22 July 2002

Original: English

Eighth United Nations Conference on the

Standardization of Geographical Names

Berlin, 27 August-5 September 2002

Item 12 (c) of the provisional agenda"

TOPONYMIC DATA FILES: TOPONYMIC DATA TRANSFER STANDARDS AND FORMATS

Developing the Canadian Geographical Names Service (CGNS)

(Submitted by Canada)* *

DEVELOPING THE CANADIAN GEOGRAPHICAL NAMES SERVICE (CGNS)

Paper submitted by Canada

The Canadian Geographical Names Data Base (CGNDB) contains more tliat 510,000 records, about two-thirds of whicli are currently official naines, as approved by the Geographical Nanies Board of Canada. Geographical nanies are tlie focal point of tlie CGNDB with each nanie record having a unique identifier. Such geo-referenced data form a valuable search tool wlien linked to other databases. To take advantage of today's technology. the CGNDB is evolving in order to participate in a national initiative, the Cariadian Geospatial Data Infrastructure. The basis of this initiative is to coordiiiate Canada's numerous databases of geographic information and to make this information accessible on the Internet. A Canadian Geograpliical Naines Service (CGNS) is being developed to address the toponyniic aspects of the CGDI. The CGNS will be the next generation of techiiology for the distribution of Canadian geographical nanies.

Introduction

Geographical nanies are the focal point of the Canadian Geograpliical Names Data Base (CGNDB) with each name record having a unique identifier and at least one set of coordiiiates. Such geo-referenced toponymic records can form a valuable tool for searching wlien liiiked to other databases. The Canadian Geospatial Data Infrastructure (CGDI) is a national initiative to coordinate Canada's numerous databases of geograpliical information and to niake this information accessible on the Internet. A Canadian Geographical Nanies Service (CGNS) is being developed to address the toponymic aspects of the CGDI. The CGNS will be the next geiieration of techiiology for the distribution of Canadian geograpliical naines. Geographical names are considered an intuitive spatial reference and a basic component or fundamental layer of the country's framework data. These data will be incorporated into tlie Canadian Geospatial Data Infrastructure.

Background

The Canadian Geographical Nanies Data Base (CGNDB) contains more than 510,000 records, about two-tliirds of whicli are current official names for places and geographical features, as approved by the Geographical Naines Board of Canada (GNBC) and its predecessors. This database is inaintained by the Geographical Nanies Section of tlie Centre for Topographic Information, Geomatics Canada, Natural Resources Canada. It is a collection of the official geographical name decisions made by the various federal, provincial and territorial naming authorities of the Geographical Names Board of Canada which are forwarded to Natural Resources Canada for input into the CGNDR. The CGNDB is the fundamental national database to provide official names for mapping and charting, gazetteer production, World Wide Web refereiice, and other geo-refereiiced digital systems. At present, more than 30 attributes may be stored for any name. This

authoritative national collection of Canada's geographical names, with records dating back to 1897, is updated on a daily basis.

The forerunner of the CGNDB (the National Toponyinic Data Base) was developed in 1978 as a replacement for a growing card-index registry, which had been maintained since the creation of the original Geographic Board of Canada in 1897. The digital database was originally designed to increase the efficiency of gazetteer production and names compilation for the National Topographic Series (NTS). In 1987, the digital database was remodeled into a relational database and, in 1999, the database was re-engineered to its current form.

The GeoNames Web site was launched on the Internet in 1994 and has become a key promotional tool to disseminate geographical names information to GNBC members, other government agencies, and the general public. The number of accesses to our www site <http:// geonames.NRCan.gc.ca> has risen dramatically from 350 requests/day in the first few months of operation in 1994 to a current 5,300+ requests/day for documents, geographic location, name origins, and other information. In addition, tlie number of queries against the CGNDB itself has reached an average of 6,588 requests/day primarily for use in database and mapping programmes.

The status quo

The CGNDB is managed with ORACLE RDBMS software (Version 8.0.4), running on an ULTRASPARC IIi, using the operating system SOLARIS 2.6. The use of the character set ISO 8859 ensures the inclusion of accented characters used in Canada's French-language geographical names. Characters that include diacritics found in some Canadian Aboriginal-language names are presently beyond the scope of most computer systeins and are represented in the CGNDB by numerical substitutions. When Canadian and international standards are developed, accepted, and implemented, the CGNDB records will be modified accordingly.

The geographical name itself is the focal point of the CGNDB with each naine record having a unique identifier, an alphabetical sequence code consisting of five letters assigned to each record in the database. Such geo-referenced records form a valuable search tool when linked to other federal, provincial and territorial databases. Direct linkages have been established with related databases in several federal and provincial/territorial governments. The Government of Canada is presently using the CGNDB records as the official authority of geographical naines to be used as a reference for those filing environmental impact reports, now required by law. Work continues to associate the CGNDB toponymy with digital geospatial data of the National Topographic Data Base (soon to be the Geospatial Database (GDB)), managed by the Centre for Topographic Information Sherbrooke, and to improve links between various government departmental data bases, in the broader context of developing a Canadian Geospatial Data Infrastructure (CGDI). The source of geographical names data is not changing, but rather, the way of uploading, storing, and accessing this data is rapidly evolving.

The time had come in the area of geographical names to move from a names-driven database to one that is geospatially-driven by directly linking to feature geometry, and thus, entering into a more digitally-based work environment. In 2000, the technology to combine the best of today's technology, allowed the CGNDB to become part of the Canadian Geospatial Data Infrastructure (CGDI), a national initiative. (See the web site <http://www.geoconnections.org> for more information.) The basis of this initiative is to coordinate Canada's numerous databases of geographical information and *to* make this information accessible on the Internet.   In order to do this, a national standard logical model for toponymic databases had to be developed to incorporate the data drawn from the various provincial and territorial authorities. The CGNDR is not a datübase of federal records, rather a national focal point for storing and distributing Canadian geographical names data.

The purpose of the development of a standard is threefold:

- to provide a national standard for GIS-enabled toponymic databases ready to provide digital extent information compatible with the CGDI,
- within different federal, provincial, and territorial government jurisdictions build linkages froin their databases with standardized logical models to select toponymic framework layer data from within each of their respective jurisdictions, and
- once the linkages are developed, integrate the databases and thus build accurate toponymic data into the fundamental layers of the CGDI.

The first phase of developing a toponymic standard model was to analyze the different toponymic databases belonging to the federal, provincial, and territorial agencies and other naming authorities in ternis of their present and anticipated future needs for toponymic data within a digital GIS environment working towards a national standard.

The goals of the Canadian Geospatial Data Infrastructure (CGDI) are to coordinate Canada's numerous databases of geographical information, to enable framework data to offer a more complete, revealing, and accurate picture of the Canadian environment, and to eiiable organizations to remain autonomous while working together to provide easy, consistent and harmonized access to geographical information and services.
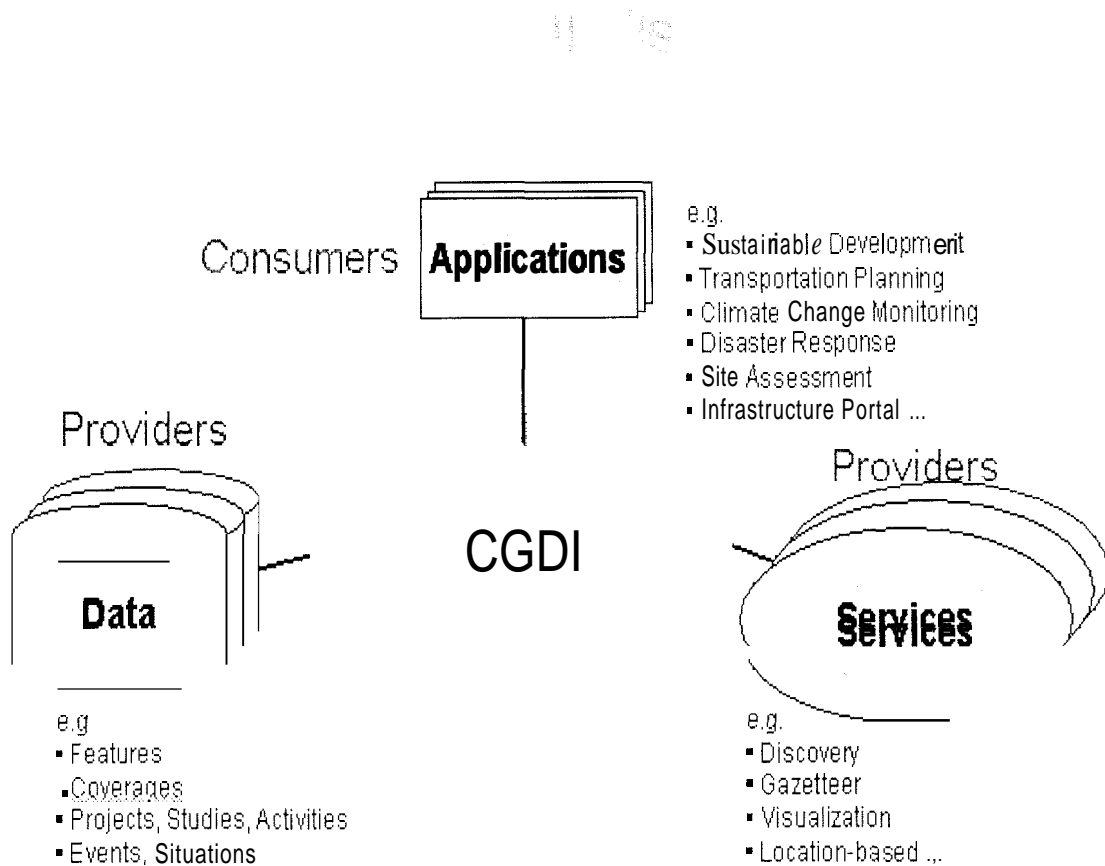
This national framework is the infrastructure required to provide the geographic data sets about Canada based upon a common reference system, and it will enable the development of related applications and value-added services.

In today's GIS-enabled world, we need accurate toponyms, linked to GIS data systems providing digital feature extent information in order to realize the full benefit of the toponymic data already available. In order to achieve this goal, a Toponymy Framework Project was initiated to develop a national standard for toponyinic data bases that would ensure that the needs of the data providers were addrcssed and that allowed for the integration of this data into the fundamental layers of the CGDI from within the dìffèrent jurisdictions. **A** review of standards, from such organizations as the International Organization of Standards, Technical Committee 211 on Geomatics (ISO TC 211) which is compliant with the Open GIS Consortium (OGC) [1] was undertaken to ensure what is developed nationally will be compliant with other national and international standards. OGC's mission is to give the world's information systeins a new connection to physical reality by making geo-referenced data behave like just another standard data type in systems of all kinds <www.opengis.org>.

The implementation strategy for the Toponymy Framework Project is an evolutionary approach that supports the CGNS with the minimum impact on existing systems. It also accommodates uncertainty in a number of areas: data providers will come on board on their own timetable: CGDI users and their requirements have not all been identified: OGC technologies continue to mature offering new and improved capabilities: and finally, the CGDI is very young and will take some time to find its final form. This leading edge initiative will influence and will be influenced by the direction and progress of others. In order to succeed, the CGNS must be flexible and strong to keep its place in this sea of change. The implementation strategy includes a pliased development of core capability, prototypes, workshops, pilots, integration support for the data provider and data user communities, and promotion of the CGNS within the CGDI development comniunity.

Core capabilities have been organized and prioritized under five headings:
* "Names" which support geographical name record transactions into and out of the CGNS
* "Maps" enables names to be generated for maps by the Web Feature Server (WFS)
* "Geometry" will support a direct link via a Feature ID to other databases that store the true geometry of each feature
* "Syndication" describes the fully distributed environment in which names are served directly by the provincial/territorial/federal authority
* "Collaboration" develops the web applications and processes for the provinces that want to replace their current systems or wish to enable participation of external agencies in the geographical naming process.

---

[1] The Open Consortium (OGC) is a non-for-profit membership organization founded in 1994 to address the iack of interoperability among systems that process geo-referenced data, and between these systems and inainstream computing systems.

In its final state, the plan is that the CGNS will be a distribution centre for a cascading iietwork of provincial/territorial/federal Geographical Names Services. The CGNS will be a distributed Names Service operated by the Geographical Names Section, of Natural Resources Canada <http://cgns.nrcan.gc.ca> and will be hosted on a server in Ottawa. The CGNS will be compliant with the Web Map Server (WMS) and Web Feature Server (WFS) specifications of the OGC. The WFS will deliver geographical names as data records formatted in SML/GML. The WMS will deliver an image of geo-referenced labels (geographical names) that can be used to overlay other map layers in the construction of a web map.

The toponymic data delivered to the CGNS is spatially enabled – presently as points, later to be stored in the database or via link(s) to other databases and will support a variety of query techniques for finding and filtering the names data. Given the differeiices in technology amongst the various GNBC jurisdictions and the readiness seen across the country, this will take time to achieve. To satisfy the immediate requirements of the CGDI community, a flexible approach for data uptake has been proposed. A CGNS warehouse has been created as a central data store that each province can use until their Names Service is installed. A web transaction process will be defined through which each province can load their names into the warehouse. Provincial systems will have to be extended with a capability to write transactions to the CGNS WFS. In advance of this, the CGNS implementation team will develop this capability for the CGNDB. This will allow an immediate flow of all naines from the CGNDB to the CGNS. Using a combination of data flows from the various provincial/territorial databases, the CGNS will provide a national view of names, and will support the transition timetable of each province as we work towards the full implementation of the CGNS.
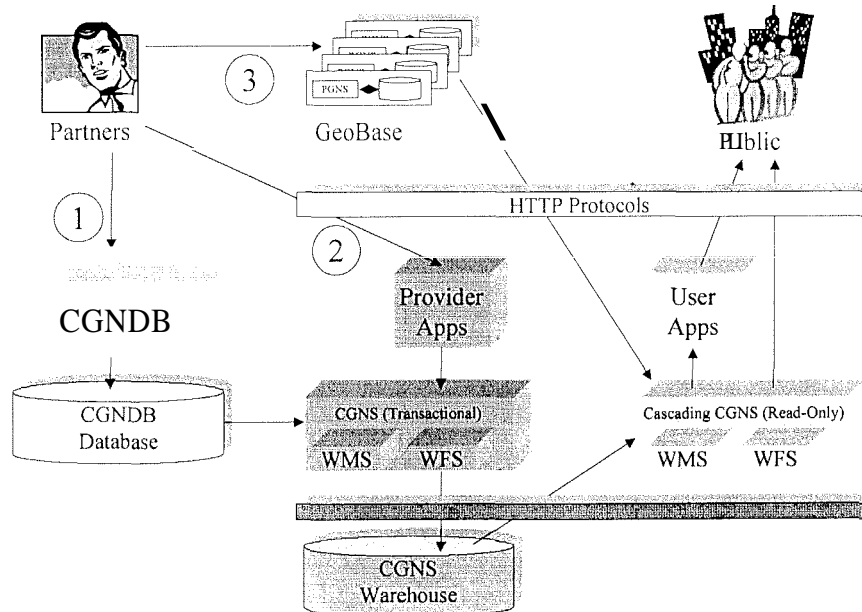
Holonics Data Management Group Ltd

1. Toponymic data transfer from Providers (GNRC) to CGNDB to CGNS warehouse
2. Toponymic data transfer from Providers (GNBC) directly to CGNS warehouse
3. Toponymic data transfer from Providers (GNBC) to a cascading CGNS

To ensure comprehension and support of the CGNS initiative, a Data Providers Workshop was held in Deceniber 2001 for Geographical Names Section staff and members of the Geographicai Names Board of Canada and their technical assistants. The workshop was designed to educate the provincial/territorial/federal representatives about the CGNS, to define a data upload process from provincial databases to the CGNDB/CGNS, and to start a CGNS community. 'The workshop was organized to meet the needs of both the business and technical communities within the group. Copies of all presentations and workshop suinmaries as well as a variety of documentation and prototypes concerning this project are available by accessing the web site: <http://cgns.holonics.ca/>.

As part of the implementation strategy, a Pilot Project was initiated with the Province of Manitoba in Deceniber 2001 with a planned completion date of March 2002. The purpose of the pilot was to denionstrate the data upload function, to uncover the issues related to security and serving naines to the CGNS, to create an opportunity to explore the general CGNS capabilities, to use the results of the pilot to update the Data Upload Process/CGNS Architecture Documents, and finally to support a communication and promotion strategy for the CGNS. Manitoba was chosen as it has a geographical nanies programme capable of moving quickly, their profile is representative of the technology seen in Toponymic Units across the country, the names unit lias direct access to its department's mapping organization, and its staff are highly committed to the success of the CGNS.

The future

It is hoped that the other provincial/territorial data providers will follow suit as their technology and resources permit. The investments needed to successfully participate in this national initiative are the necessary human and financial resources, up-to-date technology, continued commitment to data quality, and a determination to raise the visibility of toponymy, nationally and globally.

The development and implementation of the CGNS will also facilitate our participation in GeoBase. GeoBase, the National Framework Data Base, is governed by the Canadian Council on Geomatics, and is the primary responsibility of the Federal Government of Canada to populate and disseminate. The intention of this "quality base" of geographical information is to provide reference and context to a wide variety of theniatic data and applications to enable users to map Canada through the cooperative acquistion, sharing, integration, and effective use of quality geographical information across Canada for all Canadians. GeoBase has been defined as the fundamental geographical information that describes the Canadian landmass above and below water. The GeoBase data adheres to the concepts and definitions of the CGDI framework data. Collaboration has been effected from the beginning through a set of well-defined roles and responsibilities between Federal and the Provincial-Territorial Governments.[2]

---

[2] Canadian Council on Geomatics Working Group on "Base Data Quality Issue" A Vision for Quality Base Geographic Information in Canada "GeoBase" January 8, 2002.

We are at the leading edge of these initiatives, and will hopefully learn by other's efforts to integrate these new technologies.

Useful Websites

\<http://geonames.NRCan.gc.ca\>
\<http://cgns.holonics.ca\>
\<http://geonnections.org\>
\<http://geoconnections.org/english/tap/index.html\>
\<http://geoconnections.org/english/framework.index.html\>
\<http://geoconnections.org/english/access/index.html\>
\<http://maps.nrcan.gc.ca/HAL_report\>
\<http://www.opengis.org\>
\<http://www.cubewerx.com\>

References

Open GIS, 1999, OGC Consortium, Inc., web site \<http://www.opengis.org\>.

An Executive Overview of the CGNS Initiative, Michael Panyszak, Holonics Data Management Group Ltd., November 2001, contract deliverable - Toponymic Data Architecture Standard for the CGDI.

The Canadian Geographical Names Data Base (CGNDB) in 2001, Geographical Names Section, Geomatics Canada, Natural Resources Canada, *Canoma* Vol. 27, No. 1, July 2001.