer that specifies a data model, and any other arrangements needed for data submission. This data model identifies format/contents and the logical constructs used by the Producer and how they are represented on each media delivery or in a telecommunication session. (OAIS)
SIP	Submission Information Package	An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information. (OAIS)

Acronym	Term	Definition
	Technology-independent Authentication	The authentication of records based on the use of administrative procedures to establish a presumption of authenticity or, if necessary, a verification of authenticity, especially through comparison of the evidence compiled about a record's identity and integrity and the procedural controls exercised over its creation, use, maintenance and/or preservation with the requirements for authentic records. (IP2)
	Trusted Custodian	A preserver who can demonstrate that it has no reason to alter the preserved records or allow others to alter them and is capable of implementing all of the requirements for the preservation of authentic copies of records. (IP2)
	Trusted Preservation System	The whole of the rules that control the preservation and use of the records of the creator and provide a circumstantial probability of the authenticity of the records, and the tools and mechanisms used to implement those rules. (IP2)

## C. Acts and Regulations Specific to the Preservation and Management of Government Information

Legislation	Obligations	Limitations/Challenges
<b>Library and Archives of Canada Act</b> (2004, c. 11)	<ul style="list-style-type: none"> <li>▪ Manage data for possible accession</li> <li>▪ No data disposal without written consent</li> <li>▪ Submit data collected for the purposes of public opinion research</li> <li>▪ Legal deposit of two copies of all publications</li> <li>▪ Deposit of data considered to be of documentary heritage</li> </ul>	<ul style="list-style-type: none"> <li>▪ How dynamically changing data sets are treated as records</li> <li>▪ Treatment of data sets with contributions of multiple producers</li> <li>▪ Treatment of dynamic/interactive digital maps created with web services</li> <li>▪ Distributed archiving mechanisms and controls</li> <li>▪ Preservation of digital information accessible through web portals</li> <li>▪ Limited institutional support to distributed archives</li> </ul>
<b>Legal Deposit of Publications Regulations</b> (SOR/2006-337)	<ul style="list-style-type: none"> <li>▪ Direction on submission of digital publications</li> </ul>	<ul style="list-style-type: none"> <li>▪ The system is not conducive to the deposit of data</li> </ul>
<b>Copyright Act</b> (R.S., 1985, c. C-42)	<ul style="list-style-type: none"> <li>▪ Data ownership and use rights and responsibilities follow the data into the archives</li> </ul>	<ul style="list-style-type: none"> <li>▪ Need to manage use and access rights</li> </ul>
<b>Access to Information Act</b> (R.S., 1985, c. A-1)	<ul style="list-style-type: none"> <li>▪ Data access restrictions follow the data into the archives</li> </ul>	<ul style="list-style-type: none"> <li>▪ Need to manage use and access rights</li> </ul>
<b>Privacy Act</b> (R.S., 1985, c. P-21)	<ul style="list-style-type: none"> <li>▪ Personal information may be disclosed to LAC for archival purposes</li> </ul>	<ul style="list-style-type: none"> <li>▪ May need to aggregate and/or anonymize some data to enable access</li> </ul>
<b>Privacy Regulations</b> (SOR/83-508)	<ul style="list-style-type: none"> <li>▪ Direction on management of personal data when in the archives</li> </ul>	<ul style="list-style-type: none"> <li>▪ Time limitations on when personal information may be accessed for research or statistical purposes</li> </ul>
<b>Personal Information Protection and Electronic Documents Act</b> [2000, c. 5]	<ul style="list-style-type: none"> <li>▪ Applies to personal information collected, used and disclosed by private sector organizations.</li> </ul>	

Legislation	Obligations	Limitations/Challenges
<b>Canada Evidence Act</b> (R.S., 1985, c. C-5)	<ul style="list-style-type: none"><li>▪ Establish authenticity of archived data introduced as evidence (solid metadata, change tracking over time, properly operating systems, manage and maintain systems to ensure authenticity of the records they contain – security protocols, access protocols, etc.)</li></ul>	

## D. Information Management Policies and Directives of the Treasury Board of Canada Secretariat (TBS)

Objectives	Obligations
<b>Policy Framework for Information and Technology (2007)</b>	
<ul style="list-style-type: none"> <li>▪ Provides guiding principles to sound information and technology management practices across government</li> <li>▪ Key guiding principles:               <ul style="list-style-type: none"> <li>○ <b>Stewardship</b> – data must be rigorously managed throughout its life cycle, regardless of medium or format</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>▪ Creators have a responsibility to manage records</li> <li>▪ The legislation and regulation in Table 1 are to be a part of a records life cycle</li> <li>▪ Efficient management of information in an organization, from planning and systems development to disposal or long-term preservation</li> </ul>
<b>Policy on Information Management (IM Policy 2007)</b>	
<ul style="list-style-type: none"> <li>▪ To achieve efficient and effective IM:               <ul style="list-style-type: none"> <li>○ support program and service delivery;</li> <li>○ foster informed decision making;</li> <li>○ facilitate accountability, transparency, and collaboration; and</li> <li>○ <b>preserve</b> and <b>ensure access</b> to information and records for the benefit of present and future generations</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>▪ Geospatial data should be part of an integrated information management strategy along with other information</li> <li>▪ If geospatial data are used to support operations, policies and programs, it is important to ensure that they are maintained as part of a record set. Metadata, description of data collection methodologies, data quality parameters, contextual information and any other attributes deemed necessary by the creators and maintainers should assist with the understanding. If used in an experiment, then the parameters of the experiment should also accompany the geospatial information.</li> <li>▪ Creation of data discovery and access mechanisms</li> <li>▪ Ensuring that all data are managed to respect user 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</li> <li>▪ Ensuring departmental participation in setting government-wide direction for data and recordkeeping</li> </ul>

Objectives	Obligations
<b>Policy on Management of Information Technology (2007)</b>	
<ul style="list-style-type: none"> <li>▪ 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</li> </ul>	<ul style="list-style-type: none"> <li>▪ This policy is quite limited</li> <li>▪ The following is however recommended even though not part of the policy:                             <ul style="list-style-type: none"> <li>○ 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</li> </ul> </li> </ul>
<b>Directive on Information Management Roles and Responsibilities (2007)</b>	
<ul style="list-style-type: none"> <li>▪ Identify the roles and responsibilities of all departmental employees in supporting the deputy head in the effective management of information (data) in their department</li> </ul>	<ul style="list-style-type: none"> <li>▪ Ensure that appropriate management direction, processes and tools are in place to retain the quality of data throughout the information life cycle: planning; the collection, creation, receipt, and capture of data; their organization, use and dissemination; their maintenance, protection and preservation; their disposition; and evaluation.</li> <li>▪ Decisions on retention</li> </ul>
<b>Directive on Recordkeeping (2009)</b>	
<ul style="list-style-type: none"> <li>▪ Ensure effective recordkeeping practices that enable departments to create, acquire, capture, manage and protect the integrity of information (data) resources of business value in the delivery of GC programs and services</li> </ul>	<ul style="list-style-type: none"> <li>▪ Ensure that methodologies, mechanisms and tools are in place to support departmental recordkeeping requirements throughout the data life cycle                             <ul style="list-style-type: none"> <li>○ identify, establish, implement and maintain repositories of data of business value are stored or preserved in a physical or electronic storage space;</li> <li>○ establish, use and maintain taxonomies or classification structures to facilitate storage, search, and retrieval of data of business value in all formats;</li> <li>○ establish, implement and maintain retention periods for data of business value, as appropriate, according to format;</li> <li>○ develop and implement a documented disposition process for data; and</li> <li>○ perform regular disposition activities for all data.</li> </ul> </li> </ul>

Objectives	Obligations
<b>Standard for Electronic Documents and Records Management (EDRM) Solutions (2010)</b>	
<ul style="list-style-type: none"> <li>▪ Support efficient and effective management of information through the use of EDRM solutions and reduce overall cost through standardization and economies of scale</li> </ul>	<ul style="list-style-type: none"> <li>▪ Automated systems to:                             <ul style="list-style-type: none"> <li>○ manage, protect and preserve data from creation to disposition, ...;</li> <li>○ maintain appropriate contextual information (metadata); and</li> <li>○ enable organizations to access, use and dispose of records (i.e., their retention, destruction or transfer) in a managed, systematic and auditable way.</li> </ul> </li> </ul>
<b>Standard on Metadata (2010)</b>	
<ul style="list-style-type: none"> <li>▪ Increase the use of standardized metadata and value domains in support of the management of information resources:                             <ul style="list-style-type: none"> <li>○ improve data accessibility, sharing, authenticity, reliability, and integrity across departments; and</li> <li>○ increase the ability of programs and services to share data efficiently and effectively between systems and across departments.</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>▪ Adopt and implement metadata standards</li> </ul>
<b>Standard on Geospatial Data (2009)</b>	
<ul style="list-style-type: none"> <li>▪ Support stewardship and interoperability of data by ensuring that departments access, use and share geospatial data efficiently and effectively to support program and service delivery</li> </ul>	<ul style="list-style-type: none"> <li>▪ Implement ISO 19115 Geographic information – Metadata standard</li> <li>▪ Implement ISO 19128 Geographic information - Web map server interface standard</li> </ul>

# E. Geoarchiving Business Planning Guidebook Highlights

This guideline was developed by the Geospatial Multistate Archive and Preservation Partnership (GeoMAPP) to assist state and other users to evaluate their current situations and needs, determine the costs and benefits of a geoarchives program, and to provide the tools and operational framework for organizations to successfully design, create and implement a [geoarchiving](#) program. The Guidebook advocates an approach whereby plans are assembled through an iterative process of outreach facilitated group discussions, research, drafting, and review including participation or guidance from key stakeholders in the GIS, archival organization, and IT staff. It utilizes a list of questions to help facilitate these discussions and to generate content for each section of the plan. The following table summarizes the “Highly Recommended” elements within each section of the Guidebook.

Guidebook Section	Essence
<b>MAIN SECTION</b>	
1. Executive Summary	What outcomes are you proposing to accomplish? Why do you need to do it?
2. Geoarchives Self-Assessment	What are the current conditions and assets?
3. Customers and Stakeholders	Who is this for and who is making the case?
4. Program Goals	What are the specific ‘Programmatic Goals’ for this Business Plan? For each goal, what are the ‘Success Factors’ (or supporting objectives)?
5. Benefits and Justification	What is the primary reason ‘why’ you need to do what you are proposing? What benefits and return on investment will accrue if it is done?
6. Requirements and Costs	What is your organizational approach? What are the estimated total costs of your proposal?
7. Implementation Overview	Phasing and milestones Budget Plan
8. Measuring Success and Feedback for Recalibration	Establish cost and benefit metrics and process for regular update/review
<b>APPENDICES</b>	
A. Geoarchiving Business Planning Process Map and Checklist	An overall graphic summary of the Geoarchiving Business Planning process and a checklist to be used during the planning process
B. Comprehensive Cost Benefit Analysis Guidance Document	A guide to developing an analysis of project storage needs, overall project costs and cost benefit calculations for a Geoarchiving project

Guidebook Section	Essence
C. Geoarchiving Comprehensive Cost-Benefit Analysis Toolkit	An Excel workbook to assist the user to project storage needs, overall project costs and create cost-benefit calculations for a user-determined period of years
D. Guidance and Templates for Supporting Business Documentation	A guide to the development of Geoarchiving business cases related to geoarchiving
E. Bibliography	Geoarchiving publications and reference materials