Database of Exonyms and Other Variant Names (EVN-DB) within the EuroGeoNames Service

Roman Stani-Fertl

DOI: 10.2436/15.8040.01.261

Abstract

Within the geographical names service “EuroGeoNames” a supplement database was established to cover toponyms not contained in the national names databases (i.e. exonyms and variants of endonyms). The EVN-DB is populated with exonyms and variant names from official lists of exonyms, compilations of toponyms of national and international names authorities and publications by toponymic experts. All names are linked with their respective standardized endonym.

The database holds a wide range of attributes and information that has never been available in one single database. It is a synopsis of a wide range of published and unpublished documents. EVN-DB provides for each exonym and variant name elements like language, status, gender, grammatical number, and others. The report shows the structure of the database and introduces its elements. Data can be retrieved from any user via the homepage of EuroGeoNames.

*****

EuroGeoNames and EVN-DB

EuroGeoNames (EGN) is a project to build a common European platform for geographical names. National Mapping and Cadastre Agencies (NMCAs) of participating countries provide EuroGeoNames with their national names data.

These national names databases cover the territory the national authority is representing, in the official languages of the respective country. The core of the project is the EGN Central Service in which the data of the connected NMCAs (= EGN Local Services) are retrievable.

The boxes in Fig. 1 labelled "EGN Local Service" stand for each country connected to EuroGeoNames and having provided its national names data to the EGN Central Service.

Every user can find toponyms of these connected countries via the Reference Application. Data are stored in the Central Service. The database of exonyms and other variant names is part of the EuroGeoNames project.

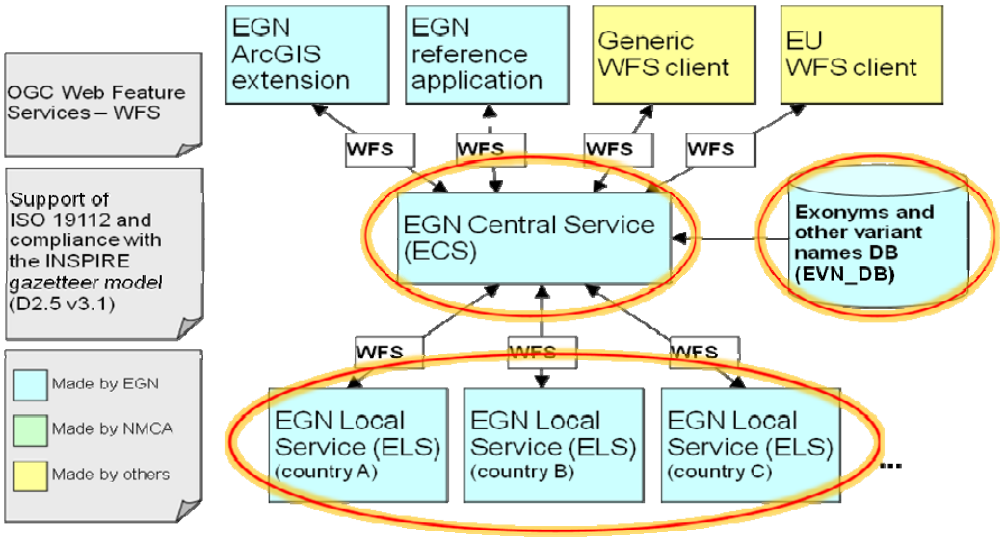

Fig. 1: Architecture of EuroGeoNames

Objectives of the EVN-DB

Not all geographical names in common use are part of the national names databases. There is a wide range of toponyms which are used in everyday language but which cannot be found in toponymic databases of NMCAs. Databases of the NMCAs do not cover all geographical names referring to geographical features in their respective country. With few exemptions databases of NMCAs include only toponyms

- ✓ of geographical features within their national territory,
- ✓ in the officially recognized language(s) of the country, and
- ✓ in their accurate standardized and approved form.

That means that specific types of toponyms other than those mentioned above could not be found in the databases of NMCAs. Certain name forms are missing. There are no exonyms and other variant names for features of the respective territory in the national names databases. If someone is searching for toponyms like exonyms or certain variant name forms of geographical feature one would not get any results.

To be able to provide the user of EuroGeoNames with such names the database of exonyms and other variant names was created.

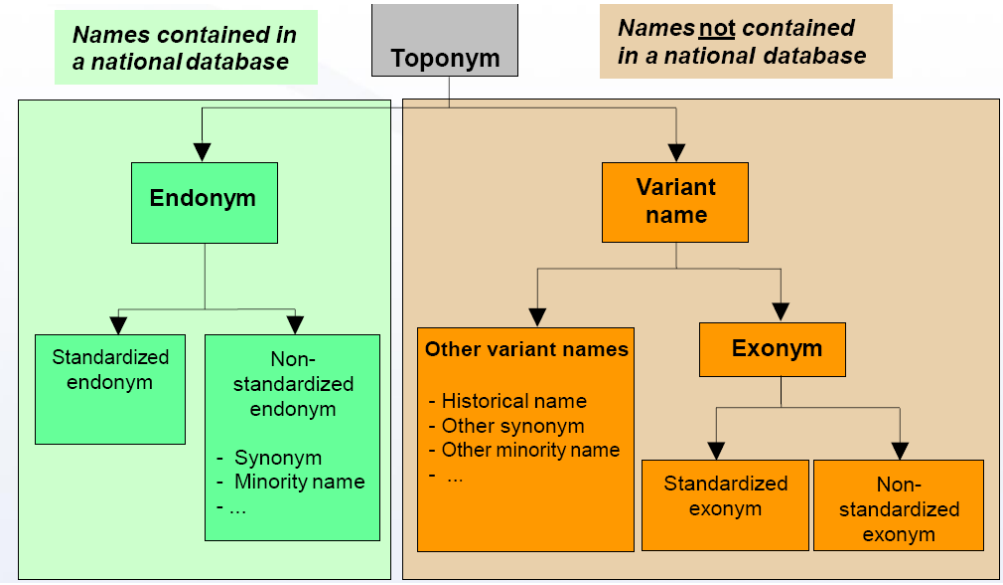

Fig. 2: Types of toponyms according to the EuroGeoNames requirements

EGN classification of toponyms

Fig. 2 shows two types of toponyms under the aspect of EuroGeoNames. The left, green box contains all names maintained by the NMCAs in their national names databases. The right, orange box contains all names not part of the national names databases. They are called variant names. The largest group of these variant names are the exonyms. However, there are still other groups like historical names, synonyms, or certain minority names which should be searchable too.

To handle the problem of toponyms, which are not available in the national names databases but are necessary for a satisfying information retrieval a separate database was developed - the database of exonyms and other variant names. In this database one can find

- ✓ exonyms in common use for standardized endonyms,
- ✓ endonymic variant name forms but not standardized, and
- ✓ historical endonyms for standardized endonyms.

The EGN Names database of variant names and exonyms includes all these types of orange toponyms.

Database model

In the following, there are introduced selected elements of the database of exonyms. A detailed documentation of the data model of the EVN-DB, the structure of the tables, and a description of their attributes can be downloaded from the website of EuroGeoNames: www.eurogeonames.org.

Three tables form the core of the database:

- **TAB_Variant_Names** - This table contains all relevant attributes concerning a certain exonym or variant name.
- **TAB_Metadata** - The table metadata provides information about the source of the name form and its attributes.
- **TAB_GazetteerEntry** - The table GazetteerEntry contains the necessary information to identify the appropriate standardized endonym for the respective exonym or variant name and it contains the UID (Unique Identifier) to link the name forms with the EGN Central Service.

Contents of the EVN-DB

At the moment the EVN-DB is populated with names from 18 countries in 29 languages.

For each exonym and variant name a set of attributes is stored in the database. This set of attributes comprises

- name of the feature = exonym or variant,
- variant name ID = a unique identifier for each dataset,
- language,
- grammatical gender,
- grammatical number,
- status,
- the country where the feature is located,
- if necessary the transcription system used to Romanize the toponym,
- metadata = information about the source where the name and its attributes were derived from,
- and additional remarks.

Language Code*)	National	English	French	Spanish	German
baq	euskera; euskara	Basque	basque	vasco	Baskisch
cat	català	Catalan	catalan	catalán	Katalanisch
cos	corsu	Corsican	corse	corso	Korsisch
cze	český jazyk	Czech	tchèque	checo	Tschechisch
dsb	Dolnoserbski	Lower Sorbian	bas-sorabe	bajo sorabo	Niedersorbisch
dut	Nederlands	Dutch	néerlandais	neerlandés	Niederländisch
eng	English	English	anglais	inglés	Englisch
est	n.a.	Estonian	estonien	estonio	Estnisch
fin	suomi; suomen kieli	Finnish	finnois	finés	Finnisch
fre	français	French	français	francés	Französisch
frr	Frasch	Northern Frisian	frison septentrional	frisio norteño	Nordfriesisch
frs	-	Eastern Frisian	frison (oriental)	frisio oriental	Ostfriesisch
fry	frysk	Western Frisian	frison (occidental)	frisio occidental	Westfriesisch
ger	Deutsch	German	allmand	alemán	Deutsch
glg	galego	Galician; Gallegan	galician	gallego	Galicisch
gre	elliniká	Modern Greek (post 1453)	grec moderne (après 1453)	griego moderno	Neugriechisch
hrv	hrvatski	Croatian	croate	croata	Kroatisch
hsb	hornjoserbšćina	Upper Sorbian	haut-sorabe	alto sorabo	Obersorbisch
hun	magyar nyelv	Hungarian	hongrois	húngaro	Ungarisch
ita	italiano	Italian	italien	italiano	Italienisch
lav	latviešu valoda	Latvian	letton	letón	Lettisch
lit	lietuvių kalba	Lithuanian	lituanien	lituano	Litauisch
mul	-	different	différents	diferentes	verschiedene
nno	nynorsk	Norwegian Nynorsk	norvégien nynorsk	nynorsk	Nynorsk
nob	bokmål	Norwegian Bokmål	norvégien bokmål	bokmål	Bokmål

Fig. 3: The table shows a selection of languages spoken in countries, which are connected to EuroGeoNames. For these languages and others more exonyms are available. The languages are sorted by their three letter ISO code.

The attribute “Language”

At the moment the EVN-DB covers 29 languages. These languages comprise not only national languages but also regional languages used in one of the 18 participating countries.

That means that the service includes exonyms in all national and most minority languages of all participating countries. For instance the database comprises names in two Sorbian languages, in three Frisian languages or in four Sami languages.

The attribute “feature classification”

EGN und therefore also EVN-DB has to handle many different national classification systems for geographical features. To solve this problem and to provide users with homogenous information for all participating countries there has been developed a feature classification. It consists of one main level and one sub level. The main level comprises eight feature classes as shown in Fig. 4 and indicates the main category the feature (geographical object) is assigned to. The 2nd level is a sub level and splits the main feature class into sub categories.

add new record

	Code	Feature Class	Description	Examples
Edit Delete	1	Countries, administrative units, and other areas	Countries, territorial units of a country for administrative purposes and other manmade areas	country, commune, tourist area
Edit Delete	2	Populated places	Buildings for housing of any category like cities, towns, villages, etc.	capital, city, village, hamlet, farmhouse
Edit Delete	3	Non-residential structures and buildings	Any kind of structures and buildings, except populated places, transport, telecommunication, and hydrographic features	mine, park, incinerator, church, survey point
Edit Delete	4	Transport and telecommunication features	Structures related to transports and telecommunications	airport, pier, road, byway, tunnel, antenna site
Edit Delete	5	Terrain features	Land features of natural environment including vegetation. In general they are natural elements but can be modified by man	island, cape, mountain, peak, forest
Edit Delete	6	Hydrographic features	Natural or man-made features related to water	bay, lake, brook, polder, spring, dam, glacier
Edit Delete	7	Conservation areas	Terrain or hydrographic conservation areas of natural environment and World Heritage Sites	nature reserve, national park, world heritage site
Edit Delete	8	Miscellaneous	Other type of features not included in subclasses 1 to 7	

Fig. 4: 1st level of the feature type catalogue developed and applied by EuroGeoNames

E.g. code 7 "Conservation areas" has two sub categories namely:

- World Heritage Sites (World Heritage sites)
- Conservation areas of natural environment (nature reserves, national parks, wilderness areas)

Where do the data come from?

The main sources for exonyms and variant names are lists compiled by national names authorities and toponymic experts. They provide more or less comprehensive lists of standardized exonyms and other variant names in the language(s) they are representing. The span of sources for variant names and exonyms ranks from officially approved lists, published by governmental authorities to semi-official gazetteers and tables compiled by toponymic experts. These data form the core of the EVN-DB. The number of names entered and their attributes depend highly on the quality of sources provided by authorities and experts. Four types of data sources represent different kinds of data quality.

1. Officially approved lists of exonyms issued by ***governmental bodies*** for use in official correspondence / documents;
2. Lists of standardized exonyms issued by public or semi-public bodies (***names authorities***) but not officially approved;
3. Lists of variant names / exonyms of a high scientific standard published by ***research institutes*** and ***experts*** on toponymy;

4. Other collections of exonyms;

The quality of these four types mirrors in the field “status” and “classification” in the EVN-DB.

Status of Exonyms and other variant names

The attributes *classification* and *status* describe the quality of the different types of sources used for the EVN-DB:

Toponyms are differentiated into three classes: endonym variant, standardized exonyms, and not standardized exonyms.

Additionally standardized exonyms are distinguished by their status into

- approved and other (= not approved).

What does “**standardized**” mean? A toponym is “standardized” if it is sanctioned by a name authority. All other name forms of geographical features which are not sanctioned by a name authority are not standardized. They are classified with status “other”.

What does “**approved**” mean? A toponym is approved if it is enacted by a legal body, for instance a governmental authority or the Ministry for foreign affairs or others. These approved name forms have a legal status and their use is obligatory in selected situations and documents.

Classification	Status
n.a.	n.a.
endonym variant	approved
exonym, standardized	approved
endonym variant	other
exonym, not standardized	other
exonym, standardized	other

Fig. 5: Five occurring combinations of *classification* and *status* used in the EVN-DB

While in the EVN-DB the classification in combination with the status allows five types the Web Service of EGN the Reference Application summarizes it and shows just three types, namely “other_endonym”, “exonym_standardized” and “other”. Nevertheless, the indication of the status of a toponym is one of the major milestones of EGN.

TAB-Metadata

Information about the data source and its attributes is compiled in the table TAB_Metadata. In this table one can get information about the source where a specific variant name or exonym came from. Depending on the type of source different attributes are available. The table comprises the following fields:

- title and subtitle of the publication,
- publisher and editor,
- name and contact of the body or person providing the data,
- accuracy or version of the data.

Connecting the two data models

EVN-DB is a non-public and standalone database, which handle their data separately. To make EVN-DB data accessible for EGN users these two data models must be linked. The data of the exonyms database must be integrated in the EuroGeoNames Central Service.

But how this is done? Each dataset in the EVN-DB has a unique Spatial Object Identifier and this ID of an exonym or variant name is connected to its appropriate standardized endonym in the EGN Central Service.

The "Spatial Object UID" is linked with the > Geographic Identifier of the LocationInstance in the EGN data model and the > EGN-Variant-ID of the EVN data model.

By using the "Spatial Object UID" in the EVN-DB and linking this identifier with the appropriate "Geographic ID" in the EGN data model the exonym is connected with the appropriate geographical feature and therefore with its appropriate national standardized endonym. (Fig. 6)

This is the way to integrate the data from the EVN-DB into the Central Service of EGN.

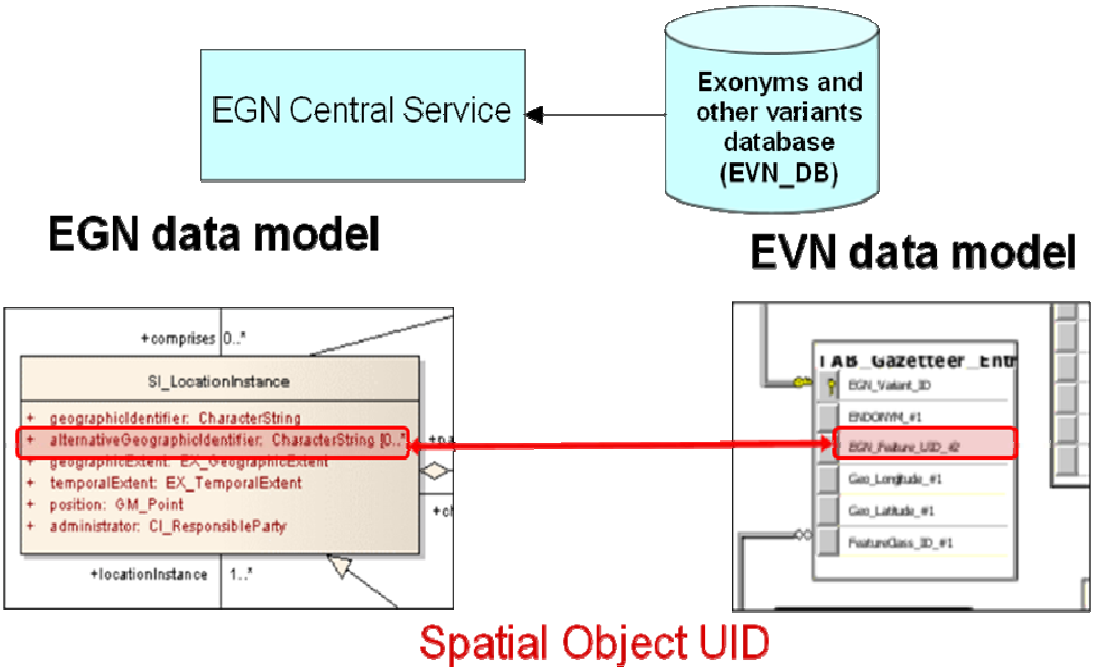

Fig. 6: The integration of the Database of exonyms and other variant names into the EuroGeoNames Central Service is done by the Spatial Object Unique Identifier

Accessibility

Fig. 7 is a screenshot made from the EGN web service – the Reference Application. It shows how data are made accessible within EuroGeoNames. The example shown is the exonym Lipsko.

Searching for Lipsko results in three hits. The first name is the Slovakian, the second the Croatian and the third the Czech exonym for the German town Leipzig.

By clicking on one blue name, in our example on the Czech exonym, the window "Feature Information" opens and all available attributes are displayed.

Within this feature information the following elements come from the EVN-DB:

- the name, the exonym or name variant itself – Lipsko;
- Custodian – it shows that it is an entry in the EVN-DB;
- Status – standardized exonym ;
- Language – cze – It's the Czech exonym for the German town;
- some additional alternative name forms if available - Lipsk, Lipsia and others;
- and the Data Source – Index ceskych exonym.

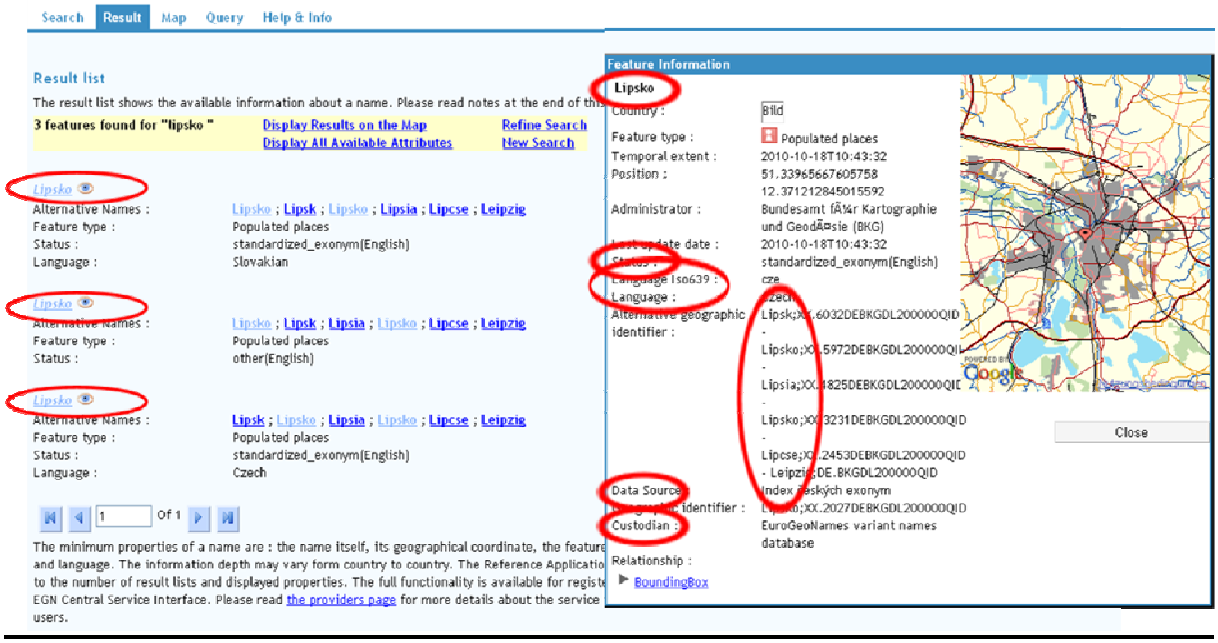

Fig. 7: Screenshot of the EGN Reference Application with the search results for *Lipsko* and the available attributes.

Participating Countries

At the moment exonyms and variant names in national languages of all countries connected to EuroGeoNames are available. These countries are Austria, Belgium, Croatia, Cyprus, Estonia, Finland, France, Germany, Latvia, Lithuania, Norway, Slovenia, Spain, and The Netherlands. Additionally there are variant names and exonyms for the Czech Republic, Slovakia, Hungary and Turkey in the EVN-DB.

References

EuroGeoNames: www.eurogeonames.org

Abbreviations

CS	Central Service
EGN	EuroGeoNames
EVN-DB	Database of Exonyms and Other Variant Names
ID	Identifier
n.a.	not available
NMCAs	National Mapping and Cadastre Agencies
UID	Unique Identifier

In summary

The database of exonyms and other variant names (EVN-DB)

- ✓ is a standalone non-public database
- ✓ contains exonyms and variant names not found in the national names data bases which are provided by NMCAs (EGN Local Service)
- ✓ provides a wide range of attributes for each dataset: e.g. status, language, feature type, meta data, etc.

Datasets are linked with the geographical feature of the national names databases and therefore with their standardized endonyms as long as a standardized endonym is available in the EuroGeoNames Central Service.

The EVN-DB is a supplement database to the national names databases, which adds additional information to geographical features and therefore raises the value of the standardized endonyms of NMCAs.

At the moment it contains about 7,600 datasets for 18 countries in 29 languages.

Roman Stani-Fertl
Feldstraße 88
A-3420 Kritzendorf
Austria
roman.stani-fertl@univie.ac.at