Encoding

Text Encoding:

At the 18th session of the United Nations Group of Experts on Geographical Names in Geneva in 1996 a Working Group on Toponymic Data Exchange Formats and Standards was formed to investigate and recommend requirements, standards and formats which were available for the encoding, processing, international exchange and promotion of nationally standardized geographical names for international use.

The Working Group completed a survey of requirements for the text encoding of nationally approved and standardized geographical names and a review of the suitability of existing standards. The International Organization for Standardization (ISO) in conjunction with the Unicode Consortium had developed the 16-bit International Standard ISO/IEC 10646 as a solution for the problem of encoding text in a multilingual information processing environment.

The Working Group concluded that most romanized geographical names could be encoded using this standard.

For more information on text encoding and Unicode, click here

Statistics Canada's 2011 Census identified more than 60 Indigenous languages in Canada, and the number of officially recognized geographical names in these languages is constantly growing. The Secretariat of the Geographical Names Board of Canada (GNBC) manages and maintains the Canadian Geographical Names Database (CGNDB), and has developed solutions for displaying these geographical names. While many languages are properly represented using the Latin alphabet (used for English and French), other Indigenous languages require the use of diacritics or syllabics in order to properly spell and represent the geographical names used by Indigenous communities. The attached paper (see below) describes efforts by Natural Resources Canada (NRCan) to represent Indigenous geographical names in Canada. Efforts were expanded tremendously when the CGNDB was converted to the UTF-8 encoding (or Universal Coded Character Set Transformation Format – 8-bit).

How UTF-8 revolutionized the Writing of Indigenous Geographical Names (in English)

UTF-8 a révolutionné l'écriture des langues autochtones (in French)