

**Twenty-ninth session
Bangkok, Thailand, 25 – 29 April 2016**

Item 9 of the Provisional Agenda

Activities Relating to the Working Group on Toponymic Data Files and Gazetteers

Advances in Spatial Data Management in Canada's National Geographical Names Database

Submitted by Canada*

* Prepared by Kristina Kwiatkowski, Toponymy Specialist, Geographical Names Board of Canada Secretariat, Natural Resources Canada (Canada)

Advances in Spatial Data Management in Canada's National Geographical Names Database

SUMMARY

Over the past two years, the Canadian Geographical Names Data Base (CGNDB), Canada's authoritative national names data base, has undergone significant technical and structural advances. This paper highlights the migration of the CGNDB data model from an attribute-based model to a geospatial-based model, the management of toponymic relationships within the CGNDB, development of a new web-based application to access the database, and the dissemination of a suite of Open Data products from the CGNDB.

Background

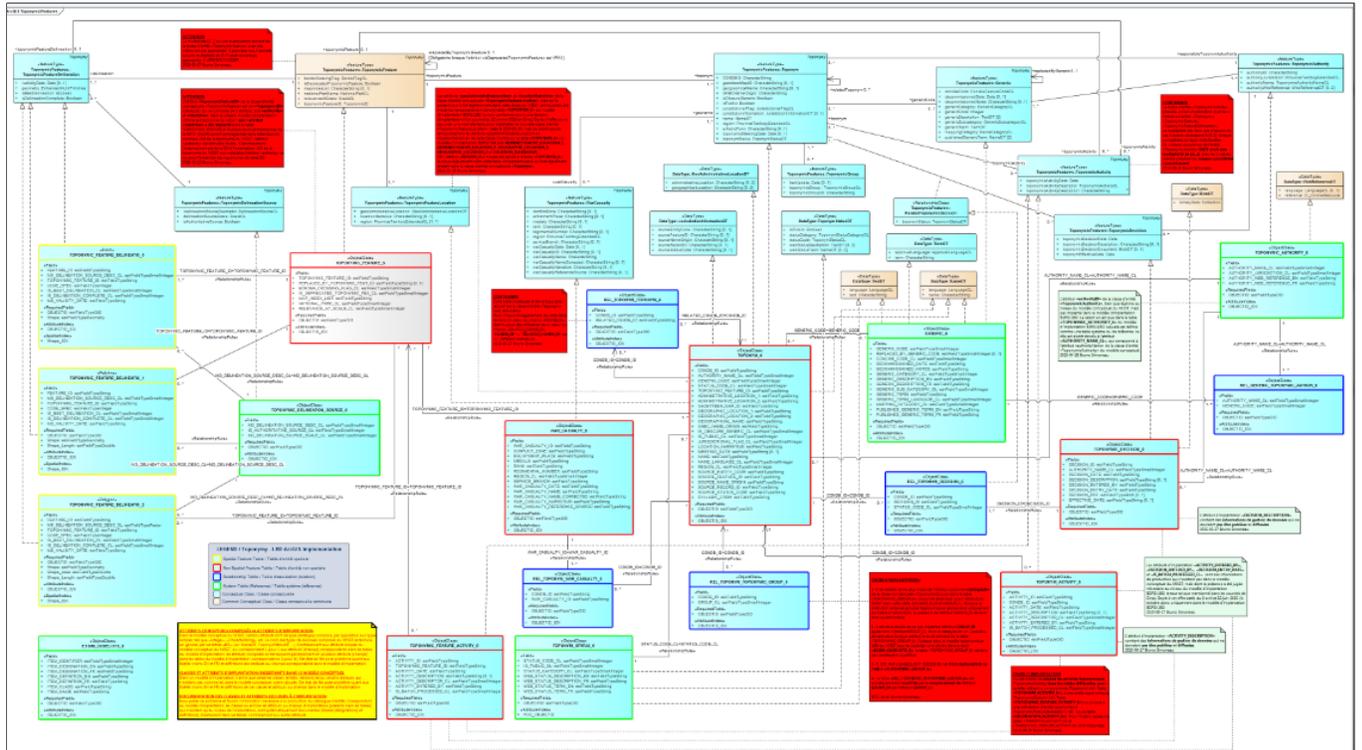
The Geographical Names Board of Canada (GNBC) is Canada's national coordinating body responsible for standards and policies on place names. The GNBC comprises federal, provincial and territorial departments and agencies, each with specific authority and responsibility for their respective jurisdictions. Working together as a multi-jurisdictional national body, GNBC members coordinate efforts to ensure that geographical names are consistently managed.

The GNBC is supported by a Secretariat provided by Natural Resources Canada (NRCan), a department of the Government of Canada. NRCan provides infrastructure and support for the Canadian Geographical Names Data Base (CGNDB) as the authoritative national data base of place names, and a key component of Canada's Spatial Data Infrastructure. NRCan consolidates geographical names data, spatial delineations of features, and new naming decisions provided by the naming authorities of the GNBC in the national database. As of February 2016, over 550 000 place names are managed in the CGNDB, 390 000 of which are official names.

Data Model

In 2015, the data model for the CGNDB underwent a fundamental transition, evolving from an attribute-based model to a more efficient and interoperable geospatial-based model. The new data model was designed using a relational ISO standard model and vastly improves the functionality and interoperability of the national database. The previous data model contained one large table with over 90 fields. The new geospatial data model contains 45 tables, and is capable of handling spatial and relational data. The transformation of the data model enables relationships between toponyms and spatial delineations for features. There are over 133 000 spatial delineations contained in the CGNDB, predominantly for hydrographic features such as rivers and lakes. The new model can manage multiple formats including PDF decision documents, shapefile delineations, and sound files. It also enables better support for data validation, database monitoring, and statistical reporting.

FIGURE 1: CGNDB 2015 DATA MODEL



CGNDB Feature Identifiers

The key to the transformation of the CGNDB from attribute-based to geospatial-based was the inclusion of Feature Identifiers (FIDs). FIDs were developed to uniquely identify each named feature; they are implemented as a Universal Unique Identifier and are 32 alphanumeric characters in length. FIDs remain associated with a feature regardless of the name changes of the feature. FIDs allow for retrieval of all of the names with the same spatial extent. However, in circumstances where there is a significant change in the spatial extent of the feature, a new FID would be assigned. For example, if a lake were to evaporate and become two separate smaller lakes, two new FIDs would be assigned, one to each of the two new lakes.

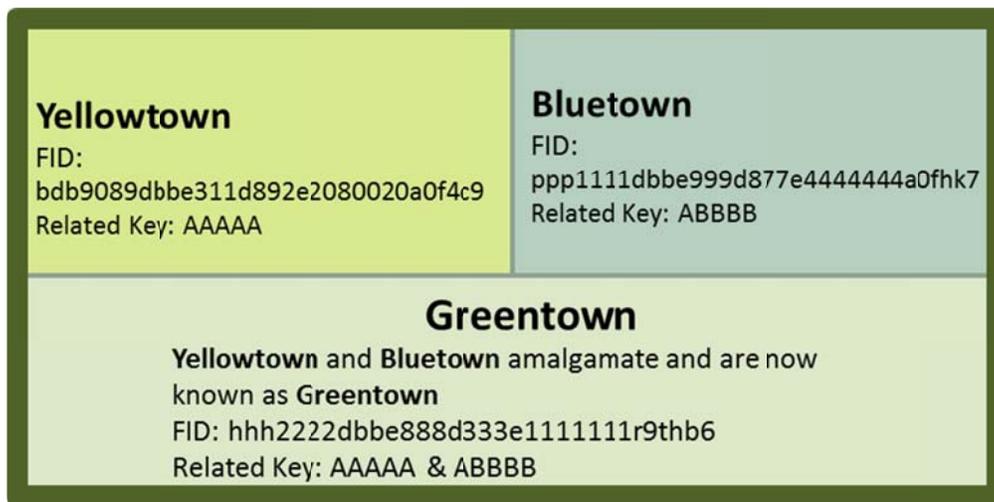
In 2015-2016, NRCAN completed the work to add FIDs to all toponyms in the CGNDB. The process of adding FIDs to the records was semi-automated. Automatic assignment requirements were created to assign FIDs to non-complex relationships. For example, a related currently official and formerly official name that has the same generic definition would have been automatically assigned the same FID.

To handle records that did not meet the automatic assignment requirements, a special Working Group of GNBC members discussed various scenarios, and created a set of FID assignment rules. A thorough examination of the historical relationships and spatial extent of the records was carried out by geospatial technicians at NRCAN in consultation with the provincial and territorial naming authorities to ensure that the correct FID was assigned.

Maintaining Historical Relationships

Prior to the use of FIDs, geographical name relationships were handled using a five character Related Key. The Related Key follows the “history of the name”, meaning it follows the various name changes of the feature over time, but not necessarily the exact spatial extent of the name. For example, when two municipalities merge or amalgamate and the new municipality takes the name of one of the previous, there would be three separate FIDs, and the Related Key would historically link all three of the names together.

FIGURE 2: FID ASSIGNMENT SCENARIO FOR AMALGAMATED TOWNS



Geographical Names Web Application

To support the development of the new data model for the CGNDB, and to facilitate queries, edits and updates by GNBC members, NRCAN has also updated a web-based application tool called GNApp-II. The new application was launched in February 2016, and supports improved interaction by the GNBC members directly with the national database. The new application was developed with the input of the members through extensive requirements gathering and usability testing. The application has improved display and searching functionality, as well as a map visualizer. Through the GNApp-II application, GNBC naming authorities can now submit naming decision documents to the database, as well as upload delineations of named features to the CGNDB in shapefile format.

CGNDB Data Dissemination

Canada’s national geographical names database may be accessed by the public through a search tool on the NRCAN public website, and downloaded in various formats..

Web-based queries through the NRCAN website may be based on the name, feature type, province or territory, latitude/longitude coordinates, rectangular area, or using a unique key.

Users may also search for names containing aboriginal language characters (Inuktitut syllabics or extended Roman alphabet). A query returns the feature type, region, unique identifiers of both the name and the feature, latitude and longitude, the date when the name was approved or changed status, and a spatial delineation of the name's location (if available) overlaid on a base map. Official and historical names of Canadian geographical features, including populated places and undersea features and can accessed via the NRCan website at: <http://www.nrcan.gc.ca/earth-sciences/geography/place-names/search/9170>

The suite of geographical names digital products delivered from the CGNDB was recently re-evaluated, reformatted, updated, and standardized to meet current data needs. Geographical names product files covering all Canadian provinces and territories may be downloaded in Shapefile, KML, GML, and CSV format from NRCan's GeoGratis portal (<http://geogratias.gc.ca/>) as well as from the Government of Canada's Open Data Portal (<http://open.canada.ca/en/open-data>).

In addition, NRCan offers an Application Programming Interface (API) as a means for direct access to the CGNDB for customized searches.

Looking Forward

In 2016-2017, NRCan will develop additional functionality in the GNApp-II application, such as enabling GNBC members to add sound files containing the pronunciation of names, as well as adding batch upload/download capabilities through the web application interface.

NRCan will work in collaboration with the naming authorities of the GNBC to undertake an in-depth review of the various code lists within the CGNDB that define the feature type and status of the toponym. The goal of this review is to determine if further efficiencies and advancements can be made through strengthened validation rules for data entry in the GNApp-II Web Application.

NRCan will continue to work with GNBC naming authorities to define optimum dissemination formats for geographical names data, and how to handle some of the specific challenges of names dissemination, for example including multiple official names for features, and names in multiple Indigenous languages. The goal is to provide consistent, authoritative and updated geographical names from the CGNDB as part of the suite of open geospatial data for Canada.