EGN (EuroGeoNames) Service based on OGC Gazetteer service AP

Pekka Latvala

Finnish Geodetic Institute
Overview

• Background
• Original EGN service
• Renewed EGN service
• Schema mappings
• Future
Background

• The EuroGeoNames (EGN) project (2006 – 2009) was implemented by EuroGeographics as a part of the eContentPlus project

• EGN service is a Gazetteer service that serves geographical names data from several European countries

• EGN service consists of:
  • Several national EGN services (hosted by NMCAs)
  • Central EGN service (hosted by BKG Germany)
Background

- Currently EGN service contains geographical names data from 14 countries. 4 countries are ready to connect to the service.

- Italian data also available
- Norway hasn’t joined yet
- Greece is connected with no data
- Estonia hasn’t connected but they have provided data
- The short term target is to expand the EGN service into all 27 EU Member States
Reasons for the service renewal

- The original schema was developed before INSPIRE specification were in place.
- The original output schema was ambitious but most of the countries couldn’t provide data for it.
- New countries have been difficult to convince to join to the EGN service.
- Need to simplify the schema in order to get countries more interested.
- Simpler schema suits also better for web application use.
- Simpler schema makes database easier to understand and manage.
- Old database schema was quite complex and hard to manage.
  - Data update procedures were difficult to handle that ultimately resulted in data corruption in the central service.
Original database schema
Original output schema
Original EGN service - architecture

- Central EGN service and national EGN services
  - Database: PostgreSQL + PostGIS
  - Web Feature Service interface with deegree WFS application 2.x
- Updates are handled periodically through WFS-T requests together with additional database scripting
- The whole dataset of a certain country is updated at the same time by retrieving the data from the national EGN service and replacing fully the corresponding data in the central service
The renewed EGN Service

- The EGN central service is to be renewed and the national EGN services are left untouched
- The renewed service is based on the ISO19112 schema from the OGC Gazetteer Service AP Best Practice document
  - Simpler schema in comparison to the original schema
  - Will be used as the main input and output schema
  - Database schema is derived from this schema
- Data is loaded from the existing National EGN services via a schema transformation process
  - Exonym database will be merged into the central service
  - In future, the INSPIRE Geographical Names services are seen as source for the EGN data
- The EGN central service is established as a Cloud Service-based instance with its own service database
  - The production database will be kept as a separate instance
New output / input Schema

OGC Gazetteer AP

FINNISH GEODETIC INSTITUTE
New Database Schema

- **si_location_instance**
  - PK: location_instance_id
  - geographic_identifier
  - position
  - date_modified
  - spatial_object
  - location_type

- **si_gazetteer**
  - PK: gazetteer_id
  - name
  - scope
  - territory_of_use
  - custodian

- **si_location_type**
  - PK: location_type_id
  - name
  - identification
  - definition
  - territory_of_use

- **gazetteer_location_type**
  - PK: gaz_id
  - parent
  - feature_type

- **alternative_geographic_identifier**
  - PK: alt_geogr_identifier_id
  - loc_id
  - name
  - language

- **location_type_location_type_id**
  - PK: location_type_link_id
  - url_link

**Relationships:**
- location_type = location_type_link_id
- loc_instance_id = location_instance_id

**FKs:**
- PK: location_instance_id
- PK: gazetteer_id
- PK: location_type_id
- PK: alternative_geographic_identifier_id
Architecture of the new service

Production environment

Amazon Cloud Service

[Diagram showing the architecture with WFS FrontEnd, degreeWFS, Production database, National EGN nodes, Exonymdatabase, INSPIRE GN nodes (in future), and Service database.]
Data upload process

• The data was downloaded from the national EGN nodes, transformed into the new schema and uploaded to the production database with a custom Java process.

• The data from the exonym database was loaded separately into the central database because it had completely different structure.
Schema mapping

Original schema

- ComplexGeometry
  - geometry
  - spatialObjectUID
- LocationInstance
  - geographicIdentifier
  - locationType
  - position
  - geographicExtent
  - lastUpdateDate
- LocationInstanceName
  - name
  - language
  - status

New schema

- spatialObject
- SI_LocationInstance
  - geographicIdentifier
  - locationType
  - position
  - geographicExtent
  - dateModified
- alternative
  - GeographicIdentifier
    - name
    - language
    - nameID
    - type
    - primary
- URL
Schema mapping

• Left out
  • complexGeometry
  • classificationScheme

• status (to be indicated by type/primary)

<table>
<thead>
<tr>
<th>status</th>
<th>type</th>
<th>primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>official_endonym</td>
<td>official</td>
<td>TRUE</td>
</tr>
<tr>
<td>other_endonym</td>
<td>official</td>
<td>FALSE</td>
</tr>
<tr>
<td>standardized_exonym</td>
<td>variant</td>
<td>TRUE</td>
</tr>
<tr>
<td>other</td>
<td>variant</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
Future

• The data update procedures are to be specified
• Use of the INSPIRE GN application schema
• Other output formats
  • JSON
  • KML
  • etc...
Thank you!

Questions?