An Implementation of the OGC's WFS Gazetteer Service Application Profile

CASE: The EuroGeoNames Central Service Renewal

Lassi Lehto, Pekka Latvala and Jaakko Kähkönen
Department of Geoinformatics and Cartography
Finnish Geodetic Institute
Masala, Finland
lassi.lehto@fgi.fi, pekka.latvala@fgi.fi, jaakko.kahkonen@fgi.fi

***Abstract*—Gazetteer Service is a Web Service that provides access to geographical names data sets. The service interface has been recently specified by the Open Geospatial Consortium (OGC) as an Application Profile (AP) of the well-known Web Feature Service (WFS). The service interface was implemented in the context of the EuroGeoNames Central Service renewal project. This work represents one of the first real-life implementations of the specification and serves as a test of its adaptability in the European multi-lingual environment. The experiences are encouraging.**

Keywords-Gazetteer Service; geographical names; geocoding; Web Services.

I. INTRODUCTION

A. EuroGeoNames data set

The EuroGeoNames (EGN) data set has been originally put together in a major EU-project [1]. The main players of the project consortium included for instance the National Mapping and Cadastral Agencies (NMCA) of Germany, Austria and Slovenia and the European NMCAs' co-operation body EuroGeographics. In the initial inventory phases altogether 28 European countries took actively part in the project providing information on their place names data sets. In the end the number of countries officially committed to grant access to their national source data sets fell to a bit lower figure of 15. Still, this number of countries provides a significant share of the European geographical names content [2]. The current EGN data coverage is presented in Figure 1.

The main guiding principles of the EGN project include support for multilingual place names. A special emphasis has been put on the collection of a separate exonym database that contains names of certain important European geographical features in up to 17 different languages. The name sources of the EGN service are regarded as the most authoritative in Europe, as they are maintained by the official national agencies.

B. EuroGeoNames service architecture

The service architecture used in the provision of the EGN content is based on a distributed model, in which each participating NMCA sets up and maintains a national EGN service instance. The centralised EGN service updates it combined geographical names data base from these local nodes. The communication between the local service nodes and the centralised service is based on the Open Geospatial Consortium's (OGC) plain Web Feature Service (WFS) access interface specification. The service architecture was initially seen as a hierarchy of cascading WFS nodes, but was eventually modified for performance reasons to include a centralised European level cache. The EGN Central Service has been operated the German NMCA, contracted by the EuroGeographics [3].

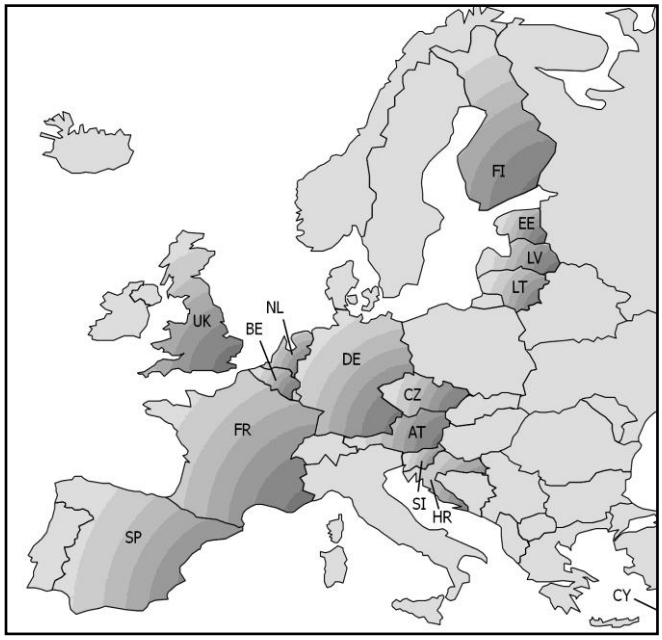

Figure 1. The current EGN coverage.

EuroGeographics has contracted the Finnish Geodetic Institute (FGI) to develop a renewed solution for EGN Central Service [7]. The main results gained so far in the still ongoing project are described in the following chapters. Motivation for the work is described in Chapter II. In Chapter III the standardization process of the Gazetteer Service interface is reviewed. Chapter IV focuses on the renewal of the EGN data model and the setup of the new Cloud Service-based EGN Central Service. Chapter V discusses in more detail the service interfaces used and explains the challenges encountered while implementing the WFS Gazetteer AP. Chapter VI contains the outlook detailing some plans for future work on the EGN data set and the service infrastructure. Finally, conclusions are drawn in Chapter VII.

 ISBN: 978-1-61208-251-6

II. MOTIVATION

The existing EGN Central Service implementation has been found to be outdated and in need of a renewal. The main problems with the existing Central Service are related to its highly complex database structure that is difficult to understand and hard to manage. Because of the complicated data structures, the data update procedures are difficult to execute. This has occasionally resulted in data corruption in the Central Service database, because of a fault in the update operation. In addition, the Service's output schema is considered to be too heavy for applications that operate in the Web Service environment. These difficulties have led into situation where it is difficult to convince new European countries to join the EGN service, partly because of the very demanding data structures present in the current national EGN databases. Also the use of the data set has been limited by the overly complex data structures and the too complicated access interface.

The renewed implementation of the EGN Central Service tries to alleviate the problems identified with the earlier EGN implementation. The main changes in the new Central Service are: adoption of a simplified data model and new input and output schemas, a Cloud-service based implementation architecture and the provision of new Web-friendly service interfaces.

III. GAZETTEER SERVICE

The OGC has been working on the service interface specification for a Gazetteer service for several years. Already in 2001 a Discussion Paper was published on the topic with a title: "Gazetteer Service Draft Candidate Implementation Specification" [4]. In 2006 an official OGC Specification was published describing the Gazetteer Service as a profile of the WFS service interface, denoted by the acronym WFS-G. Already in this early version of the specification the applied information model was based on the work of the ISO TC211 on spatial referencing by geographic identifiers that has been published as an ISO international standard 19112 in 2003[5]. The latest revision of the Gazetteer Service specification has been released by the OGC as an Application Profile (AP) of the WFS in Feb 2012 [6].

According to the ISO 19112 data model a gazetteer is comprised of a set of location instances that are of a specific location type and are referenced by one or more geographical identifiers, i.e. place names. This basic principle of taking place as the basic concept and attaching potentially several individual names to a single place is in the core of the ISO Gazetteer model and subsequently one of the main challenges in the original EGN data model and database schema development.

The Gazetteer Service access interface is based on the WFS standard with a few adjustments that specify, how one can conclude that a certain WFS actually is a Gazetteer Service. The AP makes it easier to access place names data from a Gazetteer Service by explicitly defining the data model in which the place names data is to be delivered.

IV. EUROGEONAMES DATA MODEL

The original EGN data model was developed in course of a large EU-project. Due to some design principles it grew overly complex. This was caused partly by tight adherence to a standardized way of handling multilingualism and partly because of an all-encompassing information model; a typical result of a team work in large project consortiums.

The database schema of the renewed EGN Central Service is based on the abstract ISO 19112 schema that has been fine-tuned and made concrete in the OGC's Gazetteer Service AP Best Practice document. The same ISO 19112 schema is also used as the main data input and output schema in the renewed service. The fact that the same schema is used internally in the service and also for its input and output ensures the best possible query performance and easy data maintenance.

The original EGN database structure was very complicated with many interdependencies among the database tables. As a result of the redesign and simplification the number of database tables was reduced from 22 to 6. Although a significant number of data items have been dropped, the remaining content seems to satisfy most of the typical query requirements encountered in gazetteer-related applications.

The database structure is built around the two main concepts present already in the original ISO 19112 abstract model: a location (`si_location_instance`) that can be referenced with potentially several geographical names (`alternative_geographic_identifier`). Although simple, this fundamental relationship makes the data structure to go beyond what is supported by the Simple Features data model in OGC specifications. These two object types form the core of the database, stored in two big database tables. In addition the database contains a table for storing location type information (`si_location_type`) and another table for the general information relating to the whole gazetteer data set (`si_gazetteer`). Two auxiliary tables are needed for linking the gazetteer table to location type table (`gazetteer_location_type`) and for linking the location types to an URL, from which the corresponding type classification information can be retrieved (`location_type_link`).

Multilingualism has been in the core of the EGN development since the beginning. There are three different levels in the support for different languages in the EGN service: 1. Exonyms are included into the data set for the most important European geographical features, 2. The most important attributes related to the place names are available in all major European languages, 3. The user interface of the EGN reference application is multilingual.

In the renewed EGN Central Service solution the language support has been simplified and streamlined. The exonym data set was integrated to the newly structured database unchanged. However, the original structure closely reflecting the ISO-standardized PT_Free_Text model was replaced by much simpler mechanism of maintaining language information inside the `si_location_type` table. Some multilingual data items were removed because of the

 ISBN: 978-1-61208-251-6

simplified data model. On the service interface level the INSPIRE-standardized mechanism involving a dedicated LANGUAGE-parameter is taken into use. Location type information is encoded into the corresponding XML-element of the output schema in the requested language. In addition, the response message contains a URL-reference from which the client application can retrieve a full description of the location type, again in the originally requested language. LANGUAGE-parameter does not affect the treatment of the exonyms-related content.

User interface development for the renewed EGN Central Service is currently ongoing.

V. SERVICE ARCHITECTURE

The renewed EGN Central Service is built according to the OGC WFS Gazetteer Service AP. It is based on the recently published OGC's Best Practice paper on the subject [2]. The national EGN Services are still hosted by the NMCAs of the EGN data provider countries for storing the geographical names data on the national level. The EGN Central Service is now maintained by the FGI as a Cloud Service instance and it contains the combined geographical names data that are collected from the national EGN Services. The Central Service is implemented as the OGC Gazetteer Service node and uses the deegree software platform as the WFS implementation and the PostgreSQL database with its spatial PostGIS extension as the database platform. The initial data upload from the national services is performed by retrieving the full datasets from the national EGN Services and replacing completely the corresponding data in the Central Service database.

Architecturally the renewed EGN implementation consists of the original national EGN service nodes, the new centralized EGN production database and the Cloud Service-based EGN Central Service that contains its own service database. Clients can connect to the service via various different access interfaces on the application layer. The EGN service architecture is thus built according to a four-layer model, see Figure 2.

The national EGN Services and the original EGN schema that they support are left untouched and their data contents are integrated to the new centralized production database that is maintained by the EGN Operator. The corresponding production database is kept as a separate data store, on which periodic updates are run by a semiautomatic process from the national EGN Services. The update process is carried out incrementally via monitoring software that queries the WFS interfaces of the national EGN Services for any updated or added data, transforms such data into the OGC Gazetteer Service AP schema and uploads the update into the production database.

The EGN Central Service is established as a Cloud Service-based OGC Gazetteer Service instance that provides access to the EGN content that is stored in the service database (EGN Cache). The service database is updated periodically by copying the production database to the Cloud. Before the update process is executed, quality checks are done to make sure the correctness and internal consistency of the production database.

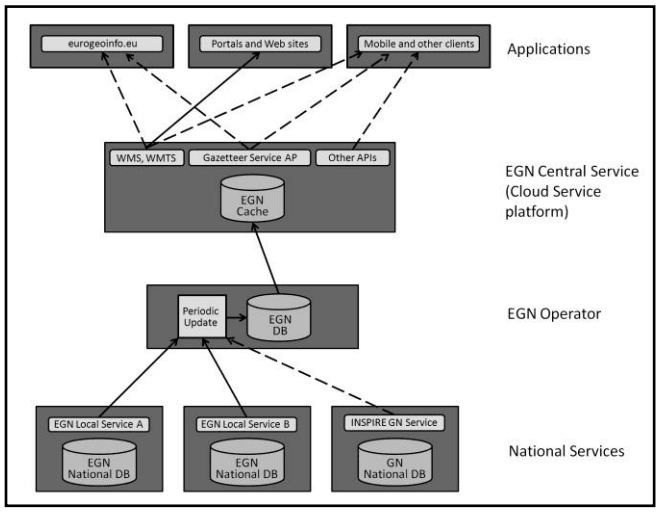

Figure 2. The new Cloud service-based EGN service architecture. The EGN Central DB maintained by the EGN Operator is periodically copied to the Cloud Service platform, from which various different client applications are supported.

VI. FUTURE DEVELOPMENTS

In future, the EGN Central Service is foreseen to work also as an integrated source for European-wide provision of INSPIRE Geographical Names (GN) data, retrieving its data content from the national INSPIRE GN Download Services. The GN data can be updated to the EGN Central Service through a similar update procedure that is used with the local EGN nodes; by first transforming the data from the INSPIRE GN schema into the OGC Gazetteer Service AP schema and then updating the transformed data into the production database. The output will be similarly configured to support INSPIRE GN schema through an internal schema transformation process.

Initial tests on running the GN to EGN schema mapping seem promising. Most of the current simplified EGN schema can be filled in from the GN data. For update purposes it is critical that the values for the lifecycle –related attributes (beginLifespanVersion, endLifespanVersion) are available. Matching the rather coarse location type classification of the INSPIRE GN schema to the much more detailed EGN classification will be a challenging task.

The main access interface to the EGN Central Service is the interface as defined in the OGC Gazetteer Service AP. Visualisation of the EGN data can be accessed via a Web Map Service (WMS) interface. The EGN Central Service will be expanded also with other access interfaces. The probable candidates include JavaScript Object Notation (JSON)/JavaScript Application Programming Interfaces (APIs) and Keyhole Markup Language (KML) encodings.

In Figure 3 a simple client application accessing the renewed EGN service is shown. This application is connected to the service both through the Web Map Service (WMS) for the visual map display of place names, and the WFS Gazetteer Service AP interface for the results of the query.

 ISBN: 978-1-61208-251-6

Figure 3. A simple visualisation, querying and geocoding UI for accessing the EGN Central Service. The map shows EGN content displayed on top of the Google Maps satellite image backdrop. An exonym query and its results are also shown.

VII. CONCLUSIONS

The initial experiences gained from the EGN Central Service renewal suggest that the OGC Gazetteer Service AP is good candidate for a standardized access to geographical names data content. The adopted approach of using the WFS interface as a basis for accessing geographical names content seems to be well-justified. The basic geocoding functionality is readily supported and at the same time all the available additional attribute information related to the place names can be accessed through the same interface.

The adopted ISO 19112-based information model seems to satisfy all the data requirements deemed as critical in the pre-project data simplification analysis done by the EGN Project Board. Name status attributes were relatively easy to map from the classification applied in the original EGN data to the value spaces adopted in the Gazetteer Service AP. The location type classification of the original EGN data set was used as such. A significant simplification was achieved in the database structure by reducing the number of database tables from 22 to 6.

It is seen as promising solution for the future EGN update process that the national INSPIRE GN Download Services could be used as the data source. The schema mapping from the INSPIRE GN schema was tested and it seems feasible. EGN Central Service can be also seen as a promising solution for the provision of the integrated INSPIRE GN data services with full European coverage.

Multilingualism can be supported by applying the INSPIRE-introduced LANGUAGE-parameter. This enables much more elegant and efficient way for supporting different languages than the originally applied approach of returning data in all available languages and letting the client to sort out the desired one. Additional information on the location type is available in the requested language by following a URL-link present in the response and referring back to the service.

Experiences gained in the process of deploying EGN Central Service on the cloud service platform have been entirely positive. The service has run uninterrupted for the whole test period and it is easy to make it adapt to a varying level of user demand.

The work on the EGN Central Service renewal continues. New access interfaces are being developed for the EGN Central Service. These are expected to improve the accessibility of the service and further increase its use in various application areas. The data harmonisation continues with the help of new guidelines to be developed. Coverage of the data set is expected to expand, partly helped by the new simpler data model, partly due to the use of the upcoming INSPIRE GN services.

ACKNOWLEDGEMENT

The EuroGeoNames renewal project has been funded by the EuroGeographics Aisbl.

REFERENCES

[1] P-G. Zaccheddu, “EuroGeoNames (EGN) – developing a European geographical names infrastructure and services,” Proc. INSPIRE Conference, Jun 23-25, 2008, Maribor, Slovenia, CD-ROM.

[2] P-G. Zaccheddu and D. Overton, “EuroGeoNames (EGN) – Implementing a sustainable European gazetteer service,” UNGEGN Working Paper No 38, 2011, at: http:// http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/26th-gegn-docs/WP/WP38_EGN_item%209_UNGEGN26.pdf [accessed: 2012-12-17]

[3] EGN Central Service, at http://www.eurogeonames.eu [accessed: 2012-12-17]

[4] R. Atkinsson [ed.], “Gazetteer Service Draft Candidate Implementation Specification”, at: http://portal.opengeospatial.org/files/?artifact_id=1041 [accessed: 2012-12-17]

[5] ISO TC211,”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 [accessed: 2012-12-17]

[6] J. Harrison and P. Vretanos [eds.], “Gazetteer Service – Application Profile of the Web Feature Service”, OGC Best Practice, 2012, at: http://portal.opengeospatial.org/files/?artifact_id=46964 [accessed: 2012-12-17]

[7] P. Latvala, L. Lehto, J. Kähkönen, A. Jakobsson and S. Urbanas, “Renewal of EuroGeonames Central Service,” Proceeding of the INSPIRE Conference 23-27 June, 2012, Istanbul, Turkey at: http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2012/index.cfm [accessed: 2012-12-17]

 ISBN: 978-1-61208-251-6