

The Renewed Implementation of the EuroGeoNames Central Service

Pekka Latvala
Finnish Geodetic Institute
Geodeetinrinne 2
Masala, Finland
pekka.latvala@fgi.fi

Lassi Lehto
Finnish Geodetic Institute
Geodeetinrinne 2
Masala, Finland
lassi.lehto@fgi.fi

Jaakko Kähkönen
Finnish Geodetic Institute
Geodeetinrinne 2
Masala, Finland
jaakko.kahkonen@fgi.fi

Abstract

The EuroGeoNames service is a gazetteer web service that serves harmonized, pan-European geographical names data. This paper describes the EGN central service renewal process where the EGN central service database implementation and the related Web Feature Service implementation were replaced with new solutions. In addition, the new WFS implementation was enhanced with a custom front-end WFS application that adds a LANGUAGE-parameter to the WFS interface that can be used for requesting language-specific information. The finished EGN central service implementation was copied from the development environment into a cloud service environment for operational use. The new service implementation was further expanded with new web-friendly service interfaces, including support for JSON, KML and WMTS outputs. In the future, the INSPIRE Geographical Names download services are seen as a potential input and output format for the renewed EGN central service.

Keywords: Gazetteer Service, Web Services, Geographic Names.

1 Introduction

In recent years, there has been an increasing interest in the geospatial field for creating harmonized and integrated pan-European spatial data sets. In Europe, each country has traditionally collected their own spatial data separately from other countries from the viewpoint of their own national requirements. In most cases, the data sets that different countries produce are not directly compatible with each other. A common way of harmonizing heterogeneous data sets is to create a common data schema, to which the different data sets are transformed.

Another growing trend and demand in Europe is to increase the utilization of the collected spatial data. A common requirement is that the data must be easily available to users via the Internet. The requirement can be achieved efficiently through the use of spatial web services that use standardized interfaces for handling the communications between the client and the server. The Open Geospatial Consortium (OGC) has created many of these kinds of interfaces, including the Web Map Service (WMS) interface [1] that can be used for serving rendered visualizations of the spatial data and the Web Feature Service (WFS) interface [2] that can be used for serving the actual spatial data.

The EuroGeoNames (EGN) service [3], created by Eurogeographics [4] is one of these kinds of spatial web services that serve combined and harmonized data content. The EGN Service is a gazetteer type of web service and it serves pan-European geographical names data that have been collected from several European countries.

During the years that the EGN service has been operational, its implementation has been found to be in need for a renewal. This paper describes the renewed EGN central service implementation. The original EGN service implementation is introduced in the second chapter. The third chapter presents the renewed EGN central service implementation. The paper ends with discussion and conclusions of the project.

Some examples of similar types of gazetteer services are the SCAR Composite Gazetteer of Antarctica [5] that serves the geographical names data that are combined together from different Antarctic territories, the USGS Prototype Web Feature Service Gazetteer [6] that contains data from the United States of America and uses the OGC's Web Feature Service Gazetteer (WFS-G) Application Profile as its data model and the JRC Fuzzy Gazetteer [7] that serves global geographical names data.

2 The Original EGN Service

The EGN project was originally executed between the years 2006 and 2009 as part of the European Commission's eContentPlus programme. The EGN service's data contents consist of geographical names data that covers 14 European countries [8]. The data are classified into 34 feature classes that form 8 top-level feature categories 26 sub-categories.

One key requirement for the EGN service has been the ability to support multilingualism and to deliver the geographical names data in several languages. For this purpose, a separate exonyms and variant names database [9] has been created in the EGN project that contains multilingual geographical names data in up to 17 languages from a number of important locations.

2.1 Architecture

The original EGN Service is implemented with distributed service architecture that consists of one central EGN service that is hosted by BKG Germany and several national EGN services that are hosted locally in the EGN data provider countries. The data is maintained at the national services and the central service is used for collecting all data from the national services into one database, from where the data is served to the end-users. The data contents of the central

service database are updated periodically by replacing wholly the data sets from each of the national services.

All EGN service nodes use a common data schema to ensure that different countries produce compatible data in a harmonized form. The EGN data schema is based on the abstract ISO 19112 schema [10] that has been extended further. The service's database schema and its input and output schemas are based on this common data schema.

2.2 EGN service renewal

The need for the EGN service renewal has arisen from issues that have been identified with the existing service implementation. The main issues are related to the complexity of the service's database structure and its output schema. The database structure is highly complex, containing many interdependencies between tables. The complexity makes the database structure hard to understand and the data update procedures difficult to execute. The situation has ultimately led to the execution of faulty update procedures that have caused data corruption in the EGN central service database. The service's output schema is also unnecessarily complicated and it is considered to be too heavy for web application use.

Another aim in the renewal project is to enhance the service's output capabilities with new web-friendly service interfaces that are based on modern web technologies. The adoption of new service interfaces will broaden the capabilities for the EGN data utilization.

The problems identified with the existing EGN service implementation have led into a situation where it has been difficult to convince new countries to join into the EGN service. The service renewal aims to get more countries to provide data for the EGN service. In the future, the target scope is to expand the number of EGN data providers into all of the 27 EU Member States.

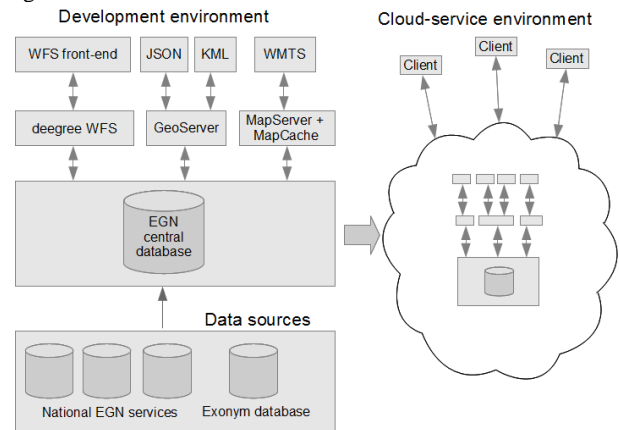
3 The Renewed EGN Central Service

In the EGN service renewal process, only the EGN central service implementation was renewed and the existing national EGN service nodes were left unmodified. In central service, the database implementation and the related WFS implementation were replaced with new implementations. The new WFS implementation is the main access point of the renewed service. It consists of two components; the deegree WFS application and the additional front-end WFS application. The EGN service's access capabilities were also enhanced with a number of new web-friendly service interfaces that produce output in JSON, KML and WMTS formats (Figure 1).

Architecturally, the renewed central service implementation consists of two identical service environments; the development environment and the cloud-service environment. The development environment is maintained by the Finnish Geodetic Institute and it is used for service development and testing of the data update procedures. The cloud-service environment is used for running the operational service instance and it has been set up in the Amazon EC2 service. The cloud-service database is updated periodically by copying

the contents from the development database into the Cloud Service. The operational cloud service environment allows the service's hardware configuration to be changed easily if the use requirements for the service are changed.

Figure 1: Architecture of the Renewed EGN Central Service



3.1 Data Model

The data model of the new EGN central service is based on the OGC's Gazetteer Service Application Profile of the WFS (WFS-G) Best Practice document [11]. The WFS-G document specifies an XML schema realization of the same abstract ISO 19112 schema, to which the original EGN schema is based on.

The WFS-G Application Profile schema is used as the basis for the new EGN central service's database schema and also for the new WFS implementation's main input and output schema. The use of the same schema internally in the database and externally in input and output makes the data maintenance easy and ensures high query performance.

The WFS-G Application Profile schema defines feature types *SI_LocationInstance*, *SI_Gazetteer* and *SI_LocationType* that are used as output feature types in the new WFS implementation. The *SI_LocationInstance* feature type represents the locations and contains the geographical names that are related to the locations. The *SI_Gazetteer* feature type contains metadata about the gazetteer service. The *SI_LocationType* feature type represents the feature classification of location instances.

The new WFS-G schema contains some important changes when it is compared with the original EGN schema. First, the new schema establishes one-to-many relationship between the location instances and the geographical names, whereas the original EGN schema had specified this relationship to be one-to-one. This change follows better the abstract ISO 19112 schema where this relationship has been modelled to be one-to-many.

Other important changes in the new schema are related to the individual schema elements that do not have places in the new EGN schema. The new schema does not have an element for representing the original schema's *complexGeometry* element and this information was left out from the new central service database. This is not a significant problem, because the EGN database is not the most appropriate place for storing

the complex geometry information that was already missing from many of the national EGN services. Furthermore, the *complexGeometry* information is not maintained in the EGN service database, and if the information is needed, it can be retrieved with feature identifier from the database where it is stored and maintained.

Another change is that the new schema does not have a corresponding element for the *classificationScheme* element. This element had a fixed *EuroGeoNames* value in the original EGN database and it can be easily left out because it does not contain any important information.

Yet another important change is that the new schema does not have a corresponding element for original schema's *status* element. In the original schema the *status* element can have 4 different values and it was mapped into *type* and *primary* elements (Figure 2).

Figure 2: Mapping of the *status* element into *type* and *primary* elements

status	type	primary
official_endonym	official	true
other_endonym	official	false
standardized_exonym	variant	true
other	variant	false

3.2 The new Database Implementation

The new EGN central service database was implemented with PostgreSQL/PostGIS database. The structure of the database reflects the feature types that are defined in the WFS-G schema. The database consists of six tables (Figure 3). The *si_location_instance* and *alternative_geographic_identifier* tables correspond together to the WFS-G schema's *SI_LocationInstance* feature type. The *si_location_type* table reflects the *SI_LocationType* feature type and the *si_gazetteer* table reflects the *SI_Gazetteer* feature type. The *location_type_link* and *gazetteer_location_type* tables contain links to HTTP GET queries for *SI_LocationType* features and they are used for retrieving the feature classification information for language-specific queries when creating the WFS output.

The data contents of the new EGN central service database were downloaded mostly from the existing national EGN services. Some countries that had not implemented the national WFS services provided the data directly as files. The source data were transformed from the original EGN schema into the new EGN schema and loaded into the database with a custom-made Java application that executes the transformations with Extensible Stylesheet Language Transformations (XSLT) [12] scripts. The mapping for the transformations is presented in Figure 4.

The data contents of the new EGN database were further expanded with the contents of the exonyms and variant names database that was inserted manually into the new EGN database from the Microsoft SQL Server database.

3.3 The New WFS Implementation and Other Web-Friendly Access Interfaces

The basis of the new WFS implementation was created with the deegree WFS 3.1.2 [13] application by specifying the mappings between the new database and the WFS-G output schema.

In order to support multilingualism better, a separate Java-based front-end WFS application (Figure 1) was created on top of the deegree WFS application. The WFS queries are routed through the front-end WFS application, which modifies both the client's WFS queries and the background WFS service's response documents. The front-end application executes also other WFS queries that fetch additional information that is inserted to the WFS response documents.

The main function of the front-end WFS application is to add a LANGUAGE parameter to the WFS interface. The LANGUAGE parameter can be used for retrieving language-specific information about the gazetteer service's feature classifications. The values of the LANGUAGE parameter are three-letter language codes and the supported language codes are listed in the WFS service's *GetCapabilities* response document. The use of the additional LANGUAGE parameter for the WFS interface is a familiar method from the INSPIRE Download Services where it is used for specifying the language of the WFS response [14].

Another function of the front-end WFS application is to finalize the schema mapping in order to bypass some difficulties that were encountered when the mapping was created between the database and the WFS-G output. The difficulties were related to the creation of *xlink* and *xml:lang* attributes that are now added to the WFS output in the front-end WFS application.

The new EGN central service was also extended with new web-friendly access interfaces. The new interfaces that were created include the JavaScript Object Notation (JSON) and the Keyhole Markup Language (KML) interfaces that were created with the GeoServer application and the Web Map Tile Service (WMTS) interface [15] that was created with MapServer and MapCache applications. The WMTS interface was further utilized in the demonstrative OpenLayers-based client application that can be used for testing the new EGN central service.

4 Discussion

The new WFS-G Application Profile schema seems to be suitable for the provision of the EGN data content, because all essential information can be presented with it. Although some elements had to be left out from the new schema they are not considered to be essential in the EGN context. The new database implementation simplifies the database by reducing the amount of tables from 22 to 6 and makes an important structural change by establishing a one-to-many relationship between the locations and the relating names. The new output schema is structurally lighter than the original EGN schema and the new web-friendly access interfaces allow wider utilization of the EGN data.

The renewed EGN database contains approximately 2.9 million names from 2.85 million locations and it covers 17

countries. In the future, the EGN Central Service is planned to be expanded with data that are taken from the INSPIRE Geographical Names (GN) services. This would greatly increase the amount of data in the EGN central service as many countries that are not currently EGN data providers are required to create the INSPIRE GN services. The INSPIRE GN schema can be potentially used also as EGN service's additional output schema, which would make the EGN central service a centralized INSPIRE GN data store and the existing national EGN service data might be used for delivering the INSPIRE GN content.

The use of the INSPIRE GN schema as input or output for the EGN service requires the implementation of schema transformation procedures between the INSPIRE GN schema and the new EGN schema. Some initial tests were executed in this project for transforming the data between the INSPIRE GN schema and the new EGN schema. The results of these transformations are encouraging and the executing of these transformations seems feasible.

5 Conclusions

The new EGN central service implementation replaces the original EGN central service database with new and simpler solution and replaces also the existing WFS implementation that is the service's main access interface. The new service uses the WFS-G Application Profile schema as its main input and output schema and it is able to represent all the essential data elements from the old EGN schema. The support for multilingualism is handled in the separate front-end WFS application that adds a LANGUAGE parameter to the WFS interface and it can be used for retrieve language-specific feature classification information. The new web-friendly service interfaces enhance the possibilities for the utilization of the EGN data. The new EGN central service implementation was moved for operational use into a cloud-service environment. In the future, the INSPIRE GN schema is seen as a potential input and output format for the new EGN service.

References

- [1] OGC, (2012). Web Map Service. At: <http://www.openeospatial.org/standards/wms>
- [2] OGC, (2012). Web Feature Service. At: <http://www.openeospatial.org/standards/wfs>
- [3] Eurogeographics, 2012, EuroGeoNames. At: <http://www.eurogeographics.org/eurogeonames>
- [4] Eurogeographics, (2012) , Eurogeographics home page. At: <http://www.eurogeographics.org/>
- [5] Scientific Committee on Antarctic Research (SCAR), 2012, Composite Gazetteer of Antarctica. At: <https://data.aad.gov.au/aadc/gaz/scar/>
- [6] U.S. Geological Survey, 2012, USGS Prototype Web Feature Services Gazetteer (WFS-G). At: <http://frameworkwfs.usgs.gov/wfs/>
- [7] Joint Research Centre of the European Commission, 2012, JRC Fuzzy Gazetteer. At: <http://dma.jrc.it/services/fuzzyg/>
- [8] Zaccheddu, P-G. and Overton, D., (2011), EuroGeoNames (EGN) – Implementing a sustainable European gazetteer service. UNGEGN Working Paper No. 38, 2011. At: http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/26th-gegen-docs/WP/WP38_EGN_item%209_UNGEGN26.pdf
- [9] BKG, 2005, Exonyms guidelines synopsis. At: http://www.eurogeographics.org/sites/default/files/080205_D5.1_Exonyms%20guidelines%20synopsis.pdf
- [10] ISO TC211, 2003, ISO 19112:2003 Geographic Information - Spatial referencing by geographic identifiers. At: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=26017&commid=54904
- [11] Harrison, J, Vretanos, P.A., (ed), 2012, Gazetteer Service – Application Profile of the Web Feature Service Best Practice. Version 1.0 Available At: https://portal.openeospatial.org/files/?artifact_id=46964
- [12] W3C, XSL Transformations (XSLT) Version 1.0. At: <http://www.w3.org/TR/xslt>
- [13] deegree, 2012, deegree project home page. At: <http://www.deegree.org/>
- [14] INSPIRE, Technical Guidance for the implementation of INSPIRE Download Services At: http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_Download_Services_3.0.pdf
- [15] OGC, 2012, OpenGIS Web Map Tile Service Implementation Standard. At: <http://www.openeospatial.org/standards/wmts>

Figure 3: The structure of the renewed database

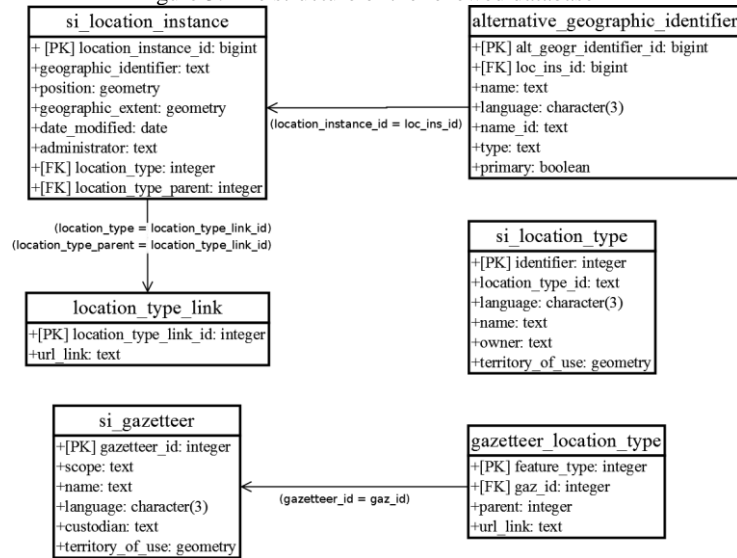


Figure 4: The mapping from original EGN schema into the WFS-G AP schema

Original EGN schema type	Original EGN schema element	WFS-G AP schema type	WFS.G AP schema element
egn:LocationInstance		is019112:SI_LocationInstance	
	egn:spatialObject/egn:ComplexGeometry/egn:spatialObjectUID		is019112:geographicIdentifier
	egn:position		is019112:position
	egn:geographicExtent		is019112:geographicExtent
	egn:lastUpdateDate		is019112:dateModified
	egn:administrator		is019112:administrator
	egn:spatialObject/egn:ComplexGeometry/egn:spatialObjectUID		is019112:spatialObject
	egn:locationType/egn:FeatureClassification/egn:featureClassificationID		is019112:locationType
egn:LocationInstanceName		is019112:SI_LocationInstance/ is019112:alternativeGeographicIdentifiers/ is019112:alternativeGeographicIdentifier	
	egn:name/egn:LocalisedCharacterString/egn:plainText		is019112:name
	egn:name/egn:LocalisedCharacterString/egn:language		is019112:name/@xml:lang
(egn:LocationInstance)	egn:geographicIdentifier		is019112:nameID
	egn:status		is019112:primary
	egn:status		is019112:type