Final Report

Good Practices in Regional-Scale Information Integration

Prepared for:

Prepared by:

HICKLING ARTHURS LOW
TECHNOLOGY MANAGEMENT, STRATEGY, AND ECONOMICS

March 28, 2008

HAL Ref: 7729

Executive Summary

The intent of the Canadian Geospatial Data Infrastructure (CGDI) is to reduce the efforts, often considerable, required by geospatial data stakeholders to access and integrate data, both within their organization and with their partners. When implemented successfully, human intervention activities in the data integration process will be facilitated as end-users and solution developers embrace CGDI-endorsed data standards, and organizations deploy these standards and the infrastructure required to make them work effectively. An important motivation for this study was that, while there is a significant opportunity to support public policy decision makers through CGDI-compliant regional atlases and decision support systems, there is limited awareness of the numerous challenges in performing regional scale information integration and the means for addressing them.

In this study, Hickling Arthurs Low conducted detailed case studies of four GeoConnections-sponsored projects, one from each of its four priority areas: Environment and Sustainable Development, Public Safety and Security, Aboriginal Matters, and Public Health. These projects were chosen as they met a set of criteria individually and collectively, including that they had engaged participants from federal, provincial, regional, municipal and community-based organizations, and they had adopted a broad range of technology platforms. Workshops were held in each of the cities of the project lead organizations to discuss case study findings and identify good practices.

Findings

The study reports findings in three areas:

- Anecdotal evidence of the broad range of data integration practices that were reported by those consulted;
- Good practices in regional scale data integration stemming from the analysis of the selected projects; and
- Data integration challenges and opportunities that were reported in each of the four priority areas.

While the original intent of this study was to identify good practices in regional scale information integration, it quickly became apparent that it would be more appropriate to determine good practices to reduce the effort related to the often difficult and time-consuming work by technology managers and systems engineers to integrate data. Such a preventative approach is consistent with the vision of CGDI that end-users should be able to access atlases and decision support systems that call up, through CGDI-endorsed services, and integrate the

required geospatial data from various systems on various platforms in various projections and scales from various organizations without the need for intervention by systems professionals.

To achieve this, good practices were identified for organizations seeking to share data:

- Develop and endorse common data standards, as data stored according to such standards is much more readily integrated;
- Prepare Geospatial Data Profiles of the datasets sought from other organizations to be better able to anticipate, and mitigate, data integration challenges once data sharing occurs;
- Establish regional CGDI data providers, which would be existing organizations with large data warehouses but with the mandate and resources to deliver trusted data and services according to CGDI-endorsed standards;
- Build Service-Oriented Architectures, so that open standards may be used to access geospatial data holdings of organizations in spite of their many different technology platforms and business processes;
- Establish data integration service centres within CGDI data provider organizations so that CGDI end-users may readily receive guidance and support when they are attempting to access and then integrate CGDI data into their own business processes;
- Undertake rapid prototype development in building atlases and decision support systems, to validate Geospatial Data Profile information and to identify data integration challenges to be addressed, as early in the process as possible.

Detailed practices are described in each of the above areas, and a Good Practices Check List was prepared and is provided. As well, a Geospatial Data Profile Tool was developed that identifies the metadata that would assist GIS technicians, systems analysts, and systems architects in their efforts to scope, design and implement a regional atlas or decision support system that delivers and/or accesses CGDI data content.

Also reported are a number of challenges and opportunities that were identified from the projects examined and that affect the success with which CGDI is being deployed in each of the four priority areas. Although up to four projects were reviewed in each priority area, and each of these projects involved multiple stakeholders from their communities of practice, it should be noted that the findings are based on evidence from a relatively small sample size.

Conclusion and Recommendations

The study concludes that there are a number of factors that would contribute to the further deployment of CGDI and they include: the public sector establishing stable funding mechanisms for regional CGDI data providers to become partners in building the CGDI; the geomatics industry adopting and enhancing CGDI-endorsed standards for data access and exchange in preference to closed proprietary methods; and communities of practice recognizing the benefits

of, and implementing the required policies, procedures and infrastructure for CGDI to work effectively.

Finally, the study makes recommendations to accelerate the delivery of trusted applications and data to end-users by enhancing the collaborative framework within which data sharing organizations operate, and that takes into consideration the very high inter-organizational dependencies that are created when organizations embrace the CGDI. The recommendations are summarized as follows:

- *Application Compliancy with CGDI*: Create a process and criteria for recognizing whether an organization's application is compliant with CGDI so that other organizations may more readily determine that the application may be integrated into their business processes.

- *Regional Partner Compliancy with CGDI*: Create a process and criteria for recognizing whether an organization's data warehouse is compliant with CGDI so that other organizations may more readily determine that the data may be trusted and may be integrated into their business processes.

- *Data Standards*: Clarify the mandate of GeoConnections in the process of designing and endorsing data standards in communities of practice, including those of the four priority areas.

- *Data Access and Exchange*: Continue efforts to deploy CGDI-endorsed standards for data access and exchange, and increase efforts to encourage inter-organizational data access and exchange through various means, including CGDI-endorsed standards.

Table of Contents

1. Introduction

"Integration of local/regional data using web services, within regional atlases, is a new and evolving concept. GeoConnections wishes to speed both the innovation and integration of regional scale information through this good practices study that will serve as an integration guide for current and future regional atlas partners, as well as the CGDI community at large." (GeoConnections, April 2007).

An important motivation for this study was GeoConnections' understanding that while there is an important opportunity to support decision makers through Canadian Geospatial Data Infrastructure (CGDI) compliant regional atlases, there is limited awareness of the numerous challenges in performing regional scale data integration and means for addressing them.

The need to improve the level of awareness was apparent from a review of submissions to GeoConnections' Announcement of Opportunity (AO) for Developing Regional Atlases for Decision-Making, July 6, 2006. Twenty-six proposals were received and reviewed[1] by a committee whose members comprised officials from various federal, provincial and municipal government agencies, including GeoConnections, each with a high proficiency in geomatics. An assessment of the proposals revealed the lowest scores were achieved on the CGDI-related criteria, including the criterion for Data Integration[2].

1.1 Study Objective

The objective of this study is to facilitate the integration of regional scale information through identifying good practices for CGDI stakeholders that:

- Enable agencies to cooperate for data sharing, keeping data closest to authoritative source and accessing distributed data via web services to reduce duplication;
- Enable agencies to establish their technology infrastructure in a manner that supports CGDI deployment;

[1] Proposals were assessed against the following major criteria:
1. Impact criteria
2. Relevance criteria
3. CGDI Criteria

[2] Data integration involves gathering and compiling data from disparate sources to support decision-making; a further definition is provided in Appendix D.

- Enable technology stakeholders within agencies to support data integration requirements of analysts;
- Identify means for users to assess the quality of integrated data that may have come from different scales or projections, or may have different alignments;
- Assist technicians to integrate data in various formats;
- Assist analysts to deliver and access current and relevant data;
- Demonstrate the value of taking a methodical approach to addressing the technical aspects of data integration using the CGDI.

The means for achieving this objective is to conduct case studies of four regional atlas and/or decision support system development projects to identify the challenges faced, and capture good practices for addressing them.

1.2 Study Approach and Methodologies

The approach to this study was to conduct it in the following phases and using the methodologies indicated:

- *Phase 1 - Scoping*: From a document review, we prepared a draft approach to conducting the study which led to the preparation of a scoping interview guide. This guide explored such issues as which GeoConnections sponsored projects should be selected for case study (see below) to explore good practices, to which audiences the practices should be targeted, and which of the many data integration activity areas should be focussed upon. Interviews were conducted with GeoConnections staff from each of the four program priority areas (Environment and Sustainable Development, Public Safety and Security, Aboriginal Matters, and Public Health), as well as technical experts involved in the design and implementation of CGDI. This phase concluded with the preparation of a Study Design Report that was approved by the Project Steering Committee established for the study;
- *Phase 2 – CGDI Framework*: We established a systematic approach for examining a CGDI application development project, and this framework served to determine the extent to which a project employed relevant CGDI-endorsed standards (See Appendix B - CGDI-endorsed standards Examined);
- *Phase 3 - Case Studies*: Based on a document review, we prepared Case Study Working Papers (see Appendix A - Case Study Working Paper Structure) that included a) the policy context for each of the selected projects, b) use cases for the applications that illustrated the need for data from multiple organizations, c) the challenges that were encountered in accessing and integrating data from different sources, d) data flow diagrams for each use case characterizing how data was accessed and delivered, and e)

revised data flow diagrams characterizing how this could have occurred if all relevant CGDI-endorsed standards were employed;

- *Phase 4 – Interviews*: We then conducted interviews with a minimum of five project stakeholders for each project to validate the information in the case study report, to review challenges faced, and to identify the practices followed to address them;

- *Phase 5 – Workshops*: Workshops were then held with up to 20 stakeholders in four communities to further explore the challenges faced over the course of each project and to arrive at a consensus on good practices for addressing those challenges. As practices were established they were fed into subsequent workshops to build on those results;

- *Phase 6 – Reporting*: This final report was drafted summarizing the study findings, and a presentation on those findings was made to the GeoConnections Project Advisory Committee.

1.3 Selection of Projects for Case Study

Considerable attention was given to selecting projects for case study because of their important effect on the findings. The selection was made based on a review of documents and websites for a range of projects identified by GeoConnections and the study team. Based on this review, a project profile template was prepared and is attached in Appendix C. Ten projects were identified by the study team as candidates, and which were further narrowed down to four using the following criteria:

- Lead's Jurisdiction: These could be federal (F), provincial or territorial (P), or local (L), the latter including regional governments. While consideration was given to having leads from each jurisdiction, assessments based on this criterion also took into consideration whether the solution had linkages to the other jurisdictions;

- Geographic Coverage: These could be national (N), provincial or territorial (P), or local (L). Assessments based on this criterion also sought to have an appropriate representation of each type;

- Technology Understanding/ Interest of Lead: Given the importance of the project lead to the case study methodology, consideration was made of the understanding by the lead of such issues as Service-Oriented Architectures, data integration, and data standards development;

- Extent of Challenges Faced: This was assessed based on the overall scope of the projects and their relevance to the study, including the extent to which the solution was operational and being used;

- Priority Geodata themes: A user needs assessment study[3] identified data themes that are a priority to geodata stakeholders in each of the priority areas, projects were sought that used this content;

- Practices in Technical Data Integration: This was assessed based on whether issues of data standards, data of varying projections and scales, and other technical matters arose through the project.

From this assessment, the study team recommended four projects for case study, the Project Advisory Committee approved this selection, and the leaders of those projects kindly agreed to participate. It should be noted that findings in this report are not attributed to specific projects as we assured the project leaders when they agreed to participate that the study would serve no other purpose but to inform the development of general good practices that would apply to any CGDI regional scale information integration project.

GeoConnections and Hickling Arthurs Low gratefully acknowledge the following project leaders and their teams for their insightful contributions and full cooperation during the studies of their CGDI projects:

- Steve Botham, Regional District of Fraser Fort George, British Columbia

- Silvia Strobl, Ontario Ministry of Natural Resources

- Robin McGinley, Cree Outfitting and Tourism Association; Rick Cuciurean, Cree Trappers Association; Valter Blazevic, Strata360, Quebec.

- Margaret Parkin, Regional of Waterloo, Ontario.

[3] Environics Research Group, Survey of Geographic Information Decision-makers, October, 2006.

2. Study Findings

In this Chapter we present findings in three areas:

- Anecdotal evidence of the broad range of data integration challenges and practices that were reported by those consulted;
- Six good practices in regional scale data integration stemming from the analysis of the case study projects; and
- Data integration challenges and opportunities that were reported in each of the four priority areas.

It should be noted that while the original intent of this study was to identify good practices in regional scale information integration, it quickly became apparent that it would be more appropriate to determine good practices to avoid the need for the often difficult and time-consuming work by technology managers and systems engineers to integrate data. This is consistent with the vision of CGDI, which is that end-users should be able to access atlases and decision support systems that call up and integrate the required data from various systems on various platforms in various projections and scales without the need for intervention by systems professionals.

2.1 Broad Range of Data Integration Challenges and Practices

After gaining an in-depth understanding of the 10 GeoConnections projects, it became apparent that a wide variety of technology and data strategies and practices were used to support data integration. These strategies reflected differing degrees of emphasis on CGDI-endorsed standards, which is best demonstrated by some of the views expressed by those consulted:

- "Technical data integration is the least of our challenges, gaining access to data in whatever form is our challenge."
- "The challenges to adopting data standards are not technical, they have more to do with personalities and their willingness to collaborate."
- "We will take the data in whatever form we can get it that is compatible with the vendor application we purchased, and we will provide it to the user as is: our role is a

coordinating one, we do not have a mandate to perform quality assurance on the data from other organizations."

- "Our users would rather have access sooner to geomatics data that is not integrated according to good practices than wait for it to be available in accordance with good technical practices because this would take a very long time; our stakeholders are local players with limited resources."

- "Some of the key challenges have been around getting orthos to display in real time, which required slicing large image files into millions of small tiles; we can't use CGDI now to access this data since these images are not yet being served by any other organization, so we serve it to ourselves using CGDI-endorsed standards."

- "We encountered rendering challenges with data coming through WMS services i.e., making data appear appropriately on top of colour air photos so they were legible."

- "The project data we need is not available from a WMS service for all of the project area so blank areas appear on the map. In order to avoid confusion, this dataset has been temporarily excluded from the map."

- "The implementation of WMS using a new data engine that lacks WFS services limits the value of data we can acquire."

- "While our solution can be construed in the context of a Service-Oriented Architecture (SOA), the reality is that our systems and those of our partners are not operating in such an environment."

- "Some key datasets for us are held in organizations with few resources, in our jurisdiction data is an asset that is guarded and traded as a precious resource."

- "It is our intent to make our data available to CGDI but we are focusing on accessing data of others through CGDI at this time."

- "Developing a data integration engine to serve up data does not mean that the information housed in it could be used. That was apparently not well understood. Data ownership issues and data filtering issues quickly arose and put the whole project in jeopardy."

Consultations with study stakeholders also revealed the need for terminology that pertained to clarifying:

- What is intended by "Authoritative Source";

- The different types of atlas that may exist;

- Distinctions between data integration and systems interoperability;

It is noted that these notions are not new, and they have been largely addressed in the broader information technology community. However, the definitions presented in Appendix D -

Terminology may be useful to the CGDI community because they are intended to clarify the roles and activities of those involved in regional scale data integration.

2.2 Good Practices in Regional Scale Data Integration

In this Section, we report the six good practice areas that were identified over the course of the study and that would mitigate the likelihood of data integration challenges:

- Develop and Endorse Data Standards
- Prepare Geospatial Data Profiles
- Establish Regional CGDI Partners
- Build Service-Oriented Architectures
- Establish Data Integration Service Centres
- Undertake Rapid Prototype Development

Each of the above practice areas is further developed in Appendix E – Good Practices Check List; as noted earlier, each of these areas was explored by positing good practices and validating them through the interviews and workshops.

2.2.1 Develop and Endorse Data Standards

An important foundation for effective data exchange among CGDI organizations is the data standards adopted by CGDI data providers. The extent to which organizations are able to share their data is in large part determined by the extent to which their data is maintained in compatible data standards. Clearly, common standards for data schemas, data validation, and data quality will greatly facilitate the data integration process, and possibly enable automated data integration processes between organizations.

There are, however, many factors that result in organizations adopting differing, and often inconsistent, data standards. These include organizations having differing mandates and jurisdictions, differing business processes, differing business information systems that may be built on proprietary data standards, and differing levels of security and access to confidential data. These differences may result in data schemas being fragmented across different technology platforms and systems, classification of entire large datasets as confidential because certain data elements within those datasets are confidential, or adoption of differing data quality control procedures so that, for example, addresses, or even postal codes, are not simple to geocode.

Practices for adopting data standards include purchasing a system that meets end-user requirements and thereby accepting that system's data standard, undertaking extensive

consultation and analysis to arrive at a data standard, and, a method preferred by those consulted, surveying the community of practice for a recognized data standard and adapting it for their purposes.

GeoConnections has established advisory committees in each of the four priority areas, and these bodies have been considering various approaches to facilitate the migration towards common data standards within their respective communities of practice; for example the first phase of the development of a critical infrastructure data standard has recently been completed for the public safety community.

An important challenge was in identifying bodies that had the mandate for establishing and maintaining data standards that could be employed by different organizations in different jurisdictions. While GeoConnections, being part of Natural Resources Canada, has a mandate in establishing standards for framework data (e.g. road network, satellite imagery), this mandate is less clear for geolinked data, such as public health surveillance data standards. Also of note is the particular challenge of establishing data standards for positioning geolinked data relative to geodata – the simplest examples being addresses and even postal codes which may be stored in various formats and displayed in various ways (for example, to where should an ambulance be dispatched for an emergency given a farm house address, when there may be multiple private driveways leading to a house kilometres into a large acerage?).

Challenges in regional scale data integration will continue as long as organizations in communities of practice have incompatible data standards. GeoConnections has established a process for promoting CGDI technical standards (e.g. WMS, WFS) for sharing data, classifying them as endorsed, recommended, for discussion, or under investigation. However, there is not a similar process in place for data standards, although clearly such standards are essential to mitigating data integration challenges and thereby accelerating the deployment of the CGDI.

2.2.2 Prepare Geospatial Data Profiles

Repeatedly over the course of this study challenges were reported when first efforts were made to acquire, by CGDI technical standards or other means, data from another organization. In spite of having entered into preliminary discussions with intended data providers, it was not uncommon for a project stakeholder to be quite surprised by a variety of factors, such as a data reprojection service not operating as planned, large gaps in coverage, very slow response times by mapping servers, or CGDI-endorsed standards implementations that were anticipated to work "out of the box" that turned out to require customization.

These experiences, which sometimes resulted in material delays in project deployment and considerable frustration among project stakeholders, may have been mitigated by entering into a more structured dialogue with external data providers to explore certain notions at the early stages of the project. Like the Check List mentioned above, these notions were discussed through interviews and workshops and the result is attached as Appendix F - Geospatial Data Profile Table. This Table, like the Check List, is intended to assist technology managers and systems engineers to gather the required metadata to scope, design and implement a regional atlas or decision support system that delivers and/or accesses CGDI data content from another

organization. Among other purposes, this tool is intended to mitigate the likelihood of challenges in regional-scale information integration identified through the case studies before systems are constructed. Importantly, by assisting to enter into such a dialogue, the tool may also assist in establishing and maintaining linkages among CGDI participants as their systems become increasingly interdependent, particularly when managers and engineers are unfamiliar with means for acquiring the metadata using CGDI services (e.g. Discovery Portal, Getcapabilities) or the metadata is incomplete.

2.2.3 Establish Regional CGDI Partners

Numerous data holdings were identified over the course of this study that would assist regional stakeholders in each of the priority areas. It was not uncommon for local organizations, some with more limited resources, to seek information from larger organizations with major data holdings. Although the deployment of the CGDI is well underway, there were instances where the projects examined faced greater than anticipated challenges accessing data from these larger warehouses using CGDI-endorsed standards.

It is recognized that the projects selected for case study were chosen because of a variety of factors, including that they were early adopters within their priority area for innovative applications of CGDI. The challenges stemmed, however, from a lack of priority and resources among key organizations, both those directly involved in the project and third party organizations, to implement and support continuous data delivery through CGDI. Were such organizations with data under high demand to have been recognized as CGDI partners with the mandate and resources to support applications like those examined (whether their data holdings be national, provincial, or local), it is believed that the knowledge and expertise of these regional nodes, and even the data and services themselves, would proliferate more quickly and easily into a national geospatial data infrastructure.

Whether such a provider would want to take on the role of being a CGDI partner was the subject of some discussion, but the scope of this study was limited to the technical aspects of regional scale data integration. It was clear, however, that there is more work required to determine where the mandate should lie for creating, maintaining and promoting CGDI holdings both initially, and as the CGDI evolves[4].

2.2.4 Build Service-Oriented Architectures

Organizations that contribute to the CGDI take on responsibility for delivering content and services to business processes and applications of other organizations. These other organizations' processes and applications may operate on different technology platforms and use different

[4] An anecdote was mentioned by one project stakeholder seeking to compare their regional data with similar data for their region in a national database. After a number of inquiries, the stakeholder was surprised to learn that the data in the national database for their region was their data. This demonstrates that while initially the national database provider could take on the mandate of CGDI partner, in the future the regional organizations could take on this role, and the national database would be created by aggregating (i.e. making cascading requests for) regional partner data holdings.

systems architectures, and in fact often may have multiple applications and architectures within their organization. The reasons for this variability are far too many to list here, but can be a result of changes to organizational technology investment strategies and service providers over time which creates various legacy systems, or the result of reorganizations of departments (or mergers and acquisitions) of organizations using different platforms.

Service-Oriented Architectures (SOAs) are a means for organizations to support data exchange in spite of their widely varying architectures. In fact, CGDI is based on organizations contributing to it having, or emulating, an SOA.

Before continuing, it is appropriate to review what SOAs are. SOAs can be characterized in different ways; a layered view is as follows[5]:

- Layer 1 - Technology Component Architecture: The Technology Component Architecture contains vendor specific products, services and their supporting architectures and identifies the lowest level components that can be re-used … "out-of-the-box".

- Layer 2 - Service Exchange Architecture: The middle layer supports the one-to-one mapping of a component offering to infrastructure services and also the construction of services composed of other services.

- Layer 3 - Business Application Architecture: The top layer allows business owners to package a tailored selection of services to be used in alignment and support of specific business requirements.

- Context - Business & Program Design: Sitting above the three layers is a formal approach to formulating business services using Business Architecture that dovetails closely with the Service Oriented Architecture's technical layers.

Alternatively, a functional characterization is as follows[6]:

- *Universal Data Access*: The ideal data integration platform would provide pre-built connectivity to a wide variety of packaged applications, mainframe systems, relational databases, and semi-structured and unstructured data. It should provide nearly unlimited data access via traditional physical as well as virtual data integration approaches, while minimizing the cost and complexity of accessing data regardless of where it resides;

- *Metadata Repository and Services*: The data integration platform should extend beyond data to its metadata—the "glue" that describes data values and their semantic meaning. It should provide a drag-and-drop user interface that enables developers to rapidly build business logic, processes, and transformations for data and make them available for reuse. It should have at its core a scalable metadata repository that stores and manages data models, transformations, workflows, and other artifacts;

[5] Service Oriented Architecture Strategy - Statement of Direction, Chief Information Officer Branch, Treasury Board Secretariat, Government of Canada, February 2006.
[6] "Data Integration in a Service-Oriented Architecture – White Paper", Informetrica, November 2005."

- *Data Integration Engine*: At the heart of a data integration platform is a high-performance engine for delivering data integration services, which offers a variety of flexible data delivery mechanisms and scalability for large-volume data transformations and movement over multiple concurrent sessions.

Indicators that an organization does not have an SOA become evident when data transfer between organizations is done through such means as:

- Use of FTP sites or e-mail: This may be because the data is in databases that are not accessible to external users, for example, because they require a client-side application to access the data and the databases don't support CGDI-endorsed standards; or

- Use of teleterminal systems: This may be because user authentication is controlled by an application, therefore to access the data one must sign onto the system rather than access it through CGDI-endorsed standards.

It should be noted that organizations may adopt an SOA using platforms that do not support CGDI-endorsed standards but that support other standards. Such other standards may be endorsed by CGDI in the future, and could include proprietary standards, mark-up languages, or others. The GeoConnections Management Board, and its CGDI technical advisory committee, continues to review the various standards that are emerging; the ones that become endorsed as CGDI-endorsed standards will ultimately be determined by the reality of the needs of each organization participating in CGDI, and the decisions they make to address those needs after reviewing the solutions available in the marketplace.

2.2.5 Establish Data Integration Service Centres

Having established an appropriate architecture that supports data exchange among internal and external end-users, organizations will want to ensure that they have the appropriate support services in place to address data integration matters as they arise. For example, through the case studies we saw various examples of end-users being classified, in terms of their security, by a central authority, and some spoke of establishing service level agreements with their users that had production systems using the data. As a result, it is apparent that some level of central support is required to support the data integration function for both internal and external end-users.

Also, while much of the preceding discussion seeks to prevent data integration issues, support may be also required to help end users deal with such issues as: achieving horizontal and vertical alignment between datasets when, for example, data is in different projections and there is no reprojection service available; or differing symbolization on datasets because Styled Layer Descriptor was not enabled by the organization serving up the data to CGDI end-users. Resolving such matters can often only be accomplished at a minimum through a dialogue with the end users, and sometimes only through a full user needs assessment. In fact, practices in one priority area may not be appropriate for another given their inherent differences, such as information confidentiality issues pertaining to personal health information, or issues

surrounding the need for time series analyses over very long periods as is required of environmental information.

While it is beyond the scope of this study to fully explore practices in establishing data integration service centres, the following options may be considered when determining the type of centre to establish:

- Policy Service Centre: has a mandate for data integration policy only (e.g., creating policies or rules for opening or restricting access to different data sets for different users, establishing templates for service level agreements).
- Guidance Service Centre: has mandate for delivery of assistance services only, (e.g., techniques for data quality control, procedures to advise internal and external CGDI users of data and service events).
- Shared Services Centre: has mandate for both policy and delivery of assistance services; under this design, personnel from various business units are brought into one organizational unit that has a mandate to support all business units. Benefits claimed of this approach include more consistent service delivery, and cost reductions through economies of scale.

2.2.6 Undertake Rapid Prototype Development

Finally, for those organizations consulted that developed their own applications, they strongly encouraged developing simple, quick prototypes at the very early stages of the project. One suggestion at a technical level was the importance of using the "GetCapabilities" request from organizations serving data using CGDI-endorsed standards. The benefits of making such requests and using the Discovery Portal include mitigating the likelihood of a wide variety of data integration challenges that can be easily avoided and that become readily apparent when reviewing the metadata returned. At the same time, project proponents also emphasized the importance of directly engaging end-users at key junctures of the development project, including the prototype development stage, to ensure that appropriate use cases are clearly defined and understood, and that the planned application will satisfy the requirements of the use cases.

2.3 Challenges and Opportunities in Implementing Good Practices

In this Section, we report challenges and opportunities that were identified from the projects examined and that affect the success with which CGDI is being deployed in each of the four priority areas. Although up to four projects were examined in each of the four priority areas, and each of these projects involved multiple stakeholders from their communities of practice, it should be noted that the following findings are based on evidence from a relatively small sample size.

2.3.1 Public Health

Public Health is a priority area for GeoConnections, which recognizes that the focus of public health improvement is on the social, environmental and economic factors affecting health as well as on communities and locations. Public health encompasses the entire social endeavour of assessing population health status and threats to it, developing policies and strategies across the full spectrum of intervention, and assuring that health needs are met and that services meet agreed-upon standards[7]. The main functions of public health practice include health promotion, prevention, health protection, health surveillance and population health assessment.

GeoConnections is aware of the tremendous potential that geospatial analysis and the Canadian Geospatial Data Infrastructure (CGDI) can provide in addressing public health issues. CGDI can add value not only through aiding the analytic dimension to public health in support of evidence-based decision making, but also in helping to improve the efficiency with which different health jurisdictions across the country interact and share data.

The GeoConnections Public Health Advisory Committee has identified two priority issue areas:

- *Population Health Surveillance:* Public health practitioners recognize that factors outside the health care system or sector can significantly affect health. Therefore, the entire range of individual and collective factors including social, economic and environmental health determinants, as well as their interactions, can be correlated with health status. Geospatial analysis allows for the integration of these multiple factors and conditions at different scales. The public health community can utilize geospatially referenced health determinants and health status information to make correlations, and identify priorities and strategies necessary to improve health and the factors that influence it. They also require mechanisms that will allow the sharing of information amongst organizations within their respective communities and jurisdictions, as well as externally.

- *Health Emergency Response and Inter-Emergency Planning:* In order for public health professionals to be familiar with the use of tools and to have capacity to respond when major emergencies eventually surface, the tools and capacity to use them must be in place on a day to day basis. This inter-emergency use of the tools for smaller emergencies and daily outbreak and surveillance activities can also help in the monitoring required for an early identification system for communicable diseases under surveillance. Public health practitioners can utilize geospatial information to track and forecast disease outbreaks, and monitor disease events across jurisdictions and international borders. This will aid in the identification of vulnerable populations, and during emergencies, allow for the assessment of public health risks and the mobilization of emergency functions. In

[7] Excerpt from a formal definition of public health from *A Dictionary of Public Health*, John M. Last, Oxford University Press, 2006: Public Health is an organized activity of society to promote, protect, improve, and when necessary, restore the health of individuals, specified groups, or the entire population. It is a combination of sciences, skills, and values that function through collective societal activities and involved programs, services, and institutions aimed at protecting and improving the health of all people. The term "public health" can describe a concept, a social institution, a set of scientific and professional disciplines and technologies, and a form of practice. It encompasses a wide range of services, institutions, professional groups, trades, and unskilled occupations. It is a way of thinking, a set of disciplines, an institution of society, and a manner of practice.

addition, it will facilitate the creation and dissemination of reports, advisories, alerts and warnings, nationally and internationally as necessary, as well as helping to determine other health emergency action, such as the initiation of vaccine production. Building the systems and training the staff in their use for inter-emergency purposes will help fine tune them for their use in an emergency, and get optimal use out of them in the interim.

Organizational units with a public health mandate vary in terms of their geospatial data needs for a number of reasons. One challenge is that they may have mandates for geographic regions that are defined inconsistently. For example, public health practitioners in school boards use school zones, government practitioners use their respective jurisdictional boundaries, Statistics Canada uses dissemination areas, and postal code areas are also often used, for example by Ontario's Local Health Information Networks. Each of these boundaries will evolve over time; however, they evolve for different reasons. Some evolve due to population socioeconomic factors (i.e., dissemination areas), some due to political factors (e.g., jurisdictional boundaries), and some due to demographic factors (e.g., school zones). The underlying drivers for the boundary changes may be similar (e.g., the development of new population centers affects all boundaries), but the boundaries themselves are often inconsistent and overlapping.

Those consulted identified an opportunity to establish a fabric of minimum mapping units such that each of the boundary areas could be characterized in terms of a collection of these units. By having such a common framework, data integration efforts would be greatly facilitated because data under one boundary system could be transformed to another boundary system.

Perhaps the single most important challenge is establishing CGDI infrastructure within organizations when these organizations have little or no awareness of CGDI. There are public health organizations with mapping services units and GIS systems; however, the vast majority of these have little awareness of CGDI. It is not uncommon, therefore, for public health professionals to have a network of sources of authoritative data and to access this data by various means, such as via FTP or teleterminals.

In order to introduce more systematic data exchange among public health partners, jurisdiction-level strategies are required to determine respective mandates of jurisdictional organizations, and where to deploy infrastructure, including personnel, equipment and data, that would assure access to public health data. A major challenge that was noted in achieving this was such organizations may have very widely varying mandates and resources, and therefore organizational capacity becomes an essential factor to consider.

A further challenge, one that is compounded by evolving organizational mandates and resources and evolving boundaries, is to be able to perform time series analyses on public health data. Inconsistent data standards and obsolescence of data storage media were also cited as factors that contribute to data integration challenges and that therefore require direct dialogue between the data custodian and the application developer to be able to ensure that appropriate data is accessed and that it is not misinterpreted. On the other hand, increasingly public health practitioners are placing emphasis on the recognized importance of appropriately geolocating their data, and these efforts should have the desired impact of further facilitating data exchange in the longer term. It was noted that efforts are ongoing to establish performance indicators for monitoring public health in some jurisdictions, and having access to these indicators over time, when they are

appropriately defined, also offers a means for performing time series analyses while reducing the burden of collecting and analyzing the source data.

It was also recognized that good quality metadata is vital when integrating public health data sets, and the maintenance of the metadata is equally important. CGDI presents an opportunity for the public health community to formalize the exchange of this information; however, it was noted that there is a limited amount in the Discovery Portal.

An overarching challenge within the public health community is defining the data requirements for decision-making. Such data can be stored according to a wide variety of data standards, and can be accessed from a wide variety of sources, from the yellow pages for locations of health practitioners, to municipal land use planning offices for recreational walkway locations, to Statistics Canada for indicators of regional socioeconomic well-being. For atlas and decision support system developers, even establishing a data standard for profiling and locating public health service providers presents challenges, requiring direct dialogue with other organizations with recognized data standards that, once adopted often require enhancement to address regional requirements to be consistent with regional boundary definitions. There was some speculation that establishing a minimum acceptable dataset for each minimum mapping unit would be very hard to achieve; however, once achieved this would greatly facilitate the development of data standards and the integration of the data itself.

The public health community continues to tackle the major challenge of determining its data requirements, and where and how such data could be obtained. Various jurisdictional legislation and organizational policies must be taken into consideration to ensure privacy and confidentiality restrictions are enforced while still supporting the important tasks of public health practitioners. One intermediate step currently being taken by some jurisdictions that could assist practitioners to benefit from and contribute to the CGDI is the creation of local applications that provide framework data (e.g., road networks and base maps) using CGDI-endorsed standards, but that also have public health data importing and geocoding tools that enable end-users to create atlases and decision support products on top of the CGDI maps. In this way, there will likely be a natural evolution towards fewer boundary definitions and geolocating methods, while requiring less infrastructure of smaller local organizations, addressing two of the challenges noted above.

2.3.2 Public Safety and Security

Public Safety and Security is a priority area for the program as GeoConnections recognizes that there are threats and hazards that have the potential to undermine the security and safety of Canadians. These can be intentional (e.g., terrorist incidents, criminal acts), accidental (e.g., human error, technological failure) or natural (e.g., meteorological, biological, geological). In managing threats and hazards, there is an increasing need for inter-jurisdictional co-operation and information sharing. Location-based information is a key resource for coordinating and assisting agencies from all levels in making crucial decisions related to public safety and security. The vision of GeoConnections is to improve decision-making in the Public Safety and Security user community by using the Canadian Geospatial Data Infrastructure.

To support the Public Safety and Security user community in emergency management the GeoConnections Public Safety and Security advisory committee has identified two priority areas:

- *Critical Infrastructure Identification*: Canadians rely on highly connected and highly interdependent infrastructures that are essential for their health, safety, security and economic well-being. In times of crisis or disasters, they seek assurance that the country's infrastructures will remain viable and resilient. Knowing what infrastructure is of significance to the public safety and security community and where it is located in relation to the event causing the crisis is a priority to support consequence management decisions by the responders.

- *Situational Awareness and Management of Consequences*: To better prepare for, manage and respond to threats and hazards that affect the safety and security of Canadians, Public Safety & Security organizations have a pressing need to gain and maintain situational awareness. As crises evolve, improved consequence management information for decision support is critical.

Local organizations are subject to various federal, provincial and local legislation that serves to plan for, and mitigate the impacts of, natural and man-made disasters and emergencies. Among other requirements, emergency response plans are required, which include assignments of powers and obligations of emergency stakeholder organizations pertaining to the planning for, declaration of, and response to, an emergency. Ultimately such policies and plans serve to save lives, and to protect infrastructure and the environment.

A major challenge for public safety stakeholders is determining the location of a distressed caller or an emergency incident. The factors contributing to this challenge include the variety of emergency response teams (e.g., ambulances, fire departments, municipal police, RCMP) and the variety of geographic information systems that respondents use to identify the location of an incident. Essential components of such systems are the road network and address point data; these not only assist in dispatching response teams but, if address point attribute data is available (i.e., is this address a building, if so is it a school, if so how many children are at that school?) they can also assist the planning and response functions.

While there is a generally accepted road network data standard across Canada, which is largely accepted by national, provincial, regional and local governments, more work needs to be done to arrive at an address point data standard. The differing requirements of data standards stem from the differing requirements of stakeholders; for example, determining how to most quickly access a house on a large multi-acre rural property that may have multiple very long driveways requires different data compared to accessing an apartment in a large municipal apartment complex with multiple buildings, all having the same municipal address. Studies have been undertaken to define a data standard for critical infrastructure, and this would be an important element of a broader address point data standard, but more analysis is planned in some jurisdictions to establish a comprehensive address point data standard.

Further, because of the integrated nature of road networks, and that they fall under the mandate of different jurisdictions (e.g., municipal roads are maintained by municipalities, forestry roads are maintained by provinces), there are inherent challenges in ensuring that there is a common

road network and address point fabric that is maintained and available to all stakeholders. Ideally, each jurisdiction would maintain its own data and make it available to others through CGDI-endorsed standards. Then, emergency dispatchers seeking the road network for a region could have a decision support system that would issue a request from the different jurisdictions for the data and produce an integrated road map.

Many smaller municipalities, however, do not have an awareness of CGDI nor the resources for infrastructure to deliver road network data in this way. Accordingly, although they would prefer not to do so, a higher level of government may have to step in to ensure that local data is accurate and is aligned with road networks of other jurisdictions. In this instance, and echoing findings from the public health area, there is an opportunity to reduce the onus on smaller, local organizations to serve their data through CGDI-endorsed standards by assigning responsibility and providing resources to one regional organization to maintain the CGDI contributions of the smaller organizations. In this way, few if any additional requirements are made of the smaller organizations but their holdings are still made available via their regional partner to the broader CGDI community.

But, it was noted that, even with agreed upon data standards, there is often a data integration effort required for various reasons, for example to bring in municipal alleys, or the roads built by forestry companies. For this reason, many end-users of this data may prefer to acquire the integrated road network from a single provider rather than via CGDI from a variety of road network providers, especially if they have specific technology platforms that may require transformations to render the data compatible with their environment.

A challenge that arose with public safety stakeholders is establishing the respective roles and responsibilities of the hierarchy of different government jurisdictions relative to the horizontal CGDI environment. For example, and this issue arose in other priority areas as well, if a lower level of government creates a regional decision support application that accesses data from a variety of partners, to what extent does that partner take on a mandate to manage the end-users' requests pertaining to data holdings external to the regional partner? One might envision a day when CGDI-endorsed standards and applications are sufficiently robust to preclude the need for support services, just as is the case with data providers on the Internet today; however, until then many end users may have expectations of regional partners that are unrealistic.

2.3.3 Environment and Sustainable Development

The Environment and Sustainable Development advisory committee has identified two areas that are of high importance within this community of practice that GeoConnections focuses on[8]:

- *Land-Use Planning (also includes oceans and freshwater planning)*: Agencies that own, manage, and use Canada's land base have responsibility for a range land use planning issues, and are responding to numerous legislative and regulatory requirements. Two key groups are land-use planners who operate under provincial or municipal legislation, and regulatory bodies that enforce land-use regulations. In marine ecosystems, the federal government has responsibility for coordinated planning among competing interests.

[8] Environment and Sustainable Development Roadmap Draft Version 1.6, GeoConnections, June 14, 2007

- *Environmental Assessment (includes large project, regional and strategic environmental assessments)*: Coordinating environmental assessment activities is the responsibility of federal and provincial environmental assessment agencies, and complying with environmental assessment legislation and regulations is the responsibility of land developers. Regional and strategic environmental assessments are a program coordinated by the Canadian Environmental Assessment Agency (CEAA). The responsibility for implementing this program involves a mix of government, non-government and private stakeholders.

The goal of the GeoConnections program for environment and sustainable development is to support land-use planning and regulatory processes by encouraging the discovery of, access to, and sharing of geospatial data that support effective decision-making. To effectively address decision-making needs in support of sustainable development objectives, land-use planning organizations, regulatory authorities and environmental assessment practitioners require access to the best available data and the best available means to integrate and analyse that data. Furthermore, the results of such integration needs to be nested within decision-making processes that incorporate social, economic and environmental factors in pursuit of broader sustainable development goals.

The ability to meet these needs can be enhanced through the expanded use and analysis of geospatial data that is linked to broader-based information management systems and/or procedures that support decision-making. To effectively address land-use planning and environmental assessment processes requires an integrated, ecosystem scale approach. Integrated Ecosystem Management[9] (IEM) is a systematic approach that can contribute to these issues by optimizing environmental, economic and socio-cultural objectives considered over space and time and across jurisdictions. The CGDI is ideally suited to contribute geospatial data and and services to support integrated decision-making.

Within each province in Canada, significant funding and volunteer effort is going into land and ecosystem stewardship activities. In Ontario, for example, the Ministry of Natural Resources (MNR) initiated the Ontario Stewardship community-based environmental program[10] in 1995. Community Stewardship Councils made up of local landowners, resource users, and interest group representatives carry out annual work plans based on what they perceive to be the environmental priorities for their community. Each Council is eligible to receive annual seed funding of up to $10,000 from MNR, and the Councils use these funds to leverage additional resources through partnerships and collaboration. Ontario government strategic priorities are pursued through this combination of influence, support, and community empowerment.

Regional atlas and decision support systems are being developed to support such provincial stewardship programs by providing comprehensive and standardized information on stewardship

[9] [Integrated Ecosystem Management (IEM)] approaches planning from the point of view of *whole ecosystems*. IEM optimizes a broad range of economic, social and environmental objectives, and addresses a multitude of industrial, recreational, cultural and other activities (ILM Coalition, 2005). IEM is an approach that applies to terrestrial landscapes (typically called Integrated Landscape Management or ILM), watersheds, and marine ecosystems.

[10] http://www.ontariostewardship.org/ontariostewardship/dynamicImages/3305_OS_Guidelines_Final.pdf

activities and projects across their jurisdictions. In these communities, the practice has been for stewardship organizations to collect data about their ecological restoration projects using spreadsheet applications and mostly capturing data about associated costs. Few such organizations use databases for recording detailed environmental information from site visits to candidate ecological restoration sites. If an organization collects restoration data, they have often independently decided on what data is collected and to what standard, and these standards often differ from organization to organization. As a result, even if one could obtain such data records from multiple projects, it would be difficult to analyse and summarise.

Collaborating on systems for tracking ecological restoration activities assists conservation communities to more effectively deploy its resources, rather than reinventing similar applications in multiple jurisdictions and collecting data that may not be comparable. While many in the environment and sustainable development community are fairly sophisticated users of geospatial information technologies, the stewardship part of the community is populated by a very large number of small community groups and individual landowners who have had limited exposure to these technologies and their value in planning and reporting on stewardship activities.

Given the nature and structure of this community, there are considerable barriers to the effective use of geospatial data and the CGDI. There is limited experience and resources to help potential partners assess the costs and benefits of investing in a shared solution environment rather than continuing with their own standalone systems. Limited effort has been devoted to the creation of data standards, or identifying and describing the roles of various members in the stewardship Community of Practice (CoP). It is difficult to work towards achieving buy-in on a common application's use by a community of practice that comprises many organizations and agencies that often compete for resources and aren't always comfortable sharing data about where they are implementing stewardship projects. Long-term sustainability of systems is also a challenge because of the volunteer nature of this community, and because most available government funding is for only short-term projects.

The technical matters involved with CGDI that stakeholders face are also a challenge. Many partners have good data, albeit in a wide variety of standards and formats, but are not sophisticated in their use of geospatial information or are not capable of serving their data in WMS, WFS or other CGDI-endorsed standards. Concerns of stakeholders that privacy and confidentiality can be protected in a shared systems environment, when site/project data is down to the level of individual landowners, is a major impediment to buy-in in the stewardship community. This requires getting agreement on and developing the security policies and tools required for the application, including decisions on the various security levels of data that different partners can expect to access.

As noted above, a key technical challenge in these types of projects is the absence of standards for thematic data such as stewardship data. The stewardship community must undertake more work to determine user requirements to be able to establish and deploy data standards. It is important to note that the scope of geopspatial data applications in the broader Environment and Sustainable Development community of practice extend well beyond those of the project studied here. For example, much work remains to arrive a standard for water quality data given the many different community and jurisdictional stakeholders concerned.

2.3.4 Matters of Importance to Aboriginal Peoples

The Matters of Importance to Aboriginal Peoples advisory committee identified two areas that are of high importance within Aboriginal communities that GeoConnections will focus on:

- *Land and Resource Management/Community Planning:* Aboriginal leaders, managers and land planners require improved planning tools and information in order to manage communities, Aboriginal treaty and settlement lands, and co-managed lands and resources in a sustainable and effective manner. GeoConnections is addressing this issue by assisting Aboriginal leaders, managers and land planners, governments, and industry to increasingly share location-based information for improved partnerships and better land and resource co-management.

- *Geomatics and CGDI awareness:* Aboriginal organizations seek to have a better understanding of the benefits, challenges and sustainable methods for using geomatics and the CGDI for decision-making within the Métis, Inuit and First Nations communities. GeoConnections is addressing this issue by assisting Aboriginal organizations to implement geomatics and the CGDI for decision-making in an effective and sustainable manner.

Aboriginal organizations have a wide range of knowledge of geospatial information technologies, some preferring to build in-house capacity and others preferring to outsource such expertise. While the members of these communities often have extensive expertise with geospatial data using maps and plans, for example for hunting, fishing and trapping activities, they often have not dealt with such data in a computer environment. Another significant barrier has been limited access within these communities to the Internet, although this is rapidly changing. As a result, the CGDI is unfamiliar to most Aboriginal organizations.

However, circumstances are changing and the move to self-government is highlighting the need to develop better land and resource management decision-making capabilities. Like all governments, Aboriginal organizations must deal with multiple partners (their own citizens, other governments, industry, and NGO's) in a timely and effective manner. Much of the content involved is geospatial, or has a geospatial component and this geospatial component cross-cuts not just the organizations, but also the issues involved – access to resources is directly linked to legal issues, sustainable development affects health, youth issues are affected by all. In the more progressive communities where organizations have adopted geospatial technologies, duplication of effort is already occurring, and there is pressure to correct this situation, including by deploying the CGDI.

Developers of regional atlases and decision support systems for Aboriginal communities face similar technical challenges as those in the other priority communities of practice. The ability to access some data is hindered by incomplete Web services (e.g., 1:50,000 NTDB data is not completed as a WMS service for all of Canada) or shortcomings in the commercial software implementations of the CGDI-endorsed standards (e.g., zoom levels and symbology not adjustable on a layer by layer basis, inability to rename layers, long layer names in both French and English sometimes causing unnecessary horizontal scrolling in the layers window, etc.). Users are also experiencing performance problems with WMS services (e.g., slow map displays

with combined map layers from several different map files, map background generated from WMS layers not printing on large scale printouts, etc.) due to Internet performance issues noted above.

Confidentially and privacy concerns are also particularly evident in this CoP. Confidentiality concerns about traditional knowledge makes the management of data access by different users to different data content (e.g., restricting access by certain users to certain types of data and different geographical areas) particularly challenging in a shared systems environment. For example, there is concern about making traditional place names databases universally accessible, because many traditional names for geographic features contain land use information (e.g., the name of a lake implies good fishing, good hunting, etc.). This creates various technical challenges; for example to filter information for different access privileges would require considerable investment to classify the information in many large datasets.

3. Study Conclusions and Recommendations

The intent of CGDI is to reduce the efforts required, often considerable, by geospatial data stakeholders to access and integrate data, both within their organization and with their partners. When implemented successfully, the need for human intervention in the data integration process will be reduced as CGDI data exchange standards are embraced by end-users and solution developers, and organizations have adopted the data standards and infrastructure to make them work effectively.

Factors that will contribute to the speed of CGDI deployment will include: the public sector establishing stable funding mechanisms for Regional CGDI Partners; the geomatics industry adopting and enhancing CGDI-endorsed standards for data access and exchange instead of closed proprietary methods; and end-users recognizing the benefits of, and implementing, the required policies, procedures and infrastructure for CGDI to work effectively.

To accelerate this transition, this study uncovered several issues that, if addressed, would create an enabling framework for CGDI to operate more effectively. The good practices identified through this study are intended for technology managers and developers building a regional atlas or decision support system for an organization. However, more needs to be done to enhance the

framework within which such organizations operate, given the very high inter-organizational dependencies that are created when organizations adopt the CGDI.

The main issues are summarized as follows, and are developed further in each of the recommendations and considerations discussed below:

- *Regional Project Compliancy with CGDI*: Creating a process and criteria for recognizing an organization's application is compliant with CGDI so that other organizations may more readily determine that the application may be integrated into their business processes;

- *Regional Partner Compliancy with CGDI*: Creating a process and criteria for recognizing a major data provider organization's data warehouse is compliant with CGDI so that other organizations may more readily determine that the application may be integrated into their business processes. Ultimately CGDI seeks to deliver trusted applications and data to end-users, and therefore processes for establishing application and data warehouse compliance would be desirable;

- *Data Standards*: Clarifying the mandate of GeoConnections in the process of designing and endorsing data standards in communities of practice, including those of the four priority areas;

- *Promoting CGDI-endorsed standards*: Continuing efforts to deploy CGDI-endorsed standards for data access and exchange, and increasing efforts to encourage inter-organizational data access and exchange through various means, including CGDI-endorsed standards.

3.1 Application Compliancy with CGDI

There is considerable uncertainty within the community about what is required for a regional atlas or decision support system to be "compliant" with Canadian Geospatial Data Infrastructure standards. To ensure that such applications are compliant, it is timely to establish a process and criteria for assessing CGDI application compliance; such a process would facilitate and accelerate end-users participating in, and benefiting from, the CGDI as it would quickly indicate to systems designers and data analysts whether an application, and in particular its data sources and services, could be relied upon by their application. However, in the absence of a clear process and criteria defining application compliancy, the premature accordance of CGDI compliancy to an application, or claims of compliancy, exposes CGDI to undue risk through its association with applications that may fall short in some ways.

Recommendation: GeoConnections should establish a process and criteria for determining application compliance with CGDI. Such a process would need to be defined, and should address such considerations as: criteria for evaluating which of an organization's datasets should be made available to CGDI, what the legitimate bases for declining access should be; and establishing appropriate guidelines for determining what CGDI-endorsed standards are expected to be employed for providing access to data through CGDI.

3.2 Regional Partner Compliancy with CGDI

There is an important role for regional geospatial data providers[11] to deliver trusted data and services through the CGDI. The greatest use of CGDI-endorsed standards at present is for accessing and delivering framework data[12]; however, use will rise for thematic data in the community's of practice in the four priority areas. For this to occur, regional framework data provider organizations must commit to delivering such services to a common standard, and have sufficient resources to do so, particularly at the local level, in order for communities of practice to be able to build on these base maps.

Recommendation: GeoConnections should establish a process and criteria for determining Regional Partner compliance with CGDI. Such a process would need to be defined but would include consideration of the extent to which the organization's geospatial data is discoverable, accessible and reliably available to stakeholders, and authorized, organizations. Identification of incentives to encourage such organizations to take on this expanded role would be required, and GeoConnections should consider providing some.

3.3 Data Standards

An important challenge in regional scale data integration among multiple organizations is arriving at data standards for geolinked data, whether by adopting the same ones or by supporting data transformation to a common one. Stakeholders in each of the four priority areas are seeking to facilitate data exchange among organizations in their sector, and GeoConnections has been considering various approaches to facilitate this including supporting the development of standards for all thematic areas, the first phase of which has recently been completed for Public Safety stakeholders.

Recommendation: GeoConnections should undertake a review of the extent of its role in identifying, developing, endorsing, and promoting geolinked data standards. Such a review should take into consideration whether inter-jurisdictional, or appropriately mandated not-for-profit, bodies exist within the priority areas that could undertake part or all of this essential function. Such a review should also take into consideration the authority and mandate for Thematic Advisory and other related committees in matters pertaining to geolinked data standards in the priority areas.

[11] Note that regional organizations may have national (e.g. National Atlas), provincial (e.g. Land Information Ontario) or even local (e.g. City of Prince George) mandates.

[12] Here "framework data" is defined as "foundation or base geographic data used to reference the location of other datasets", from "Framework Data Defined – A Global Approach", Leah Howes, GeoConnections, July 28, 2006.

3.4 Data Access and Exchange

The geospatial mapping community in Canada has made great strides over the past fifteen years in reducing duplication and increasing the sharing of geodata that can be integrated with much less effort than in the past. While the reasons for this are far too many to list here, they include the efforts of the Canadian Council on Geomatics to harmonize data standards, the adoption of innovative technologies by departments responsible for providing mapping services, and the adoption of a data sharing, in contrast to closed proprietary, culture within the geomatics industry. As a result, it is a natural extension to deploy CGDI-endorsed standards, in spite of the important organizational transformations that are required to support interdependent geospatial information systems.

However, in the four GeoConnections priority areas, GeoConnections may wish to continue to emphasize promoting CGDI-endorsed standards, such as WFS, but place additional emphasis on promoting data exchange generally. There was some indication that extensive efforts were made by the projects examined to implement CGDI-endorsed standards, although often not successfully. This delayed projects, sometimes considerably, when alternative approaches would have enabled the deployment of the atlas or decision support system sooner. In the nearer term, the success of GeoConnections could be measured more by the amount and extent of data sharing between priority area organizations[13], and less by the means through which the data is shared.

[13] It is noted that GeoConnections II is focusing on the following outcome and output, which is not believed to be inconsistent with such a change in emphasis: ***Outcome*** - Users recognize the value of regionally integrated information in addressing numerous inter-jurisdictional issues using the CGDI; ***Output*** - Local/regional/provincial data content integrated regionally to enhance CGDI applications.

A. *Case Study Report Structure*

---

1. Introduction
   a) Case study objective
   b) Rationale for selection of case study
   c) Key issues to explore (e.g. integration of electronic vs. paper-based data)
2. Resources
   a) Staff expertise and responsibilities
   b) Cash and in-kind funding, including how it was obtained
   c) Key contacts and sources of information
3. Data integration challenges, good practices, and lessons learned
   a) Opportunities, demonstrated by real world examples (e.g. reduced duplication between organization 1 and agency 2), that the data integration initiative seeks to address
   b) Challenges faced involving data requirements and integration processes
   c) Good practices and lessons learned for addressing the challenges. These may involve, but are not limited to, the following:
      i) Development and Endorsement of Data Standards
         - Endorsed specifications
         - Recommendation papers
         - Discussion papers
         - Under investigation
      ii) Building a Service-Oriented Architecture
         - *Universal Data Access*
         - *Metadata Repository and Services*
         - *Data Integration Engine*
      iii) Establishing a Data Integration Service Centre
         - Shared Services Centre
         - Policy Service Centre
         - Guidance Service Centre

iv) Supporting Data Integration

- Establishing a data content dependency matrix (user group vs. dependency requirement)
- Defining, maintaining and publishing a data theme integration quality guide level (e.g. The integration of two CGDI endorsed data themes with identical projections and scales would result in a data integration quality guide level of "excellent")
- Defining a security and access control matrix (user group vs. which themes can access)
- Defining and publishing a data service level policy (e.g. target response times to meet requests for data, minimum notice period prior to data service shutdown, performance constraints due to data channel sizes, etc.)

v) Performing Data Integration

- Providing a data dependency notification to data service providers
- Achieving horizontal and vertical alignment between datasets
- Addressing issues with varying data projections and scales
- Addressing issues with varying data models, attribution and symbolization

4. Performance indicators for the good practices. As a possible example, if the described practice is related to arriving at a common data standard, the performance indicator could be a statement such as "The relevant federal, provincial, and/or regional body endorses the common data standard".
5. Good practices check list: TBD
6. Lexicon additions
7. Workshop Issues Explored
8. Appendices
   a) Profile of initiative
      i) Initiative horizontal policy objectives
      ii) Intended users/beneficiaries of the results
      iii) Information systems environment, including software applications (e.g. ArcInfo)
      iv) Geospatial data environment (e.g. CGDI services, geospatial data standards)
      v) Brief description of governance and management for the initiative
   b) Stakeholders consulted
   c) Cash and In-Kind budget table
   d) Standards organizations involved
   e) References
9. For more information contact
   a) Project leader contact information
   b) GeoConnections contact information

B. CGDI-endorsed Standards Examined

Three steps were taken to explore good practices in regional scale data integration using CGDI-endorsed standards, as follows:

1) Determine which CGDI-endorsed standards were included in the study scope;

2) Examine each case study project to determine where CGDI-endorsed standards were employed to access and exchange data; and

3) Identify additional opportunities to deploy CGDI-endorsed standards, if any, and then explore with project why these weren't pursued – most often for good reason.

Determine CGDI-endorsed standards in Scope

For this study, the CGDI **endorsed** standards were considered to be:

- WMS, WFS, and WCS with extended functionality from SLD, Filter encoding, GML and Web Map Context Document.

- The Catalogue Service Interface (CAT) search and retrieval of metadata (ISO Dublin Core Metadata Element Set). We consider that CAT is endorsed as it is generally understood to be replacing the Z39.50 standard for search and retrieval of FGDC compliant metadata. Note that formally, CAT is a CGDI recommended standard.

- Again, though still endorsed, the Z39.50 protocol, a client server protocol for searching and retrieving FGDC compliant metadata stored in remote databases, we consider good practices are to be established with greater consideration given to CAT.

Table B-1 presents a summary of CGDI-endorsed standards at the time of the study, and how they apply to Data Integration for a Regional Atlas or Decision Support System.

Table B-1: Summary of CGDI-endorsed standards

CGDI-endorsed standards	Status of CGDI-endorsed standard	Applicable to Regional Atlas/DSS	Considered during Data Integration	Comment, and whether applies to WMS, WFS, WCS, CAT or Z39.50
Catalogue Services Interface	Recommended	Will replace Z39.50	Will replace Z39.50	Will replace Z39.50
Metadata for Geodata	Endorsed	Yes, if connecting. to CGDI. Will be replaced	Yes, if connecting to CGDI. Will be replaced	FGDC metadata standard. Will be replaced by ISO Dublin Core Metadata Element Set
Service Registry	Recommended	Too early	Too early	Most likely will not be endorsed by CGDI
Geodata Discovery Service	Endorsed	Yes, if connecting to CGDI. Will be replaced by CAT	Yes, if connecting to CGDI. Will be replaced by CAT	This will be replaced by CAT
Filter Encoding	Endorsed	Yes	Yes	Enhances WFS, SLD and Gazetteer
Gazetteer Service	Discussion	Too early, unless supported by the WFS	Too early, unless supported by the WFS	Enhances WFS
Geolinked Data Access Service	Recommended	Too early	Too early	Not used yet
Geolinking Service	Discussion	Too early	Too early	Not used yet
Styled Layer Descriptor (SLD)	Endorsed	Yes	Yes	Enhances WMS
Web Feature Service (WFS)	Endorsed	Yes	Yes	WFS
Web Map Context Document	Endorsed	Yes	No, more for the website developer	Enhances WMS
Web Map Service (WMS)	Endorsed	Yes	Yes	WMS
Geographic Markup Language (GML)	Endorsed	Yes	Yes, if the selected OGC service supports GML	Communication language of OGC services
Web Coverage Service (WCS)	Endorsed	Yes	Yes	WCS

CGDI-endorsed standards	Status of CGDI-endorsed standard	Applicable to Regional Atlas/DSS	Considered during Data Integration	Comment, and whether applies to WMS, WFS, WCS, CAT or Z39.50
Web Processing Service	Discussion	Too early	Too early	Not used yet

Determine How Standards Were Deployed.

With the CGDI-endorsed standards identified, the technical team then examined each use case for the decision support systems selected for case study. For each step of the use cases (e.g., "load base map"), the team considered where and how the data was obtained. Typically, one might determine that a WMS standard was used to access a regional authoritative source of base maps.

Identify Other Opportunities to Deploy CGDI-endorsed standards

Finally, the technical team re-examined each step of each use case to determine whether or not there were additional opportunities to employ CGDI-endorsed standards. For example, if one step of the use case was "load road network data from FTP site", this would be flagged as an opportunity to use a CGDI-endorsed standard to access the data directly from the source of the data.

The results of the above analyses were summarized in data flow diagrams, simple examples of which are presented in Figures B-1 and B-2 below.

Figure B-1: Portrayal of Data Flow Using a Java Applet

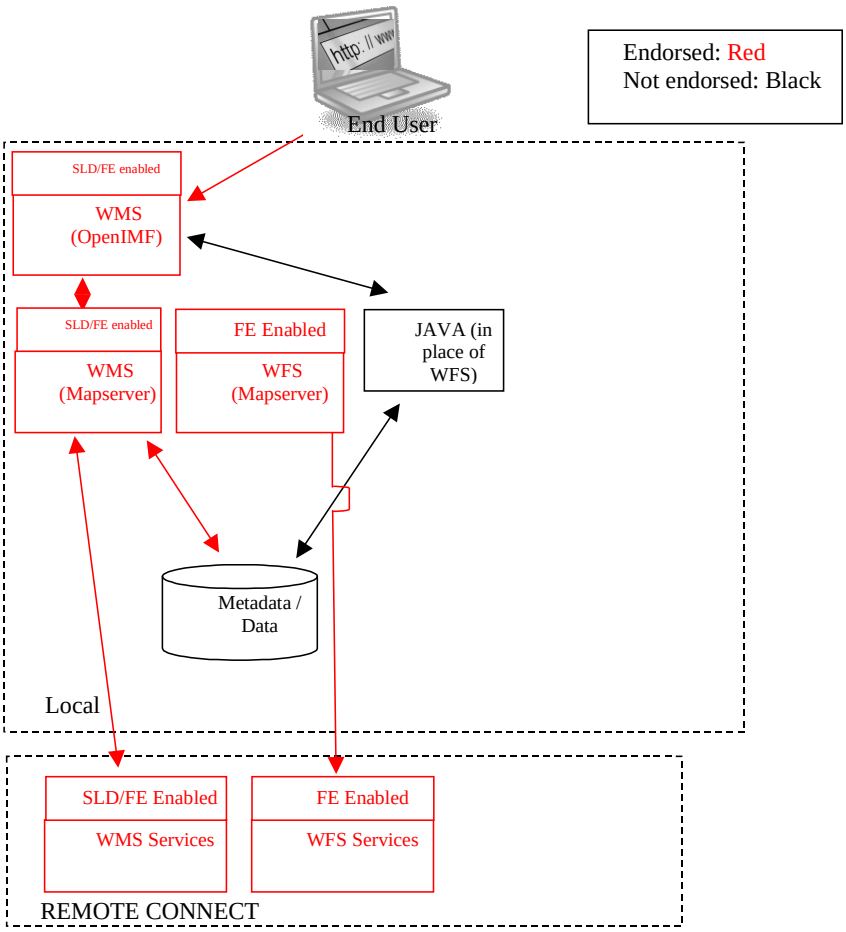
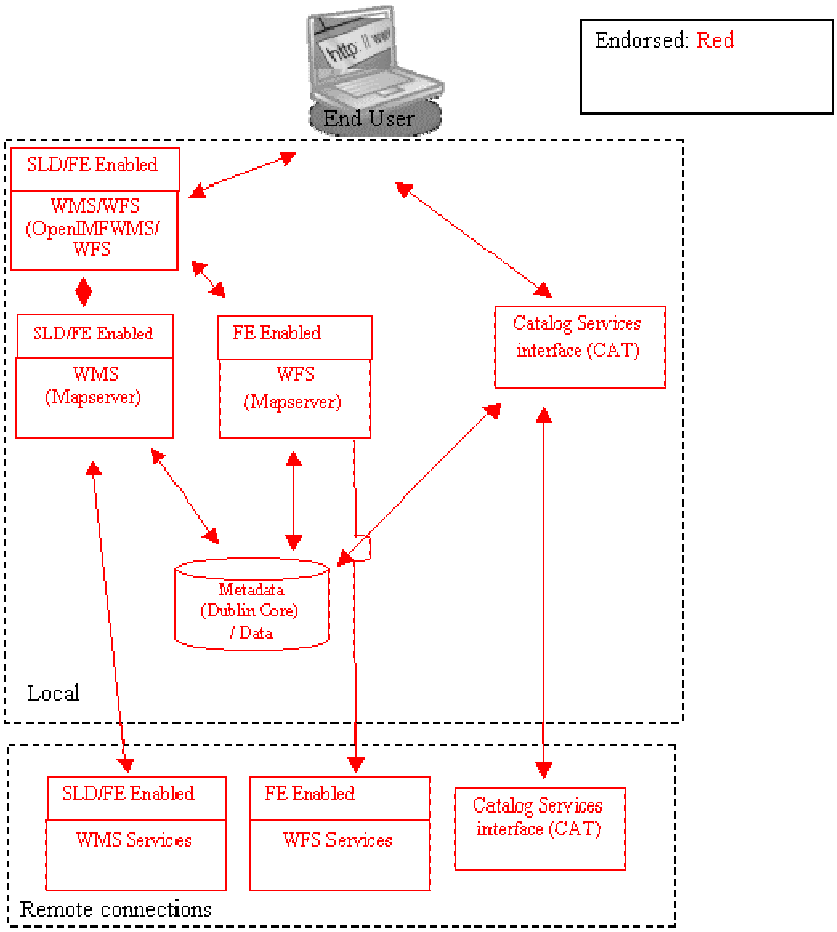


Figure B-2: Portrayal of a Data Flow Using WFS Instead of a Java Applet



C. Project Profile Tool

The following template was used in the project selection process described in Section 1.3 to collect data for 10 candidate GeoConnections sponsored projects.

Table C-1: Candidate Project Profile Information.

Initiative Characteristic	Case Study 1	Case Study 2	Case Study 3	Case Study 4
1) Description of data integration initiative				
2) Number of months initiative has been operational				
3) Lead Contact Name and Phone Number				
4) Lead Contact e-mail				
5) Initiative/Atlas Website URL				
6) Estimated Cost to Date ($)				
7) Percentage of project complete (0-100%)				
8) Federal Gov't Contributes? (1-5)*				
9) Provincial Gov't Contributes? (1-5)				
10) Municipal Gov't Contributes? (1-5)				
11) Aboriginal Community contributes?				
12) Private Sector Contributes? (1-5)				
13) Not-for-profit Orgs Contribute? (1-5)				
14) Addresses Public Health14 Priorities? (1-5)				

[14] Facilitating disease surveillance or population health analysis;

Initiative Characteristic	Case Study 1	Case Study 2	Case Study 3	Case Study 4
15) Addresses Public Safety/Security Priorities15? (1-5)				
16) Addresses Environment/Sustainable Development16 Priorities? (1-5)				
17) Addresses Aboriginal17 Priorities? (1-5)				
18) Employs watershed data? (1-5)				
19) Employs satellite imagery data? (1-5)				
20) Employs aerial photography data? (1-5)				
21) Employs land use data? (1-5)				
22) Employs socio-economic data (e.g. from Statistics Canada)? (1-5)				
23) Atlas/DSS is CGDI compliant18? (1-5)				
24) Good initiative to examine to learn Good Practices? (1-5)				
25) Good initiative to examine to accelerate innovation and adoption of CGDI? (1-5)				
26) Overall, initiative is a success? (1-5)				

[15] Facilitating emergency management and response, or critical infrastructure protection;
[16] Facilitating integrated land/marine management, including land/water-use planning, environmental assessment, and indicator monitoring
17 Facilitating land and resource management, and community planning
[18] See http://www.geoconnections.org/en/communities/developers/standards

D. Terminology

There were three areas where terminology employed by those consulted resulted in confusion over the course of the study; they pertained to clarifying:

- What is intended by “authoritative source”;
- Distinctions between “data integration” and “systems interoperability”;

CGDI Definitions

- *Geospatial*: Referring to location relative to the Earth's surface. "Geospatial" is more precise in many GIS contexts than "geographic," because geospatial information is often used in ways that do not involve a graphic representation, or map, of the information.
- *Geodata*: Georeferenced spatial data such as a road network or a satellite image. Geodata explicitly describes the spatial extent of a set of features or describes a measurable surface. It includes both geospatial data and geolinked data.
- *Geolinked data*: Data that is referenced to an identified set of geographic features without including the spatial description of those features. Geolinked data is normally attribute data in tabular data (such as population counts) that refers to a known framework (such as provinces), where the elements (the provinces) are referred to by their unique identifier (such as the province name). Geolinked data refers to all attribute data that is not directly attached and bundled with the geographic coordinates to which it applies.
- *Atlas*: A collection of geospatial and non-geospatial information (maps, charts, tables, pictures, audio, etc.) organized around a coherent theme. For example, a water resources atlas, a child health atlas, a flood risk atlas, or an adult literacy atlas.
- *Region*: A region is an area defined by the extent of pertinent subject matter and the needs of those who will use the information found in the atlas to make decisions.

Authoritative Source

A mantra within the CGDI community is that data analysts should obtain their data “closest to appropriate authoritative source”; however, there was much discussion over the course of this study on how to determine who an appropriate authoritative source is. The outcome of these discussions was that the authoritative source changes depending on the application. For example, the authoritative source for City of Waterloo road data is the city, and the authoritative source for

Ontario provincial road data is Land Information Ontario, even though LIO obtained the road data for Waterloo from the Region of Waterloo, and the Region of Waterloo obtained it from the city. Accordingly, it became apparent that there were three notions that require separation 1) the data authority, 2) the data integrator, and 3) the custodian of authoritative data.

- *Data Authority*: an organization, or an individual within an organization, that has the authority to approve access to data for an end-user external to that organization, and to approve organizational policies and procedures affecting the definition, collection, maintenance, integration, use, and archiving of the data. Before approaching a data authority, the end-user will have completed a User Needs Assessment that will have determined that the data authority's data is the most appropriate for the atlas or decision support system under consideration[19].

- *Data custodian*: an organization, or an individual within an organization, that has been delegated responsibility by the data authority to administer authorized end-users' requests for data, including entering into end user agreements, and providing end user accounts for operating systems or databases. Data custodians are also responsible for communicating and enforcing policies and procedures (for example by reducing an end-users access privileges) pertaining to the definition, collection, maintenance, integration, use, and archiving of the data. It is noted that a data custodian may or may not be employed by the same organization as the data authority.[20]

- *Data Integrator:* an organization, or individual in an organization, is a data custodian who adds value to data from multiple data authorities, who could be both internal and external, by applying their organization's policies and procedures to combine the data into information products and services that meet end-user needs. A data integrator may need to reconcile differences between the policies and procedures of the different data authorities contributing to the product, should these differences not be reconcilable the most restrictive policy shall prevail or the data having the most restrictive policy should be removed from the integrated product.

Distinction Between Data Integration and Systems Interoperability

[19] It should be noted that considerable effort was made to create a definition of the term "data authority" that would serve to determine what data is authoritative; however, this could change depending on the end-user. The challenge stems from the notion of "authority"; the authority will change depending on the atlas being developed. Using a controversial example, is the authority on whether an individual is HIV positive the person themselves, their physician, the lab that performed the analysis, or the regional health authority's Chief Medical Officer? Depending on who is determined to be the authority, the appropriate data source to create, for example, an HIV distribution map is the individual's public health record (Canada Health Infoway is developing such a record), their physician's patient records management system, one of several laboratory management systems that perform HIV testing for a region, or the regional health or provincial ministry's public health database.

[20] It was noted that some organizations, particularly large ones with major data investments, see data custodians as having a much greater mandate to support end users (see for example Data Custodianship Guidelines For The Natural Resource Sector, Integrated Land Management Bureau, Province of British Columbia, Draft Version 0.92, September 27, 2007), including providing the data custodian with the mandate to "ensure consistency of data management practices so that goals for integrated data can more readily be achieved". Given the infancy of the CGDI, it was felt that including such a mandate for CGDI data custodians would be premature; however, in the future one could anticipate a greater need for such a mandate as both demand for, and efforts to integrate, data rise.

Data integration initiatives seek to increase the value of regionally-scaled datasets by leading efforts and encouraging standardized approaches for horizontal integration of inter-jurisdictional regional geographic data to enable the information to be used most effectively. Project proponents will apply data standards that enable comprehensive integration of disparate datasets in a distributed environment. Data integration initiatives seek to ensure horizontal and vertical alignment of data with national framework data layers.

Over the course of the study, however, questions arose about whether or not data and/or systems were interoperable, and what the distinction was between systems integration and data integration. To help resolve some of this confusion, we propose the following definitions:

- *SYSTEMS integration*: an analytical activity, normally performed by SYSTEMS analysts, that defines how an application would OBTAIN data from possibly multiple data sources in possibly multiple data formats to meet user requirements.
- *DATA integration*: an analytical activity, normally performed by DATA analysts, that defines how an application would EMPLOY data from possibly multiple data sources in possibly multiple data formats to meet user requirements.
- *Data interoperability*: a data source is said to meet data interoperability requirements when the information required by a data analyst exists and is accessible for data integration purposes.
- *System interoperability*: a system providing access to a data source is said to meet system interoperability requirements when the information required by a systems analyst exists and is accessible for systems integration purposes.
- *CGDI data interoperability*: a data source is said to meet CGDI data interoperability requirements if all geospatial data is duly recorded in the CGDI Discovery Portal, all relevant CGDI approved metadata requirements are met, and all relevant CGDI endorsed data standards, if any, are supported.
- *CGDI system interoperability*: a system, or component of a system, providing access to a data source is said to meet CGDI system interoperability requirements when it supports all relevant CGDI service standards.

E. Good Practices Check List

This Appendix is a summary garnered from HAL's examination of good practices in regional scale information integration. The practices presented in Table E-1 were drawn from evidence collected through case studies of selected regional atlases, and validated by incorporating comments from each workshop into subsequent workshops.

Table E-1: Summary of Good CGDI Practices in Regional Scale Information Integration

#	Considerations	Practices for Technology Managers/Business Analysts	Technology Implementer/Programmer	Benefits	Cautions
1.	Develop and Endorse Data Standards	▪ Start if possible with data standards that are accepted within the user community (e.g. industry associations, data standards widely deployed in commercial applications)	▪ Recognize that a transformation from an existing to the desired standard may be required through scripts or other means.	▪ Designing a data standard meeting the needs of CGDI users can be a challenge; building upon the work of others may create greater buy-in even if the data standard has shortcomings	▪ If no such data standard exists, realize that there is a material risk that data in the application developed will not be easily integrated into applications of other CGDI users ▪ Do not impose a data standard as it may reduce likelihood of data sharing
2.	Prepare Geospatial Data	**Define External Data Requirements**	**Define External Data Requirements**	▪ Entering into a dialogue with other CGDI users	▪ Determining the data authority may be a challenge (e.g., the

#	Considerations	Practices for Technology Managers/Business Analysts	Technology Implementer/Programmer	Benefits	Cautions
	Profile	▪ Prepare profile of each dataset required (see data profile tool) to identify such information as the data authority and data custodian ▪ Consider the long-term sustainability of the data provider and extent of restrictive practices. **Define Local Data Offered** ▪ Prepare profile of each dataset in terms of above external data requirements ▪ Consider offering local datasets in a form that WMS, WFS and WCS can read. ▪ Place few if any constraints on the portrayal or access to the data (e.g. don't limit dimensions of requested maps) ▪ Consider providing direct access to the data layers by CGDI services (i.e. not via an application portal).	▪ Prepare profile of each dataset required (see data profile tool) to identify such information as the data access standard and version, and data styles and symbols. **Define Local Data Offered** ▪ Prepare profile of each dataset in terms of above external data requirements ▪ Implement an SLD service so client WMS can control the portrayal of the data that is rendered. ▪ If WMS server version supports this, use a Filter Enabled SLD service for the WMS client. ▪ If the WFS server version supports this, use a Filter Enabled WFS client. ▪ Datasets should normally be stored in a common projection ▪ Pre-process datasets to meet WMS, WFS and WCS functional requirements. o Tile raster data o Partition vector data o Other methods….	and data providers will improve the likelihood that application user needs will be met while leveraging significant investments in the CGDI ▪ Appropriate performance mechanisms will reduce WMS response problems when accessing large raster datasets.	data authority for regional school public health information could be the schools, one or more school boards, regional health authorities, etc.) To determine the CGDI data authority, balance application user needs, CGDI user needs, and extent data authorities embrace CGDI. Consider creating a regional node (see below) ▪ Watch for differences in versions of standards used by CGDI users and data providers (e.g., some WMS servers have partial implementations of SLD which may cause problems during data integration). ▪ Note WFS requires GML, and some organizations may have preference for GeoRSS or KML ▪ When displaying data layers, clearly identify the scale for each

#	Considerations	Practices for Technology Managers/Business Analysts	Technology Implementer/Programmer	Benefits	Cautions
					▪ WMS's getcapabilities returns from data providers do not always match the actual capacities of the WMS server. ▪ The OGC services should not be a constraint to the Regional Atlas portal. Access to the data layers by OGC services outside of the portal must be a priority using CGDI interoperability standards.
3.	Establish regional CGDI node	▪ Balance application and CGDI user needs for data consistency and data timeliness in the geographic coverage area to determine whether a regional node is required. ▪ Recognize that the data provider for complete data may not be the same as the data provider for timely data. For example, provinces will be the authority for	▪	▪ Establishing a regional CGDI node reduces the resource demands on other organizations to deliver data under CGDI (e.g., building a service-oriented architecture, responding to queries of CGDI users)	▪ Organizations that decide to be a regional CGDI node must recognize that their service may become integral to CGDI user applications that may have data quality[21] and timeliness requirements. ▪ Depending upon CGDI user requirements, regional nodes may encounter

[21] Many consider quality data to include the notion that it is timely, however in certain use cases separating these notions is important, for example when trade-offs must be made between selecting a data provider that delivers lower quality timely data over one that provides higher quality data that is available but with greater delay.

#	Considerations	Practices for Technology Managers/Business Analysts	Technology Implementer/Programmer	Benefits	Cautions
		complete road networks in a region. However, municipalities will be the authority for the most up to date road network in their region. ▪ Consider infrastructure capacity and sustainability before establishing a regional node		▪ A regional node may also facilitate the acceptance of a data standard, and acceleration of the adoption of CGDI. Of course, data providers may opt later to offer data according to CGDI-endorsed standards	an increase in requests for support from CGDI users and therefore a requirement for resources to support this. Recognize, however, that there may be a material improvement in meeting application and CGDI user needs overall. ▪ Use written data sharing/confidentiality agreements to help dispel fears about potential unauthorized access to confidential data.
4.	Build a Service-Oriented Architecture	▪ Under CGDI, organizations serving data become integral to the operational systems and services of other organizations serving their users. Consider building the application under an SOA as it provides the framework for managing inter-organizational dependencies.	▪ Register service on Catalogue Service Interface (CAT) as it is replacing Z39.50 ▪ Install the CAT service and link to the common data elements of ISO Dublin Core Metadata Element Set. Community specific profiles or geospatial based profiles can be used.	▪ SOAs provide a foundation for leveraging investments in open data standards, and the data itself, by facilitating data sharing both within and between organizations; such architectures reduce the need	▪ Beware of software patches that massage data and backdoors to transfer data. These may be symptoms of a number of possibilities, for example an inappropriate architecture (e.g., teleterminal software such as Citrix), a system component that is not CGDI compliant (e.g., java

#	Considerations	Practices for Technology Managers/Business Analysts	Technology Implementer/Programmer	Benefits	Cautions
				for patchworks of unique applications and procedures for providing and accessing geospatial data.	applets for transferring data between systems) or an organization that should be engaged but that is not (e.g., teleterminal and FTP interfaces to acquire data). ▪ Beware of tying the application under development to existing infrastructure that has limitations (e.g., the existing database may not support the full functionality of WFS).
5.	Establish a Data Integration Service Centre	▪ Based on anticipated service level demands, consider establishing a technical support service. Options to consider include Shared Services Centre, Policy Service Centre, and Guidance Service Centre. ▪ If the data provider is not delivering data through CGDI-endorsed standards, consider doing so through your service. ▪ Enter into Service Level Agreements to ensure a	▪ Document your data standard and make it available. ▪ Prepare other tools that will assist CGDI partners accessing your service (e.g., create instructions for installing basic WMS and other services for connecting to data on your system).	▪ See "Establish regional CGDI node" above	▪ See "Establish regional CGDI node" above

#	Considerations	Practices for Technology Managers/Business Analysts	Technology Implementer/Programmer	Benefits	Cautions
		common understanding of expectations between data users and providers. ▪ If the data provider is not delivering data through CGDI-endorsed standards, consider doing so through your service if this seems the only short term approach for enhancing the CGDI.			
6.	Undertake rapid prototype development	▪ Mitigate such risks as performance lags and data accessibility matters by preparing a use case for each dataset/dataset combination. ▪ Engage decision makers in testing the use case on a prototype system prior to the development of the operational system.	▪ Consider spatially enabling the database for faster query response time ▪ Consider using GML simple feature profile for simplicity. ▪ Check, and share lessons learned with, CGDI Developer's Corner, CGDI Developer's Guide and OGC cookbooks for vendor-specific information. ▪ Consider turning layers OFF and ON at appropriate scales, indexing the database, or limiting viewing area (e.g., by coordinates or by polygon) in order to reduce the processing time ▪ Consider implementing appropriate performance mechanisms which will reduce	▪ CGDI applications involve multiple organizations and multiple platforms: rapid prototyping helps confirm use case specifications will be met while ensuring data and service dependencies of external organizations are addressed.	▪ Be aware of limitations on CGDI-endorsed standards (e.g., there are many versions of WMS). Make sure the data providers have a version that is compatible with the selected WMS. ▪ Document and make available system specifications and be wary of "scope creep": this may reduce later demands from users of service.

#	Considerations	Practices for Technology Managers/Business Analysts	Technology Implementer/Programmer	Benefits	Cautions
			WMS response problems when accessing large raster datasets. ▪ Test-bed the application with different operating systems and browsers to determine which combination produces optimum performance		

F. Geospatial Data Profile Table

Table F-1 identifies the metadata that would assist GIS technicians, systems analysts, and systems architects in their efforts to scope, design and implement a regional atlas or decision support system that delivers and/or accesses CGDI data content. Among other purposes, this tool is intended to mitigate the likelihood of challenges being faced in regional-scale information integration before systems are constructed, and also may assist in establishing and maintaining technical linkages among CGDI participants.

Table F-1: Geospatial Data Profile

#	Meta Data	Data Layer 1	Data Layer 2…
1.	Layer Name for raster or vector dataset		
2.	Data authority name and contact info (authorizes access to data)		
3.	Data custodian has authorized access? (yes, no)		
4.	Data custodian name and contact info (contact for errors in data)		
5.	Registered in Discovery Portal? (Yes, No)		
6.	Has Catalog services Interface CAT service? (Yes, No)		
7.	Results of OGC GetCapabilities request for WMS, WFS and WCS (e.g. http://nsidc.org/cgi-bin/atlas_north?service=WMS&request=GetCapabilities) a. Native projection of data (e.g. EPSG:4326) b. Data projections supported (e.g. EPSG:4326, 32761, 3031…) c. Output format of request (e.g. WMS: gif, png…) d. Maximum allowable width and height of request (e.g. 2000x 2000) e. Bounding box of dataset (e.g. -180,-90,180,90)		
8.	Data standard (e.g. GeoBase: National Road Network, Version 2)		
9.	Data scale (e.g. 1:50,000)		
10.	Data accuracy (e.g. road centerline is +/- 10cm)		

#	Meta Data	Data Layer 1	Data Layer 2…
11.	Data currency (e.g. date data created and/or last updated)		
12.	Data completeness (e.g. data layer gaps in regional coverage)		
13.	Horizontal and/or vertical datum (e.g. NAD 83)		
14.	Data availability (e.g. 24/7, update monthly)		
15.	Data request response time (e.g. target 10 seconds FOB)		
16.	What is the raster data access service? (e.g. E-mail, CGDI service and version: WMS v1.2, client and server). Is WMS SLD enabled? Is SLD Filter Enabled?		
17.	What is the vector data access service? (e.g. FTP, CGDI service and version: WFS v1.1 client and server). Is WFS Filter Enabled?		
18.	What is the raster data access service? (e.g. FTP, CGDI service and version: WCS v1.1 client and server) a. Supported data formats (e.g. geotiffint16, geotiffFloat32) b. Supported interpolations (e.g. nearest neighbour, bilinear)		
19.	Other CGDI services (e.g. Web Map Content Document - WMC)		
20.	Is data layer served through cascading requests? (yes, no)		
21.	Service limitations (e.g. WMS limits dimensions of requested map, reprojection not possible, Filter Enabling not available, limits on number of requests)		
22.	User group access/update restrictions (e.g. none - public can view and update, municipal government employees only, must sign non-disclosure agreement to view…)		
23.	Custodian delegates user group access control to CGDI partners (Yes, No, or Yes, conditional on signing MOU)		
24.	Data authority has technical body that accepts external members? (Yes, No)		
25.	Until when is service assured? (e.g. planned at least until 2010)		
26.	Documentation available (e.g. data quality policy, data access policy, problem escalation protocol…)		
27.	Extent to which data access support services are available (e.g. custodian provides user and technical support, technical support only)		
28.	Notes (e.g. data has been pre-processed into 512 x 512 tiles for the following resolutions(…) in order to improve response time, examples of appropriate and inappropriate applications for the data)		

G. Selected References

Documents

A CGDI-Enabled Portal to Support Decision-Making for Community Health and Land Planning, User Requirements Specifications Document, Region of Waterloo, March 31, 2007

A CGDI-Enabled Portal to Support Decision-Making for Community Health and Land Planning, User Needs Assessment Report, Region of Waterloo, March 31, 2007

A Geospatial Portal to Facilitate Tourism Planning and Land Use Management by the COTA and CTA Community of Practice for Eeyou Istchee, Proposal, Cree Outfitting and Tourism Association and Cree Trappers Association,

Aboriginal Matters Road Map Draft Version 2.01, GeoConnections, January 2007

Best Practices in Best Practices, David Skyrme Associates. 2002 (From http://www.skyrme.com/updates/u54_f1.htm)

Data Integration in a Service-Oriented Architecture – White Paper", Informetrica, November 2005.

Environment and Sustainable Development Roadmap Draft Version 1.6, GeoConnections, June 14, 2007

Canadian Geospatial Data Infrastructure Architecture Description, v.1, December 11, 2001, CGDI Architecture Working Group

GeoConnections Interim Project Report, A Geospatial Portal to Facilitate Tourism Planning and Land Use Management by the COTA and CTA Community of Practice for Eeyou Istchee, April 30, 2007

GeoConnections 4th Interim and Final Project Report, Stewardship Tracking System Project, November 30, 2007

GeoConnections Milestone 2 Report: Functional and Technical Specifications Document, A Geospatial Portal to Facilitate Tourism Planning and Land Use Management by the COTA and CTA Community of Practice for Eeyou Istchee, January 22, 2007

Public Health Road Map Draft Version 1.0, GeoConnections, February 2007

Public Health Roadmap: Data Strategy, Version 2.0, GeoConnections, 2007

Public Safety & Security Road Map Draft Version 1.0, GeoConnections, March 2007Data Integration in a Service-Oriented Architecture – White Paper, Informetrica, November 2005."

The Dissemination Of Government Geographic Data In Canada - Guide To Best Practices Version 1.2, Tim Werschler - Statistics Canada, Julie Rancourt - Department of Justice, GeoConnections, Winter 2005

Websites

URL	Synopsis
www.iso.org	Provides numerous international standards relevant for to this study, including: ISO TS 19103:2005 provides rules and guidelines for the use of a conceptual schema language within the ISO geographic information standards. The chosen conceptual schema language is the Unified Modeling Language (UML). ISO TS 19103:2005 provides a profile of UML for use with geographic information. In addition, it provides guidelines on how UML should be used to create standardized geographic information and service models. ISO 19113:2002 Geographic information -- Quality principles ISO 19114:2003 Geographic information -- Quality evaluation procedures ISO/PDTS 19104 Principles for definition writing (ISO 704:2000)
http://www.gao.gov/archive/1995/ns95154.pdf	Demonstrates value of best practices, discusses best practices reviews, does not provide a structure for best practices.
http://www.computerpartner.nl/article.php?news=int&id=4032	Shows two kinds of best practices, well-defined processes (within study scope), and organizational design/governance (out of scope)
http://www.defenselink.mil/comptroller/icenter/learn/bestpracconcept.htm	Examines best practices through benchmarking/gap analysis, this is not approach in this study
http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/RiskManagement/rm-rcbp1_e.asp#_Toc456762775	Though the subject is not relevant (risk management), provides a framework for examining an area for best practices. Defines best practice ("A best practice is a strategy, approach, method, tool or technique which was particularly effective in helping an organization achieve its objectives for [change: managing risk to: REGIONAL SCALE DATA INTEGRATION]. A best practice is also one which is expected to be of value to other organizations. For example, a practice that was particularly helpful in establishing guidance would be of value to many other organizations, including the Treasury Board of Canada Secretariat (TBS) as the provision of guidance to federal departments is one of their important objectives."
http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/RiskManagement/rm-pps2_e.asp#_Toc456660356	Provides "Criteria For Assessing Applicability Of Best Practices To The Canadian Federal Government", a small number are relevant to this study.
http://www.just4kids.org/en/files/Publication-Twenty_States_Best_Practice_Framework-07-14-06.pdf	Another example of a best practices framework
http://isotc.iso.org/livelink/livelink/fetch/2000/2489/Ittf_Home/ITTF.htm	While beyond the scope of this study, this is the ISO template for a standard and a standard is a type of best practice

click on "Examples" and then on " ISO/IEC 12345 (E)".	