

## Section 4 Regional and global initiatives

### Chapter 9 UNGEGN World Geographical Names Database

Helen Kerfoot

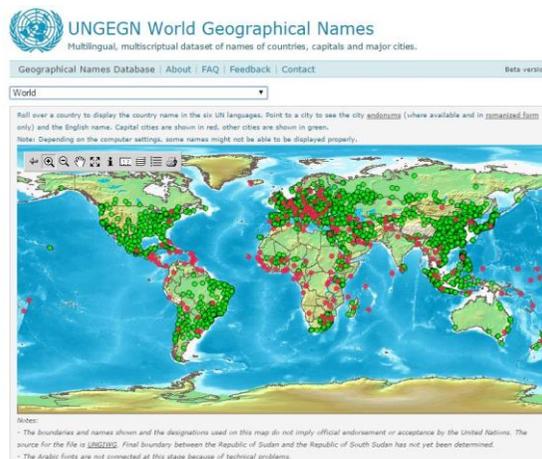
#### 9.1 Introduction

The UNGEGN World Geographical Names Database was initiated in 2004 and has continued with the support of resolution IX/6 of the Ninth UN Conference on the Standardization of Geographical Names in 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 (UN Member States) 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 (Figure 9-1). The UNGEGN Secretariat is responsible for maintaining the database and development of the web interface.

#### 9.2 History behind the development of the database

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



*Figure 9-1 World map; starting point for searching the UNGEGN World Geographical Names Database UNGEGN experts are responsible for supplying (or updating) 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 are attached to the individual entries and are available to web users.*

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 UNGEGN 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 (Figure 9-2), the Conference passed resolution IX/6, recommending that the UN Statistics Division, in cooperation with the UN Cartographic Section, the UN 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”.

UNEGN Project on Geographical Names Database	
An update on the application development	
1.	Introduction..... 2
2.	Data availability..... 2
3.	Data storage..... 2
4.	The boundary file..... 3
5.	Web interface development..... 3
5.1.	The map..... 3
5.2.	The table..... 5
5.2.1.	Countries..... 5
5.2.2.	Cities..... 6
6.	Further steps..... 7
6.1.	Data collection and processing..... 7
6.2.	Internal database maintenance..... 8
6.3.	Application development..... 8
Annex I - Example for submitting country names..... 9	
Annex II - Example for submitting city names..... 10	

*Figure 9-2 Special presentation available in text and slides at <http://unstats.un.org/unsd/geoinfo/UNEGN/unegnConf9Add.html>*

### 9.3 Geographical names data included

So far, the data 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 Term 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 Term 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 provided by the UN member state.

Convenors of the UNGEGN Working Group on Romanization Systems and the Working Group on Country Names help 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 (June 2015) the following countries have supplied city data:

A. City/town data sets - some with updates (see Figure 9-3):

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, Saudi Arabia, 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

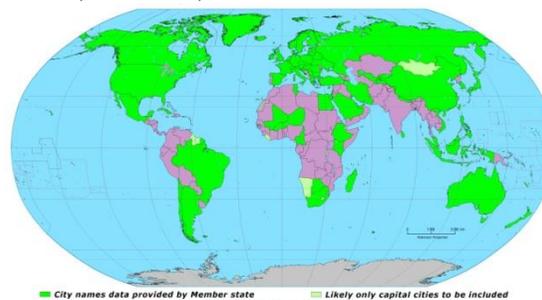


Figure 9-3 UNGEGN World Geographical Names Database – data provided for population centres over 100,000

B. Audio files for pronunciation of city/town names (see Figure 9-4):

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, Saudi Arabia, Serbia, Slovenia, Spain, Sweden, Tunisia, Ukraine

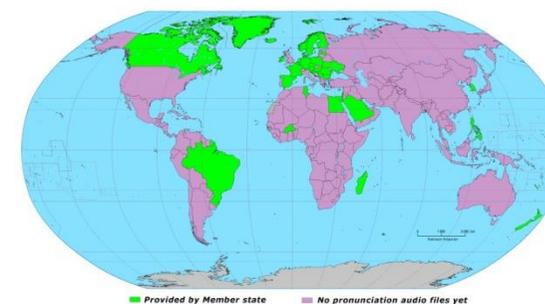


Figure 9-4 UNGEGN World Geographical Names Database – pronunciation audio files provided

The current statistics of available data for UN member states (as of September 30, 2015) are as follows:

- ❖ 5947 name records<sup>1</sup>
  - 193 countries
    - 1304 country names (including names in the six official UN languages from UN Termium)
      - 273 endonyms
      - 1031 exonyms
  - 3393 cities

<sup>1</sup> 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 8884.

- 4643 city names
  - 3437 endonyms
  - 1206 variants

- ❖ Romanization: 50 systems ... 1587 romanized forms of names
- ❖ Sound files: 1007 cities from 35 countries
- ❖ Languages: 116 (with Chinese- and English-language forms of names having the highest counts)

### 9.4 Boundaries

According to Sibylle Marxgut in her presentation to the Ninth Conference in 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.

### 9.5 Some technical aspects of the data storage and web interface<sup>2</sup>

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.

<sup>2</sup> From information from Sibylle Marxgut's presentations to the Ninth Conference in 2007 and to UNGEGN in 2009.

### 9.6 Web access

With the 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 (as shown in Figure 9-1). 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.

Queries can be made via the map interface or by using the pull-down menu.

(1) **Country names** (Figures 9-5, 9-6, 9-7, 9-8).



Figure 9-5 Rolling over a country will show the short form of the name of the country in the six languages of the United Nations



Figure 9-6 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 9-7) and below will appear tabular data (example in Figure 9-8) 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 9-7 Example of country map: Greece

Country names		
Endonyms		
Language	Short name	Formal name
Greek	Ελλάδα <i>Elláda</i>	Ελληνική Δημοκρατία <i>Ellinikí Dimokratía</i>
United Nations languages		
Language	Short name	Formal name
Arabic	اليونان	الجمهورية اليونانية
Chinese	希腊	希腊共和国
English	Greece	the Hellenic Republic
French	Grèce (la)	la République hellénique
Russian	Греция	Греческая Республика
Spanish	Grecia	la República Helénica

Figure 9-8 Tabular data providing short and formal country names: Greece

## (2) Names of capital cities

For names of capitals, rolling over the red circles displays the endonym(s) (romanized) and the name as used in English (see Figure 9-9)



Figure 9-9 Colombo, Sri Lanka, shown in Sinhalese, Tamil and English

Tabular data for capital cities (example in Figure 9-10) shows:

- The capital city endonym(s), with romanized form where applicable
- The name as used in the United Nations languages (Arabic, Chinese, English, French, Russian and Spanish)
- 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.

Capital and other major cities		
Capital city		
Lat: 6.93   Long: 79.84 කොළඹ - Kōlamba		
Endonym		
Language	City name	Source
Sinhalese	කොළඹ <i>Kōlamba</i>	UNGEGN
Tamil	கொழும்பு <i>Kōlumpu</i>	UNGEGN
United Nations Languages		
Language	City name	Source
Arabic	كولومبو	UN terms
Chinese	科伦坡	UN terms
English	Colombo	UN terms
French	Colombo	UN terms
Russian	Коломбо	UN terms
Spanish	Colombo	UN terms
Variant name		
Language	City name	Source
English	Colombo	UNGEGN

Figure 9-10 Tabular data for the capital city of Sri Lanka

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

Cities with a population over 100,000 (other than capital cities) are shown with green location symbols on the map and their names can be highlighted in the romanized form(s) of the endonym(s). (See Figure 9-11.)



Figure 9-11 A city in Cyprus with its name shown as endonyms: in Greek (romanized) and in Turkish.

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 9-12).

Other major cities		
Lat: 34.68   Long: 33.03   Λεμεσός - Lemesós		
Endonym		
Language	City name	Source
Greek	Λεμεσός 	UNGEGN
Turkish	Leymosun 	UNGEGN

Figure 9-12 Tabular data for a city in Cyprus, including the icons to click for audio pronunciation files

## 9.7 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 through UNGEGN experts, 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, modifications have been made to improve the representation of a number of map views at the country level (for example, for some island countries).

Although we would like to expand the content coverage of the UNGEGN World Geographical Names Database (for instance to include more variant names or to include major physical feature names), for the time being the

UNGEGN Secretariat and UNGEGN experts will continue to focus on fundamental aspects of the database.

The main requirements at present are:

- City data (in an Excel table with coordinates) from a number of UN member states. As well, updates are welcomed.
- Audio files (preferably .wav files) from member states to be attached to the city, capital and country name records to enhance users' understanding of the endonyms.
- Changes in names of UN member states as used by the United Nations and stored in the UN Termium database.
- Updating of boundaries as new UN member states come into existence.
- Expanding FAQs, upgrading the web interface and character representation, to be undertaken as appropriate.

## 9.8 References and further information

This article is based on Working Paper 56 presented at the 28<sup>th</sup> UNGEGN Session in 2014.

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: 25<sup>th</sup> 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](http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/25th-gegn-docs/presentations/ungegn%20db%20project%2009-04-24.pdf)

- (2) Included in 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) Included in 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