CISE ARCHITECTURE OVERVIEW

The CISE architecture document is intended to introduce the CISE technical architecture to data providers, service providers, and application developers so that they can gain an overall understanding of the services and related aspects of CISE. It provides a top end overview of the architecture, discussing the range of technology services that constitute the foundation of CISE.

It also serves as a context piece for the particular services and cross cutting themes that are being developed through the CISE projects, including metadata, Web mapping and Web services.

1 INTRODUCTION

The Task Force on the Canadian Information System for the Environment (CISE) was created as part of the Budget 2000 investment in environment and sustainable development indicators to address the need to more fully integrate economic and environmental policy decisions.

The Task Force was mandated to advise the Minister of the Environment on the design and implementation of an integrated environmental knowledge management system in support of this goal. The Task Force recommendations called for a national, more shared and strategic approach to collecting, managing, assessing and disseminating environmental information. A CISE Secretariat is now engaged in the implementation of these recommendations.

To be successful, CISE will be designed as an inclusive system, engaging data providers as well as end users in order to meet the needs of both communities. CISE will be built from the ground up as a collaborative system linking federal, provincial, municipal, industry, NGO and academic sources in response to societal priorities. This collaborative structure will be reflected in every aspect of CISE's operation from governance through to technical implementation.

From a technical perspective, the Canadian Information System for the Environment is envisioned as:

a) an authoritative online environmental portal that will provide Canadians with intuitive access to a distributed set of data, services and applications packaged to allow them to make informed environmental decisions and choices.
b) bundled web service and application components to enable partners to provide customized portal services to various clientele.

To accomplish these information management/information technology and communications goals, the CISE architecture must be developed in a way that recognizes and supports,

- specific means of strengthening regular public reporting on environmental issues, trends and progress;
- the necessity of providing easy access to the knowledge system by all users, subject to privacy and property interests, in ways suiting the needs of each type of user;
- the appropriate institutional arrangements for development, storage and dissemination of knowledge in the system to ensure long-term institutional integrity, credibility, neutrality, relevance and capacity, including an assessment of linkages between and amongst the various institutions that might house elements of the system;
- the ability to integrate and render coherent this vast array of distributed data, which remain accessible as seamless databases via the network while being maintained locally as component parts;
- the information technology and management software and hardware required to provide this coordinated access while preserving local control;
- the human capital requirements, including the skills and knowledge to develop, operate and use CISE, as well as training needs to ensure needed skills are available throughout CISE's network of partners; and
- the adaptability to respect sensitive data and the particular needs of traditional ecological knowledge by providing authentication and brokerage systems where necessary.

Through the architecture presented below CISE will have the ability to support service providers, data providers and application developers, using linked services and re-usable components. In short, CISE adopts an 'integrate, don't duplicate' philosophy. The architecture will focus on the development of underlying key services, establishing a service chaining delivery system. CISE will rely on strong partnerships, working collaboratively to build from the significant investments and leading work of other agencies such as Natural Resources Canada and international standards organizations. In order for this to happen, it is essential that Environment Canada focus on areas of authority in the environmental arena and link to other authority data and systems through the adoption of recognized standards such as Open Geographic Consortium (OGC) and ISO 19100.

2 ARCHITECTURAL OVERVIEW OF CISE

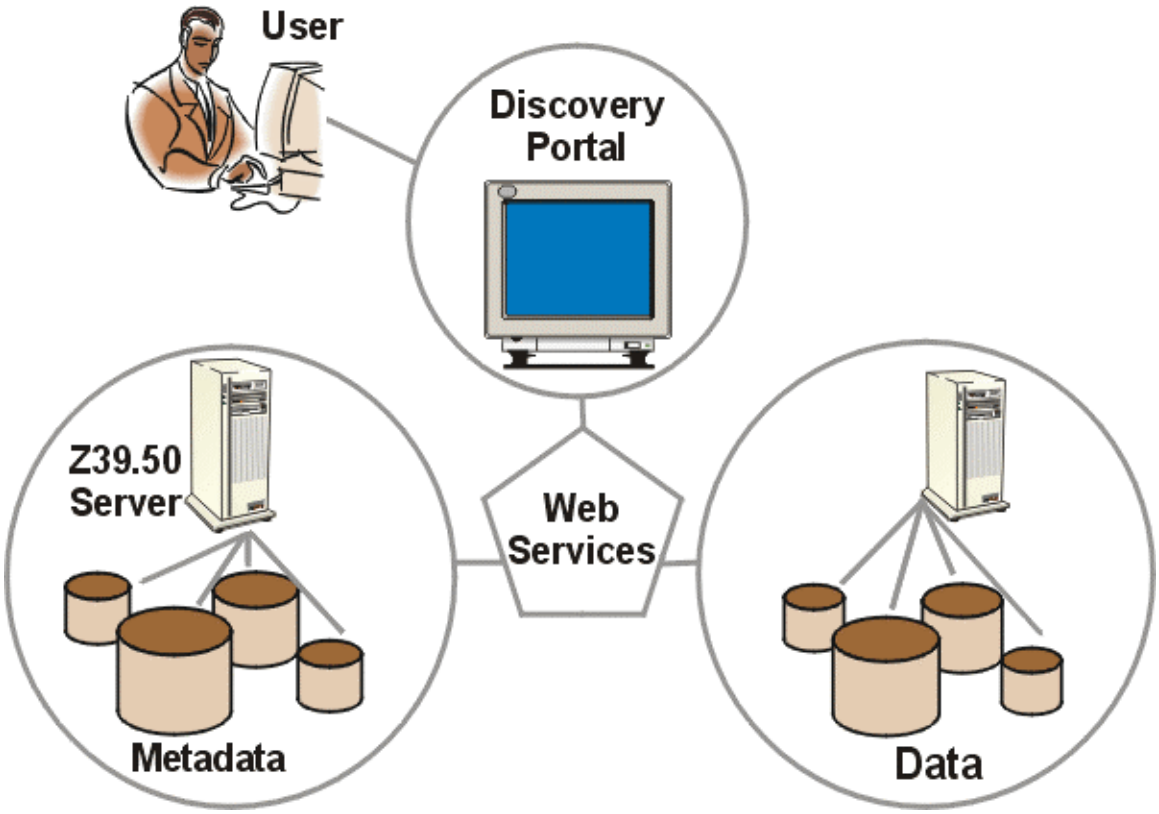
2.1 CISE Basic Architecture

Most environmental Web sites are organized thematically, providing users with access to information on a variety of topics such as water quality, water quantity, toxic chemicals, weather or birds. Although useful, this approach often fails to provide users with the breadth of knowledge required make decisions on complex environmental issues. The primary motivating factor behind the development of CISE is to respond to Canadians' 'right to know' about the quality of their environment by providing them with ready access to integrated and timely data and knowledge, empowering them to initiate change through their own choices and by influencing others. Moving from data to empowered knowledge and from sectoral to holistic perspectives requires the development of a new design and delivery paradigm. The underlying architecture of CISE must support this new paradigm, a paradigm that focuses on providing answers to the complex real world problems of today.

The basic architecture of CISE (Figure 1) consists of a network of data, knowledge and service providers. Through the use standards and specialized services users will be provided with a seamless path to the information that they need. The infrastructure will provide clients with access to information and will process and package the client feedback in an appropriate format. CISE will be built upon federal, provincial, territorial, municipal, and industrial environmental data and initiatives, and is intended to be a collaborative knowledge framework that links dynamically with other systems in Canada and beyond.

Call out:
CISE will be built upon federal, provincial, territorial, municipal, and industrial environmental data and initiatives, and is intended to be a collaborative knowledge framework that links dynamically with other systems in Canada and beyond.

Figure 1. Basic CISE Architecture



2.2 *Characteristics of the Canadian Information System for the Environment*

In order to accomplish the CISE vision, the architecture will:

a) enable easy access to the knowledge system by all users and be customizable to suit the needs of each user;
b) enable applications to discover, access and share remote online information through a distributed infrastructure;
c) enable integration of disparate information to provide Canadians with timely and relevant answers to environmental questions and issues;
d) lower barriers to content provision by providing easy solutions for mounting data sets while preserving local control and establishing authentication where necessary;
e) enable the seamless chaining of applications, data and services or combinations of these through a consistent portal;
f) provide these components and documentation as building blocks, empowering partners to build their own customized portals for unforeseen applications;
g) strengthen collaborative activities that support regular public reporting on environmental issues, trends and progress;
h) enable wide-scale interoperability by adhering to common and open information standards and specifications; and
i) facilitate the development of effective partnerships with regional, national, and international data authorities and technology development leaders.

2.3 Framework and Specific Client Data

CISE must be designed to allow seamless access to data that can be separated into two distinct categories; framework data and specific client data.

Framework data provides a context for other data by providing a common geographical reference for a country. For example, roads, water features and other information found on the National Topographic Map Series produced by Natural Resources Canada are commonly used as framework data by numerous mapping applications.

Specific client data is focused on a particular theme related to the mandate of a specific organization and is usually presented in relation to framework data. For example, client data on water quality monitoring stations may be presented on watershed boundary framework data.

These two data categories need to work together in an integrated and seamless fashion. Framework data may be used solely to provide a geographic reference or backdrop or may be extended to provide spatial querying and selection capability. For example, we may wish to select all the water quality stations within a particular watershed or find out if lakes near paved roads have higher salt content. To deal with these examples, CISE will need to access NTS map series framework data from Natural Resources Canada, watershed boundary framework data from Environment Canada, and client water quality data from Environment Canada, provincial and municipal agencies and possibly citizen based water quality monitoring projects.

Obviously, many agencies and organizations are involved in collecting the framework and client data required to answer almost any environmental question. Environment Canada is authoritative on several core environmental data sources including:

- Water Quality Monitoring Data
- Water Quantity Monitoring Data
- Migratory Bird Abundance and Distribution Data
- Species at Risk Distributions
- Weather and Climate Data
- Contaminants and Toxic Chemical Data
- Watershed and Ecological boundaries
- Critical Habitat Delineations

While CISE will pay particular attention on ensuring that Environment Canada's authority data holdings are made available to Canadians, it will also focus on developing linkages with authority data held by other organizations.

2.4 Services

The architectural design of CISE will be based on the provision of a core set of services that promote seamless connection to a wide range of distributed databases and information sources. These services will be invoked by every interaction that a user has with CISE through the use of the CISE portal, partner portals or, in some cases, specialized client software applications. In essence, CISE will be based on a series of loosely coupled, distributed and reusable services. Complex applications will be constructed from simple services that can be developed independently and chained together. These services will add to an ever expanding pool of specific tools that can be shared and re-used by others in new, innovative solutions.

CISE will be designed around three core service sets:

1. Metadata Services
2. Web Services for Data Access, Extraction and Analysis
3. Web Mapping Services

2.4.1 Metadata Services

The key to implementing a distributed knowledge system that provides seamless access to information held by many authorities, is the development of a consistent and efficient metadata reference system. By implementing a set of nested metadata/data standards including Dublin Core, Darwin Core and *CSDGM*, a series of "XML" based collections will be established that will document the numerous data holdings that comprise CISE. The use of recognized standards, in concert with well-defined taxonomies and thesauri will provide the necessary "connective tissue" to facilitate interoperability across numerous data systems. Thus a data search initiated by CISE will be able to discover relevant data holdings in other systems currently resident on the Internet.

CISE will facilitate the development of content rich metadata collections through the development and use of a number of tools (e.g. XCHAINJ, Meta Star, Metatagger, MCAT) that will help individual data owners to create and publish metadata in a consistent manner. Although these tools may differ from those being used by other agencies, adherence to accepted standards will ensure interoperability of the final metadata collections.

Although metadata is seen as a foundation service for CISE, it is quite likely it will be transparent to the end user. Data providers will be involved in the creation of their own metadata collections, whereas the CISE user will probably only see the results of a metadata search automatically invoked by the CISE portal. In this context, the metadata services provide the road map and compass to the Internet data network.

2.4.2 Web Services

As platforms and data systems become more diverse and highly distributed, technologies like XML and SOAP become critical new tools - bundled as 'middleware' to connect the content (databases) to the user (portals). XML is skilled at handling structured or semi-structured data while SOAP is used to implement robust and highly portable services. This combination of XML and SOAP enhances efficiency and knowledge sharing. The benefits of these technologies becomes clear when they are deployed within a truly distributed infrastructure such as that required by CISE. By using XML and SOAP new services can be rapidly developed based on existing applications, they distributed across numerous servers and can be easily re-used by others. Creation of new services will be highly flexible. They can be linked to existing services or take advantage of new services as they emerge.

Although XML and SOAP based services are relatively new, there are already proven examples of their functionality and flexibility in the environmental field. Recent data extraction, management and evaluation applications developed for the National Air Pollutants Surveillance Network (NAPS) have been designed using web services. The new version of Species Analyst (known as DiGiR) has also been developed using a web service approach.

The distributed nature of CISE and its approach to rely on the developments of other systems, make it a logical candidate for the implementation of a web services approach. It is envisaged that web services will be developed to provide a number of specialized services including metadata searching, linking metadata records to data records, performing data extractions and conversions, developing interpretive product and running decision support systems. Re-usable services that are already available in other systems such as the National Atlas and the Discovery Portal can be integrated into CISE creating synergy and avoiding duplication of effort.

Finally, CISE web services will empower participants to link into the international GRID system. Sophisticated solutions, such as the San Diego Supercomputing Center's Service Resource Broker, use XML/SOAP in peer to peer transactional services which incorporate authentication. Such brokering enables the public to have open access to non-sensitive records while simultaneously users authenticated directly by providers can access sensitive records. Such future

development of bi-directional web services coupled with security and authentication protocols will allow for direct client input/extraction into dynamic database systems.

2.4.3 Web Mapping Services

Canadians are influenced daily by spatial references and displays ranging in complexity from simple highway directional maps to more complex multi-dimensional presentations. Thinking spatially is truly ingrained in day to day life and therefore the development of web mapping services is an important component of CISE for the presentation and visualization of information.

Over the past few years there have been dramatic advances in the functionality of web based mapping systems and there are currently numerous on-line implementations of dynamic mapping services. Since 1999, the Open GIS Consortium (OGC) has been working with software vendors and other groups to develop open Web-based mapping specifications. Using OGC specifications, software vendors can develop open inter-operable interfaces that will provide clients with access to a large pool of Web based tools for access, display and manipulation of spatial data.

Although complete OGC compliance is yet to be a reality, a large number of vendors are investing in OGC development. CISE recognizes the need to support the overall principles of OGC development. As such, the web mapping services used by CISE will rely on vendors and applications that are committed to the ultimate realization of the OGC standards and specifications.

Web mapping services will be a core component of both the CISE system and client interface. Mapping coupled with specialized services such as Postal Code and Place Name searching will allow users to quickly zoom to their geographic area of interest and launch behind the scenes data searches and extractions. Web mapping services will also serve as a spatial presentation tool, showing metadata query results, data extraction results, leading to the presentation of dynamic custom reporting.

2.5 Client Services - My CISE Portal

Beyond the building blocks of services, CISE will also be developing a “My CISE” client application that will allow users to define and customize their connection to services. CISE will be servicing a wide spectrum of users ranging from generalists to specialists. These users will all have special requirements ranging from the extraction of relevant data for future technical analysis to the provision of dynamic, customized interpretive reports answering particular environmental questions.

The concept of “My CISE” will allow users to modify their view into CISE. Although they will always have access to the generic CISE portal they will be able to customize the portal to meet their needs. This may include focusing to particular geographic areas, selecting areas of environmental interest, automated assignment of user expertise levels and controlling instant links to particular services. The My.NBII portal developed using Plumtree software is presented in Figure 2 and may serve as a model for the “My CISE” portal.

Figure 2: A view of the MyNBII Portal

2.6 Security and Authentication

Security and authentication services will need to be developed for CISE in order to support the My CISE concept and to potentially allow users to have live input into underlying environmental databases. Even a simple service such as displaying data through a web mapping service may have some security implications related to data sensitivity and confidentiality. Wherever possible CISE will use available standards and security and authentication services.

3 ARCHITECTURE DEVELOPMENT STRATEGY

Organizing for Success

The CISE architecture is being developed primarily under the direction of the CISE Secretariat, the CISE Cross Cutting Issue Working Group and four thematic development teams concentrating on infrastructure, metadata, web services and web mapping services. In order to capture the potential synergy across the CISE projects, these cross cutting teams were established to oversee the development of standard services and approaches in those areas that were common to all projects.

The CISE projects will serve as a test-bed for the proposed architectural design presented above. By working across the water, air, biodiversity and wetlands projects, the flexibility and functionality of the proposed CISE architecture can be evaluated. The projects will become showcases for the CISE vision and will market the value and synergy that can be achieved by working collaboratively.