UNITED NATIONS GROUP OF EXPERTS	WORKING PAPER
ON GEOGRAPHICAL NAMES	NO. 21/9/Rev.1

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

UNGEGN World Geographical Names Database: an update

Submitted by Canada*

_____

* Prepared by Helen Kerfoot, Former Chair, UNGEGN Toponymist, Canada

UNGEGN World Geographical Names Database: an update
(http://unstats.un.org/unsd/geoinfo/geonames/)

Helen Kerfoot,
Former Chair, UNGEGN
Toponymist, Canada

Summary

This Working Paper provides updates on the UNGEGN World Geographical Names Database initiated in 2004, and urges UN member states to provide data (including name, romanized form, and coordinates in an Excel file, as well as audio pronunciation) on their major cities/towns to be included in the database. A more complete account can be found in WP59 presented at the 28th UNGEGN session in 2014.

Introduction

At the 28th Session of UNGEGN, Working Paper 59 presented details about the UNGEGN World Geographical Names Database, its origin and development since its initiation in 2004, to assist in responding to questions received by the UNGEGN Secretariat. This document, therefore, provides only an overview and an update. Technical details about the software, web access and boundaries can be found in this earlier document.

As recommended by the 20th Session of UNGEGN (2004) and later supported by resolution IX/6 of the Ninth Conference on the Standardization of Geographical Names (2007), the UNGEGN Secretariat has taken the lead in developing a world database to collect, manage and disseminate authoritative data on country, capital and major city names. Available on the UNGEGN website, the database helps to provide responses to toponymic questions received by the Secretariat and provides the opportunity for countries to make available the standardized forms of their city names in a worldwide context.

With the input of UNGEGN and the individual UN member states, we have a multilingual, multi-scriptual geo-referenced database, representing the reality of geographical names in a variety of languages and scripts. It is available to the general public through a web interface where names are linked to a map and their spelling and pronunciation (as audio files) can be displayed in tables, together with coordinates and romanized forms, as necessary.

Data

Data currently being collected for the database and provided for the public continues to be as follows:

    (1) Country names - formal and short forms
        a. In the language(s) and writing system(s) of the UN member state itself (source: UNGEGN Working Group on Country Names)
        b. As used by the UN in Arabic, Chinese, English, French, Russian, and Spanish (source: UN Termium database)
    (2) Capital cities
        a. In the language(s) and writing system(s) of the UN member state – with audio pronunciation (source: UN member state)

      b. As used by the UN in Arabic, Chinese, English, French, Russian, and Spanish (source: UN Termium database)

(3) Cities/towns with a population over 100,000

      a. Names (endonyms) as supplied by each UN member state in its own language(s) and writing system(s) – with audio pronunciation (source: UN member state)

      b. Romanized forms of the city/town names (where possible through systems recommended through UN resolutions) (source: UN member state)

For each capital or city name stored the following data is indicated in table form:

- coordinates of latitude and longitude (in degrees and decimal degrees),
- the language of the country in which the name is used,
- the data source (UNGEGN or UN Termium),
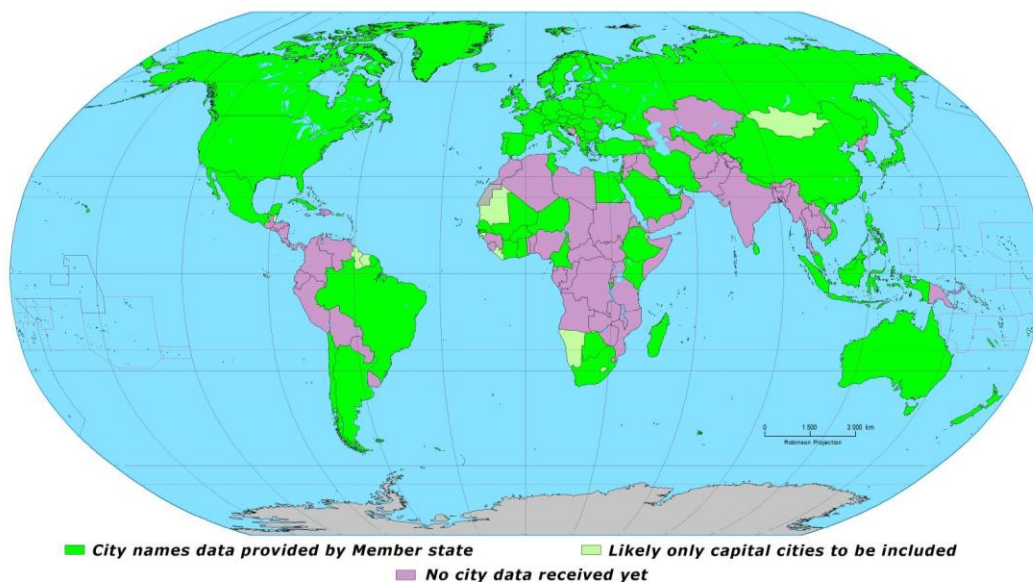- audio files for pronunciation as supplied by the UN member state.

To date the following countries have supplied city data:

A. City/town data sets - some with updates (see Figure 1):

Argentina, Australia, Austria, Belarus, Belgium, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cameroon, Canada, Chile, China, Côte d'Ivoire, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Ethiopia, Finland, France, Gambia, Germany, Ghana, Greece, Hungary, Iceland, Indonesia, Iran (Islamic Republic of), Ireland, Israel, Italy, Japan, Kenya, Kyrgyzstan, Latvia, Lithuania, Madagascar, Malaysia, Mali, Mexico, Nepal, Netherlands, New Zealand, Niger, Norway, Philippines, Poland, Portugal, Republic of Korea, Romania, Russian Federation, Saudi Arabia, Senegal, Serbia, Slovakia, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Tajikistan, The former Yugoslav Republic of Macedonia, Tunisia, Turkey, Ukraine, United Arab Emirates, United Kingdom, United States of America, Uzbekistan, Viet Nam (76)

Figure 1.

World Geographical Names Database – data provided for cities over 100,000

City names data provided by Member state   Likely only capital cities to be included
No city data received yet

In addition, a number of countries that have not submitted data will have only a capital city represented, as no other cities (and maybe not even the capital) have a population over 100,000, for example:
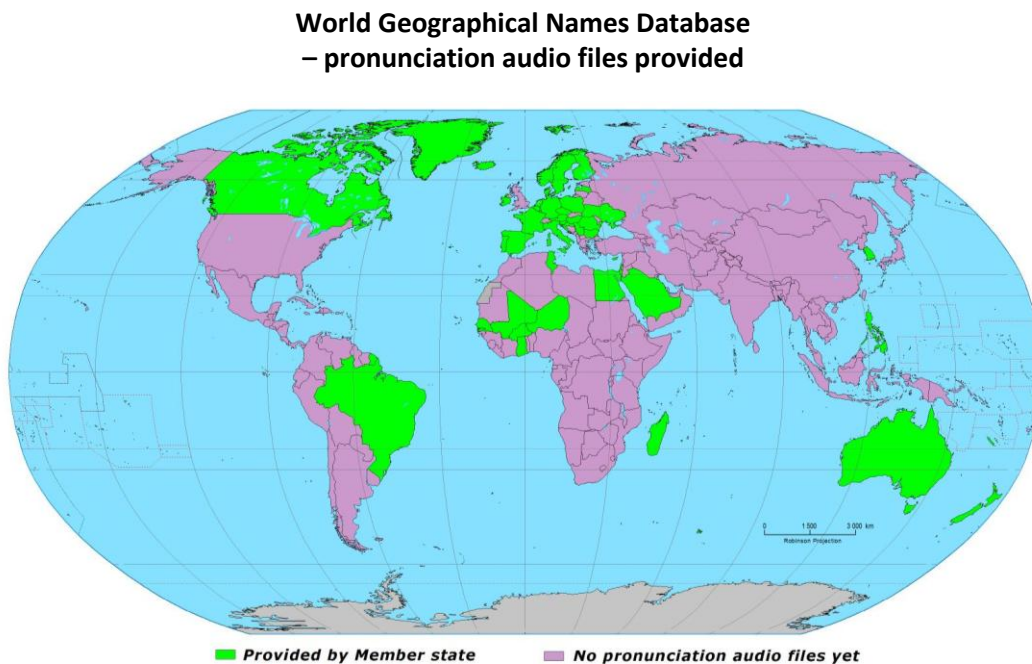
Andorra, Antigua and Barbuda, Bahamas, Barbados, Belize, Bhutan, Cape Verde, Comoros, Djibouti, Dominica, Fiji, Grenada, Guinea-Bissau, Guyana, Kiribati, Lesotho, Liberia, Liechtenstein, Luxembourg, Maldives, Malta, Marshall Islands, Mauritania, Micronesia – Federated States of, Monaco, Mongolia, Montenegro, Namibia, Nauru, Palau, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, San Marino, Sao Tome and Principe, Seychelles, Singapore, Solomon Islands, Suriname, Tonga, Trinidad and Tobago, Tuvalu, Vanuatu (44).

Audio files for pronunciation are, however, welcomed from these as well as other countries.

  B. Audio files for pronunciation of capitals and city/town names (see Figure 2):

Australia, Austria, Belgium, Brazil, Bulgaria, Burkina Faso, Canada, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Finland, France, Gambia, Germany, Ghana, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Madagascar, Mali, Netherlands, New Zealand, Niger, Norway, Philippines, Poland, Portugal, Republic of Korea, Romania, Saudi Arabia, Senegal, Serbia, Slovenia, Spain, Sweden, Tunisia, Ukraine (42)

Figure 2.

**World Geographical Names Database
– pronunciation audio files provided**

The current statistics of available data for UN member states (as of 13 April 2016) are as follows:

❖ 5987 name records[1]
   ▪ 193 countries
      o 1304 country names (including names in the six official UN languages from UN Termium)
         ▪ 273 endonyms
         ▪ 1031 exonyms

---

[1] If both short and formal country names and romanized forms of country and city names are counted as separate items, the total count of names would be over 8800.

- 3430 cities
  - 4683 city names
    - 3477 endonyms
    - 1206 variants
- ❖ Romanization: 50 systems ... 1587 romanized forms of names
- ❖ Sound files: 1100 cities from 42 countries
- ❖ Languages: 116 (with Chinese- and English-language forms of names having the highest counts)

Current requests for UN member state data

At this time we are still requesting the following information from each UN member state:

(1) If your city/town information is not yet loaded, we require an Excel file of the names of cities/towns with a population over 100,000 (in the language(s) and script(s) of the country), together with the latitude and longitude (in degrees and decimal degrees) and romanized forms of the names (preferably according to a UN-recognized Romanization system).

If your city/town information is already loaded, updates and corrections are always welcomed.
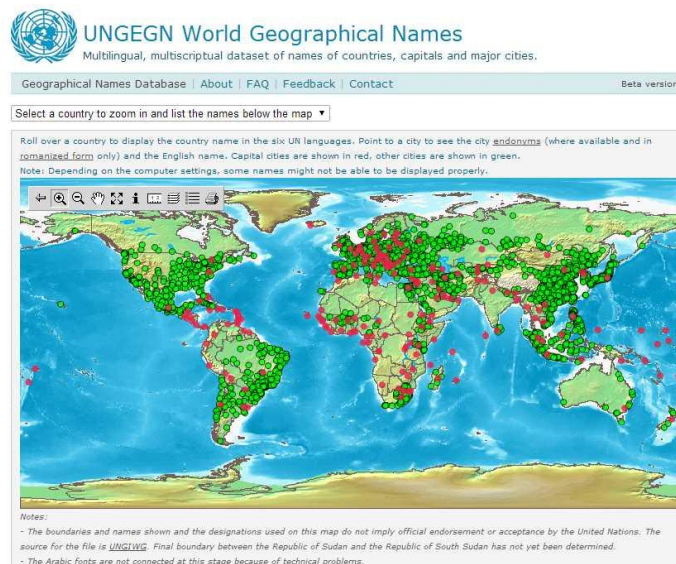
(2) Audio pronunciation files (.wav or MP3) are very useful for database users and we hope to expand this part of the database with contributions from more countries.

Changes to the names of UN member states as used by the UN will continue to be monitored in conjunction with UN Termium and the UNGEGN Working Group on Country Names.

Uploads to the database are made on a continuing basis. However, uploads to the website are made quarterly (see Figure 3 for world overview).

Figure 3.

UNGEGN World Geographical Names
- web page opening with world map access

Further references and acknowledgements

Please refer to WP 59 of the 28th Session of UNGEGN (2014) on the UNGEGN website for details of querying the online database and for references to previous documents and presentations.

Thanks are expressed to the Convenor and members of the Working Group on Romanization Systems, and to Paul Pacheco in the UN Statistics Division for updating and maintaining the UNGEGN Geographical Names Database and web access.

Annex 1

Examples of Tables

(1) Country names: Sri Lanka

Showing: (a) endonyms in Sinhalese and Tamil, with romanized forms, both as the country's short name and formal name;
(b) UN language forms for the short name and formal name.

Country names

Endonyms

Language	Short name	Formal name	
Sinhalese	ශ්‍රී ලංකා *Shrī Laṁkā*	ශ්‍රී ලංකා ප්‍රජාතාන්ත්‍රික සමාජවාදී ජනරජය *Shrī Laṁkā Prajātāntrika Samājavā dī Janarajaya*	❶
Tamil	இலங்கை *Ilaṅkai*	இலங்கை சனநாயக சோஷலிசக் *Ilaṅkai Jananāyaka Choshalichak*	❶

United Nations languages

Language	Short name	Formal name	
Arabic	سري لانكا	جمهورية سري لانكا الاشتراكية الديمقراطية	❶
Chinese	斯里兰卡	斯里兰卡民主社会主义共和国	❶
English	Sri Lanka	the Democratic Socialist Republic of Sri Lanka	❶
French	Sri Lanka [fém.]	la République socialiste démocratique de Sri Lanka	❶
Russian	Шри-Ланка	Демократическая Социалистическая Республика Шри-Ланка	❶
Spanish	Sri Lanka	la República Socialista Democrática de Sri Lanka	❶

(2) Capital city and other major cities – Finland

Showing: (a) Capital city endonyms in Finnish and Swedish, with audio pronunciation; capital city as used by the UN in the UN languages

(b) Other major cities – endonyms in Finnish and Swedish, with audio pronunciation

Capital city

Lat: 60.20 | Long: 24.92   **Helsinki**

Endonym

Language	City name	Source
Finnish	Helsinki ◀»	UNGEGN
Swedish	Helsingfors ◀»	UNGEGN

United Nations Languages

Language	City name	Source
Arabic	هلسنكي	UN terms
Chinese	赫尔辛基	UN terms
English	Helsinki	UN terms
French	Helsinki	UN terms
Russian	Хельсинки	UN terms
Spanish	Helsinki	UN terms

Other major cities

Lat: 60.21 | Long: 24.65   Espoo

Endonym

Language	City name	Source
Finnish	Espoo ◀»	UNGEGN
Swedish	Esbo ◀»	UNGEGN

Lat: 60.30	Long: 25.02	Vantaa
Lat: 60.45	Long: 22.27	Turku
Lat: 60.98	Long: 25.66	Lahti
Lat: 61.50	Long: 23.73	Tampere
Lat: 62.24	Long: 25.75	Jyväskylä
Lat: 65.01	Long: 25.47	Oulu