UNITED NATIONS   **WORKING PAPER**
GROUP OF EXPERTS   **NO. 56/9**
ON GEOGRAPHICAL NAMES

Twenty-eight session
28 April – 2 May 2014

Item 9 of the Provisional Agenda

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

UNGEGN World Geographical Names Database *

UNGEGN World Geographical Names Database

(http://unstats.un.org/unsd/geoinfo/geonames/)

Helen Kerfoot,
Former Chair, UNGEGN
Emeritus Scientist, Natural Resources Canada

Summary

The UNGEGN World Geographical Names Database was initiated in 2004 and has continued with the support of resolution IX/6 of 2007. It is a multilingual, multi-scriptual geo-referenced database containing names of UN member states, capitals, and cities/towns with a population over 100,000. All entries provide endonyms, as well as forms used by the United Nations in Arabic, Chinese, English, French, Russian and Spanish for the countries and capitals. The data, now uploaded quarterly, is accessible on the UNGEGN website at http://unstats.un.org/unsd/geoinfo/geonames/ through a world map interface and tables.

UNGEGN experts are urged to supply (or update) the city/town data from their countries together with the recognized coordinates of latitude and longitude. In addition, experts are encouraged to supply audio files for the pronunciation of each city name; these will be attached to the individual entries and be available to web users.

The UNGEGN Secretariat continues to take responsibility for maintaining the database and development of the web interface, and looks for your ongoing support with the supply of data.

_____

Introduction

At its twenty-second session in 2004, UNGEGN recommended that the Secretariat take the lead in developing a world database to collect, manage and disseminate authoritative data on country and major city names. In particular this would use the UNGEGN website to make available information that would help respond to toponymic questions received by the Secretariat and would provide a vehicle for countries to have their city names displayed in standardized form within a worldwide framework.

As a result, the Secretariat with advice from UNGEGN initiated the process of building a multilingual, multi-scriptual geo-referenced database, designed to represent the reality of geographical names in a variety of languages and scripts. The database had to be available to Experts and the general public through a web interface. Names for places would be linked to a map, and standardized names, their spelling and pronunciation, would be displayed as tables.

At the time, the database was created in SQL Server 2005 which could store all the information necessary for populating the map and providing data in tabular format (including city and country names, ISO 3-letter country and language codes, variants, coordinates, comments and pronunciation audio files).

Following a special presentation to the Ninth UN Conference on the Standardization of Geographical Names in 2007, the Conference passed resolution IX/6, recommending that the UN Statistics Division, in cooperation with the UN Cartographic Section, the Second Administrative Level Boundaries (SALB), UNGEGN and member States "further develop, populate and maintain the geographical names database" of UNGEGN, "initially containing names of countries, capitals and major cities".

Data

Data available so far includes:

(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
      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)
      b. Romanized forms of the city/town names (where possible through systems recommended through UN resolutions)
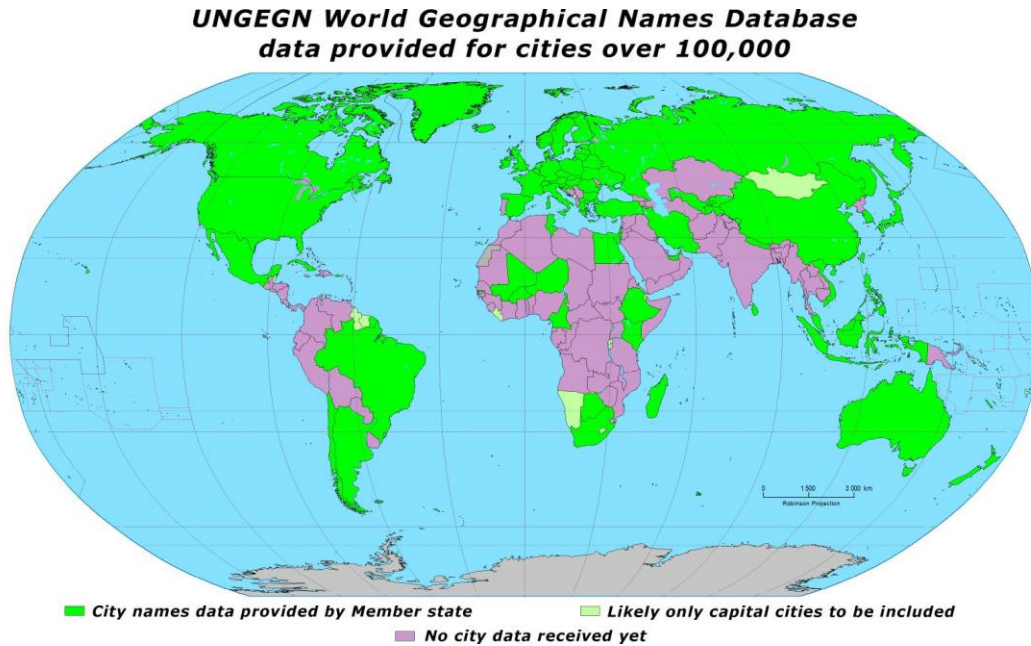
For each country or city name stored, coordinates of latitude and longitude are indicated, the language in which the name is used is supplied, the data source is noted, and audio files for pronunciation can be included if they are supplied by the UN member state.

Convenors of the UNGEGN Working Group on Romanization Systems and the Working Group on Country Names have been very helpful in verifying data before it is loaded into the database. In addition, the Working Group on Toponymic Data Files and Gazetteers has provided advice during the development stages.

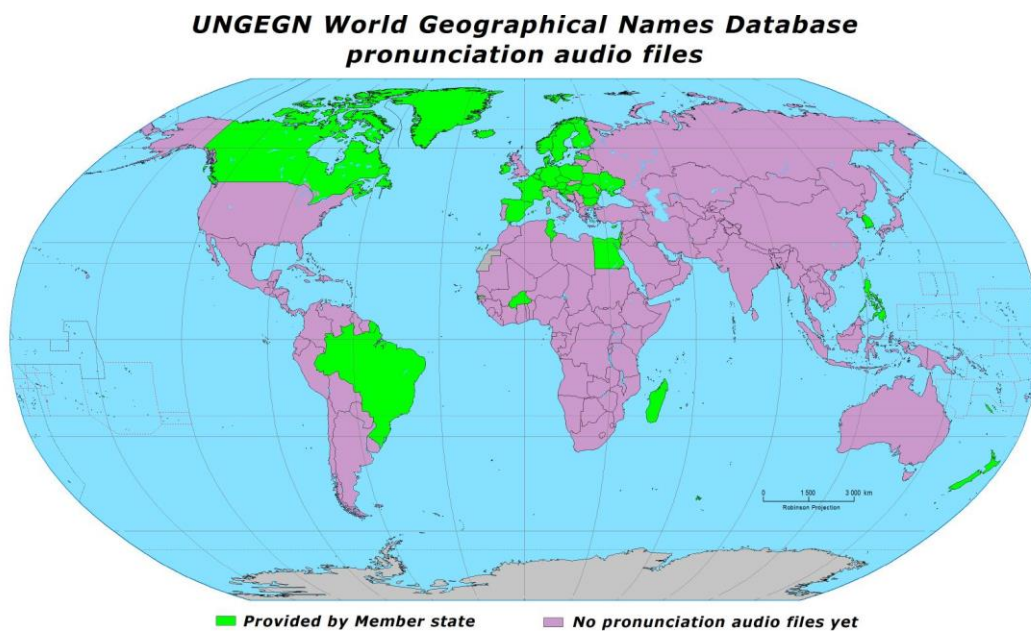To date the following countries have supplied city data:

    A. City/town data sets - some with updates (see Figure 1):
Argentina, Australia, Austria, Belarus, Belgium, Botswana, Brazil, Bulgaria, Burkina Faso, Cameroon, Canada, Chile, China, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Ethiopia, Finland, France, Gambia, Germany, 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, Republic of Korea, Romania, Russian Federation, 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

Figure 1.

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

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

B. Audio files for pronunciation of city/town names (see Figure 2):
Austria, Belgium, Brazil, Bulgaria, Burkina Faso, Canada, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Finland, France, Gambia, Germany, Hungary, Iceland, Ireland, Israel, Latvia, Madagascar, Netherlands, New Zealand, Norway, Philippines, Poland, Republic of Korea, Romania, Slovenia, Spain, Sweden, Tunisia, Ukraine

Figure 2.

UNGEGN World Geographical Names Database
pronunciation audio files

Legend:
- Provided by Member state
- No pronunciation audio files yet

The current statistics of available data for UN member states are as follows:

- ❖ 5915 name records[1]
  - 193 countries
    - o 1304 country names (including names in the six official UN languages from UN Termium)
      - 273 endonyms
      - 1031 exonyms
  - 3362 cities
    - o 4611 city names
      - 3406 endonyms
      - 1205 variants
- ❖ Romanization: 47 systems ... 1556 romanized foms of names
- ❖ Sound files: 974 cities from 33 countries
- ❖ Languages: 116 (with Chinese- and English-language forms of names having the highest counts)

Boundaries (from Sibylle Marxgut presentation, 2007)

The shapefile used is sourced from the UN Geographical Information Working Group (UNGIWG) 1:1 million dataset reflecting cartographic practices of the United Nations.

Some technical aspects of the data storage and web interface (including information from Sibylle Marxgut presentations, 2007 and 2009)

A Web Map Server (WMS) was set up to show the names on a world map. The WMS protocol is based on simple query syntax for posting a request for the desired layers and zoom window to the server, which returns a map as a standard picture. It has been developed in compliance with standards and protocols of the Open GIS Consortium.

The web interface was developed in ASP Net 2.0 and Macromedia Flash 8 to provide a map application and table format.

So far the user interface has been developed in English, although the data is multilingual.

Web access

With initial log-in, the user sees the world map of UN member states, displaying the location of capitals (red), and other cities (green) that have been uploaded with data received from the UN member states (see Figure 3). A tool bar provides map zooming and dragging capabilities, selection/de-selection of map layers, a distance measuring tool and printing icon.

Also from this view, a small amount of background information can be accessed above the map: a statement about the database, FAQs, Feedback and Contacts.

---

[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 8838.

Figure 3. World map displayed at log-in

Queries can be made via the map interface or by using the pull down menu of country names.

(1) Names of Countries

Figure 4. Rolling over a country will show the short form of the name of the country in the six languages of the United Nations
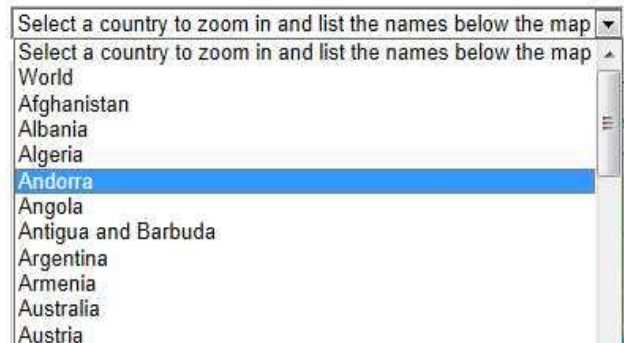

Figure 5. Individual countries can be selected from the pull-down menu

When a country is selected from the pull-down menu a map showing the country will be displayed (example in Figure 6) and below will appear tabular data (example in Figure 7) showing information on the short and formal names for the country:

- in the language(s) of the country, both in the writing system of the language and in the Roman alphabet, and
- in the United Nations languages (Arabic, Chinese, English, French, Russian and Spanish).

Figure 6. Country map enlarged

Figure 7. Tabular data on country names

Country names

Endonyms

Language	Short name	Formal name	
Greek	Ελλάς *Ellás*	Ελληνική Δημοκρατία *Ellinikí Dimokratía*	❶

United Nations languages

Language	Short name	Formal name	
Arabic	اليونان	الجمهورية الهيلينية	❶
Chinese	希腊	希腊共和国	❶
English	Greece	the Hellenic Republic	❶
French	Grèce (la)	la République hellénique	❶
Russian	Греция	Греческая Республика	❶
Spanish	Grecia	la República Helénica	❶

(2) Names of Capital cities

Figure 8
For names of capitals, rolling over the red circles
 displays the endonym(s) (romanized) and the
name as used in English

Tabular data for capital cities (example in Figure 9) shows:
* the capital city endonym(s), with romanized form where applicable
* the name as used in the six UN languages
* the latitude and longitude in degrees and decimal degrees
* the data source
* audio pronunciation files, if provided by the UN member state; they are attached to the name and can be heard by clicking on the audio icon.

Figure 9. Tabular data for capital cities

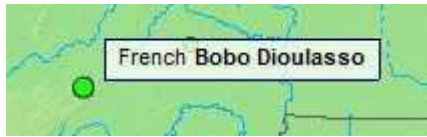### (3) Names of major cities (population over 100,000)

Figure 10. Cities with a population over 100,000 (other than capital cities) are shown with green location symbols on the map

Cities are displayed as geo-referenced endonyms in the tables. Audio files are included, if provided by the UN member state, and are accessed through the audio symbol against the city name in the tables.

Figure 11. Major cities with associated audio pronunciation files

Other major cities

Lat: 10.63 | Long: -4.77   Banfora

Endonym

Language	City name	Source
French	Banfora 🔊	UNGEGN

Lat: 11.20	Long: -4.30	Bobo Dioulasso
Lat: 11.50	Long: -3.52	Houndé
Lat: 11.78	Long: -0.37	Tenkodogo
Lat: 12.06	Long: 0.36	Fada N'gourma
Lat: 12.25	Long: -2.37	Koudougou
Lat: 12.47	Long: -3.47	Dédougou
Lat: 13.09	Long: -1.08	Kaya
Lat: 13.58	Long: -2.42	Ouahigouya

Further information

More information and progress reports on the UNGEGN World Geographical Names Database are available:

(1) Special presentations:
   (a) 2007: Ninth UN Conference "UNGEGN project on geographical names database"
       http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/special-presentations-docs/geonamesdatabase.pdf
   (b) 2009: 25th UNGEGN Session "The UNGEGN Geographical Names Database"
       http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/25th-gegn-docs/presentations/ungegn%20db%20project_2009-04-24.pdf

(2) Within Secretariat reports to UNGEGN (http://unstats.un.org/unsd/geoinfo/UNGEGN/sessions.html)
   (a) 2004   GEGN/22/9   last page
   (b) 2006   GEGN/23/11   section 3
   (c) 2009   GEGN/25/11   section on Database
   (d) 2011   GEGN/26/11   section on Database

(3) Within UNGEGN Bulletins (http://unstats.un.org/unsd/geoinfo/UNGEGN/bulletin.html)
    (a) #35   August 2008   p. 7-8
    (b) #36   March 2008   p. 6
    (c) #37   Sept. 2009   p. 6-7
    (d) #38   March 2010   p. 8
    (e) #39   Fall 2010   p. 6

Current work

The UNGEGN Secretariat continues to be responsible for the design and development of the database, as well as for entry of data that has been supplied by UN member states or comes from UN Termium. The website is now updated four times a year to reflect the information loaded on an ongoing basis into the UNGEGN database. Since 2012, adjustments have been made to improve the representation of a number of map views at the country level (for example, for some island countries).

The UNGEGN Secretariat in conjunction with UNGEGN experts will continue to focus on fundamental aspects of the database.

- City data (in an Excel table with coordinates) is still required from a number of UN member states. As well, updates are welcomed.

- Audio files (preferably .wav files) when supplied by member states will be attached to the city, capital, and country name records to enhance users' understanding.

- Changes in names of UN member states as used by the United Nations will be monitored for updating.

- Updating of boundaries as new UN member states come into existence will be investigated.

- Expanding FAQs, upgrading the web interface and character representation, will be undertaken as appropriate.

- Uploads of data to the UNGEGN website will be completed quarterly.

Thanks are expressed to Sibylle Marxgut for several years work on the UNGEGN database, and to Paul Pacheco who has taken over the data updates and web interface improvements.