

UNGEEN 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

1. Introduction

The United Nations Group of Experts on Geographical Names (UNGEGN) recommended in its 22nd Session, held in New York in April 2004, the development of a data storage system to collect, manage and disseminate names of countries and major cities with a population of more than 100.000. The system must follow the objectives of UNGEGN as specified in the resolutions of the conferences. The database can also contribute to the objective of limiting the use of exonyms in international cartography by linking them to the appropriate endonyms. Furthermore, the pronunciation of endonyms should encourage their use.

As a result, the UNGEGN secretariat initiated the process of building a database to store country names and major city names of the world in a multilingual, multi-scriptual and geo-referenced database. The format should represent the reality of existing geographical names in various languages.

The standardized geographical names should be made available for general reference through one web interface. Data should be made available for the UNGEGN experts and the public at large. Names for places in the world will be linked to a map, so that information on names, spelling and pronunciation can be accessed including links to standardized forms.

2. Data availability

Country names are available from the United Nations (UN) as well as through UNGEGN itself. The UN dataset contains the formal and short country names of the member states in the six official UN languages and an updated UNGEGN-list provides additional information, particularly country names in the official national languages of the countries.

Regarding the city names, the United Nations Cartographic Section provided the United Nations Statistics Division (UNSD) with a dataset containing all capital cities as used in English and in the national language (Romanized where applicable), including the longitude and latitude for positioning the cities on the map. In addition, UNGEGN has so far provided sample data, including sound files (Norway) and for some countries¹, corrections/updates to the city names as found in the UNSD data². In addition, some sample data was collected in various scripts for 2006 demonstrations.

3. Data storage

Based on the requirements for the application and interface design, a database was created in SQL Server 2005, which provides a comprehensive business intelligence platform for data integration, analysis and reporting.

The database stores all information necessary for populating both the map and the tabular presentation: country and city names in different languages including the ISO country and

¹ Australia, Austria, Canada, Croatia, Czech Republic, Estonia, Germany, Greece, Hungary, Iceland, Israel, Kyrgyzstan, Latvia, Malaysia, Netherlands, New Zealand, Norway, Slovakia, Slovenia, Tajikistan, The former Yugoslav Republic of Macedonia, Ukraine, United Kingdom, United States of America, Uzbekistan, Vietnam.

² <http://unstats.un.org/unsd/demographic/sconcerns/densurb/urban.aspx>

language 3-letter codes, endonyms and exonyms (for country names) and variants (in other languages for city names), the data source and the coordinates for the cities. Furthermore, comments or footnotes and pronunciation files can be added to each name record (endonyms only).

4. The boundary file

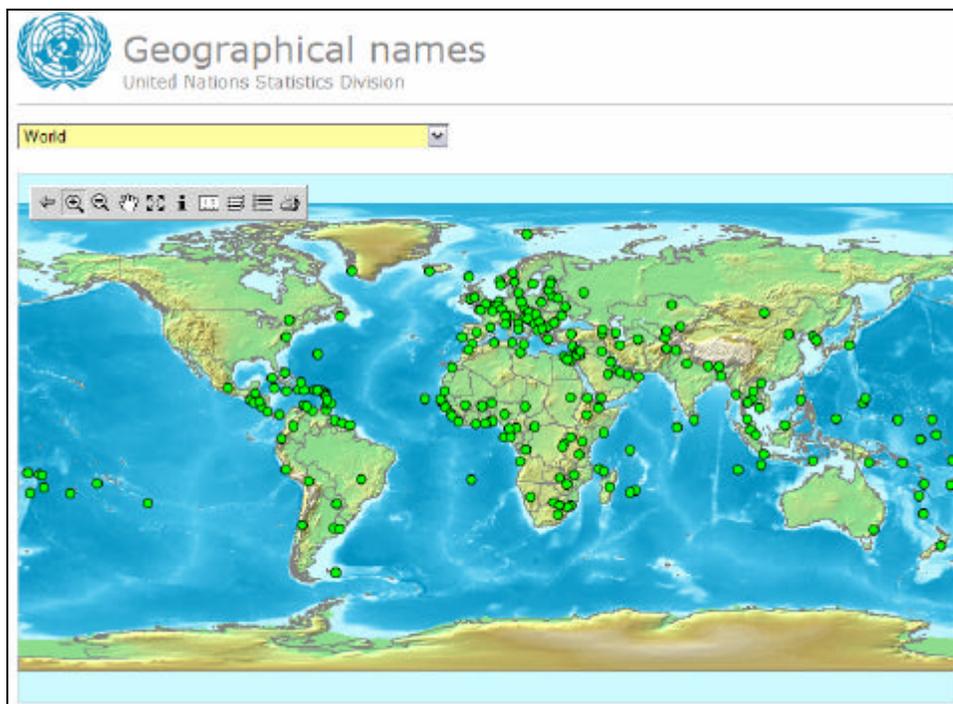
The shapefile for the boundaries is based on the United Nations Geographical Information Working Group (<http://boundaries.ungiwg.org>). The 1:1 million dataset has been chosen as the basis for the mapping application, which reflects the cartographic practice of the United Nations Cartographic Section. This dataset is intended to provide the United Nations community with a worldwide coverage of international boundaries consistent with the boundary representations that are used by the United Nations.

5. Web interface development

The United Nations Statistics Division in cooperation with UNGEGN is developing the web application to offer the various names of countries and cities in different languages on line. The web interface for the geographical information consists of two parts: an interactive map and a tabular section.

5.1. The map

A Web Map Server (WMS) has been 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 defined by the OpenGIS consortium³.



³ <http://www.opengeospatial.org/>

The map shows the countries based on the official UN boundaries and includes the capital cities of all United Nations member states. It offers zooming in and out as well as a query feature, where geographical names can be displayed in an overlaid popup-window.

The world map provides the following features:

Moving the mouse over a country opens up a popup window displaying the names for that country in the UN languages:

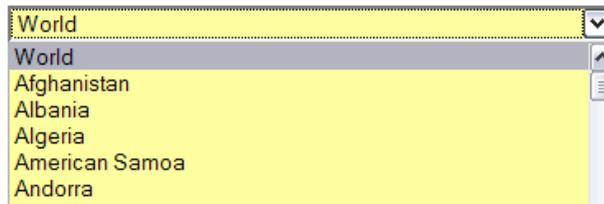
- Arabic
- Chinese
- English
- French
- Russian
- Spanish



Hovering over a city displays the name in English and if available the endonym.



A dropdown menu lists the countries that are available in alphabetical order. Selecting one zooms in to that particular area.



Clicking on a country updates the table right below the map with the complete geographical information. The table includes the geographical names in all forms and languages available for that country and major cities:

- Language
- Endonym or exonym
- Short name and/or long name
- Data source

If applicable or available:

- Romanized short and/or long name including the Romanization system
- Sound file for endonyms
- Footnotes/Comments

Country names

Endonyms

Language	Short name	Formal name
German	Österreich	Republik Österreich

Other languages

Language	Short name	Formal name
Arabic 	أستريا	جمهورية النمسا
Chinese 	奥地利	奥地利共和国
English 	Austria	the Republic of Austria
French 	Autriche (f)	la République d'Autriche
German	Österreich	Republik Österreich
Russian 	Австрия	Австрийская Республика
Spanish 	Austria	la República de Austria

City names

Language	Short name
Capital city	
German	Wien <small>Endonym</small>

5.2. The table

In addition to the map, the user can view the geographical names in tabular format. Whenever a country is selected in the list box or clicked on in the map, the table underneath the map displays the corresponding country information. All available names will be presented in this section of the application.

5.2.1. Countries

The following fields are currently available for the tabular display of country names (the number of records per country only being limited by the number of world languages):

- Short name (s)
- Formal name(s)
- Endonym or exonym
- Language
- Romanization system
- Sound file for endonyms (if available)
- Data source
- Footnotes/Comments

Country names

Endonyms			
Language	Short name	Formal name	
German	Österreich	Republik Österreich	

Exonyms			
Language	Short name	Formal name	
Arabic 	النمسا	جمهورية النمسا	
Chinese 	奥地利	奥地利共和国	
English 	Austria	the Republic of Austria	
French 	Autriche (l')	la République d'Autriche	
German	Österreich	Republik Österreich	
Russian 	Австрия	Австрийская Республика	
Spanish 	Austria	la República de Austria	

Table: Example Austria

5.2.2. Cities

The city names table contains:

- Language
- Name
- Variant
- Romanization system
- Sound file for variants (if available)
- Capital city
- Data source
- Footnotes/Comments

City names			
Language	Short name		
Capital city			
Norwegian	 Oslo	Endonym	
Other major cities			
Norwegian	 Bergen	Endonym	
Norwegian	 Stavanger	Endonym	
Norwegian	 Trondheim	Endonym	

Table: Example Norwegian cities

6. Further steps

This version of the geographical names application is the first step of the development process. Further work has to be done in the following areas:

6.1. Data collection and processing

UNGEKN has to request data on country and city names including the necessary details from the experts.

The data has to be constantly checked and reviewed by the experts of UNGEKN to ensure accuracy and reliability. It is therefore important that the secretariat and UNGEKN form a close partnership with clear responsibilities.

A template for UNGEKN to provide country and city names has been prepared so that the information flow is as easy without format changing. An overview of the template for country names can be found in Annex I, a second one for the city names in Annex II.

For a complete record the following information is necessary:

Country names:

- ISO-3 country code⁴
- Language
- ISO-3 language code⁵
- Endonym or exonym
- Short name and/or formal name
- Data source

If applicable or available:

- Romanized short and/or formal name including the Romanization system
- Sound file for endonyms
- Footnotes/Comments

City names:

- ISO-3 country code
- Name
- ISO-3 language code
- Language
- Endonym or variant in another language
- Data source

If applicable or available:

- Romanized short and/or formal name including the Romanization system
- Sound file for endonyms
- Footnotes/Comments

⁴ http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3 this comment was for you, not to include in the document!!!(a direct ISO link would be preferable on both these footnotes

⁵ http://www.loc.gov/standards/iso639-2/php/code_list.php

6.2. Internal database maintenance

The UNGEGN Secretariat is in charge of the further development of the database. The Secretariat is responsible for data editing for the time being. This includes the entering of new data, the changing and the deleting of data. At a later stage, a mechanism has to be put in place that facilitates the database maintenance.

6.3. Application development

The UNGEGN secretariat will continue to develop the UNGEGN database on geographical names. The web application will be adjusted as new information comes in. Feedback from users and UNGEGN experts is greatly appreciated and will be considered as well to make the website user-friendly and keep the information up-to-date. The comments and suggestions should be sent to zewoldi@un.org.

Once this process of collecting and publishing names of places is complete, the Secretariat could consider adding other major geographic features and their names, for example, rivers, lakes, mountains, and so forth, provided that the necessary resources are available. The ultimate outcome would be a geographic database on country and city names, and major geographic features within a country.

At this stage, the user interface will be developed in English, although the data will be multilingual. Later on the navigation can be offered in other languages as well.

Annex I - Example for submitting country names

Country ISO-3166 3-letter	Short name in non-Roman alphabet	Romanization system used - if appropriate	Romanized form of short name	For endonyms - sound file for the short name	Formal name in non-Roman alphabet	Romanization system used - if appropriate	Romanized form of formal name	For endonyms - sound file for the formal name	Language ISO-639 3-letter	Language	Endonym/Exonym	Data source	Footnotes/Comments
JPN	??	?	Nihon, Nippon	JPN_short.wav	???	?	Nihon-koku, Nippon-koku	JPN_formal.wav	jpn	Japanese	endonym	UNEGN list (Tech manual 2007)	wav file from
ITA			Italia (I')	ITA_short.wav			Repubblica Italiana	ITA_formal.wav	ita	Italian	endonym	UNEGN list (Tech manual 2007)	wav file from
JPN	??????						??????		rus	Russian	exonym	UN Terminology Bulletin	
JPN			Jaapan						est	Estonian	exonym	(1) Maailma Kohanimed, 1999 (2) http://www.eki.ee/knab/mmaad.htm	(1) From Peeter Päll (2) official Estonian website
ITA			Itaalia						est	Estonian	exonym	(1) Maailma Kohanimed, 1999 (2) http://www.eki.ee/knab/mmaad.htm	(1) From Peeter Päll (2) official Estonian website
LVA			Latvija	LVA_short.wav			Latvijas Republika	LVA_formal.wav	lav	Latvian	endonym	UNEGN list (Tech manual 2007)	wav file from
LVA			Lettland				Republik Lettland		deu	German	exonym	Duden 2000	From Joern Sievers
LVA			Läti						est	Estonian	exonym	(1) Maailma Kohanimed, 1999	(1) From Peeter Päll (2) official Estonian website

Annex II - Example for submitting city names

Country ISO-3166 3-letter	Name in non-Roman alphabet	Romanization system used - if appropriate	Romanized form of name	For endonyms - sound file for the name	Language ISO-639 3-letter	Language	For endonym Capital of country (yes/no)	Endonym/Variant	For Variant provide approx. endonym	Latitude (in decimal degrees)	Longitude (in decimal degrees)	Data source	Footnotes/ comments
BEL			Bruxelles	Brux_short.wav	fre	French	yes	endonym		50.83	4.35	Kathleen Van Doren 2006	Bruxelles=B russel wav file from
BEL			Brussel	Brus_short.wav	nel	Dutch	yes	endonym		50.83	4.35	Kathleen Van Doren 2006	Bruxelles=B russel wav file from
BEL			Brussels		eng	English		variant	Bruxelles			?	
BEL			Brüssel		est	Estonian		variant	Bruxelles			Place Names Database, Institute of Estonian Language, 2001	From Peeter Päll, 2006
EGY	Ⲅⲓⲛⲏⲛⲓ	Survey of Egypt System (SES)	El Qāhira / Al Qāhirah	EIQ_short.wav	ara	Arabic	yes	endonym		30.17	31.25	Working Group on Romanization , 2006	wav file from
EGY			Kairo		deu	German		variant	El Qahira / Cairo			Federal Foreign Office, Germany	From Joern Sievers, 2006
EGY			Kairo		nor	Norwegian		variant	El Qahira			Language Council of Norway www.sprakad.no	From Botolv Helleland, 2006
EGY			Le Caire		fre	French		variant	El Qahira			<i>Nouvel Atlas Universel</i> , Paris	From Henri Dorion, 2006
THA	??????	Royal Institute of Thailand System 2000	Nonthaburi	Nonth_short.wav	tha	Thai	no	endonym		13.85	100.53	Working Group on Romanization , 2006	wav file from
THA			Nonthaburi		eng	English		endonym				?	

