

Executive summary Report User/Business requirements (EGN WP2)

The WP2 task is to compile and describe potential commercial applications, check existing applications and analyse and evaluate these results in order to derive requirements for the EGN service (contents, quality). Contributors to WP2 are ESRI Germany, GeoDan, GeoTask and Utrecht University (coordinator), with input from EDINA.

The WP2 report shows first how names data are collected and how names files need constant updating and maintenance. It then distinguishes between different use functions and different user groups and determines specific user group needs. These are systematised in RISE use cases. The ensuing needs and requirements are put in a market perspective and then followed up by an inventory and analysis of current commercial application of geographical names. As the main objective of EGN is to create a web-based geographical names service, the scope and attributes of current free web-based names servers are analysed. Finally an information model is developed based on the use cases.

Data sourcing

Presently, geonames files are very much based on the paper maps they were taken from, as can also be seen from the fact that in some countries geoname coordinates still refer to the locations of the names on the map instead of the locations of the named objects. In future, this will be reversed and the names on the map or in gazetteers will be a selection of the national names databases or the national topographic databases, with a tendency towards the latter, where the names will be attributes of topographic objects.

Geonames are decided on by (local) authorities. Topographical fieldwork is considered the best means to have geonames information officially collected, both in the field and from local government, as this should be free from bias or commercial considerations. This state monopoly is increasingly under pressure, as mapping activities are being outsourced and there is an increased investment of commercial geospatial information suppliers in fieldwork activities, such as for yellow pages, route planning systems and tourist information systems. In future some form of symbiosis between public and private geodata collecting systems might be envisaged.

The rate of change in geographical names is considerable. It is supposed to be in between 5 and 23% per revision cycle for geonames (and between 5 and 15% for street names); the higher percentages would refer to changes in mapping procedures, changes in orthographic rules or in procedures regarding minority names.

Geonames functions and user groups

The following use groups and functions have been discerned, that should be reflected in the use cases described (see table 1). The geoname enrichment function discerned was seen as up to the VAR.

Kommentar [FO1]: from GeoTask, the health field was added to marketing, the insurance field was added to finance, the pipeline monitoring was seen as represented by distribution, lbs services were combined with emergency services

functions/ use groups	normalisation = look up	translation	indexing	geocoding ~ geoparse	geo- indexing	reverse look up
finance, insurance	xxx	xx				
web sales, tourism	xxx 1	xxx 1	xxx	xx	xx 1	
marketing, health	xxx 9	x 9		xxx 9	xxx	
media	xxx 8	x 8	xx			
distribution	xx	xxx	xx	x	xx	
spatial planning	xx 7	xx		x 7	x	
map data production	xxx	xxx	xx	x	x	++ 13
emergency services, lbs	+++	+	+ 4	+++ 4	+	+++ 12
science	+++ 5	+++ 5		+++ 5	++ 6	++ 6
individual users	+++ 2, 3	+++ 2, 3	+	+++		

Kommentar [FO2]: based on James' paper

Table 1 functions for the use groups discerned, on the basis of needs (xxx), wants (xx), nice to have (x). Numbers refer to the use cases (highlighted), described in the Sharepoint. The following use cases were described: 1) hotel booking use case, 2) EGN name server use case, 3) geoportal map application use case, 4) emergency map use case, 5) geoparsing use case, 6) historical research use case, 7) metadata search use case, 8) geonames checker use case, 9) real estate use case, 12) coordinate emergency use case, 13) query names in bounding box. The last three rows have been added and are not based on the original GeoDan questionnaire.

From the use cases, the following requirements for the EGN names service emerged:

Content: all place names and their variant names in all European languages, and their coordinates.
 -All major lakes and rivers, mountains, physical and administrative areas, all named landscape features. Named areas should have polygons. Previous, replaced names should still be in the database
 -Finance and Insurance and Use case 4 and 9 also require street addresses and major points of interest
 -Use case 8 also requires pronunciation information
Functionality: at place name input show coordinates or bounding boxes (and show adjacent names), show feature types, show variant names, fuzzy name search (also: sounds as:); show name on a map (use case 2), combined query (feature and name or feature and area or feature and coordinates)
 Use case 8 requires a pronunciation facility. Proximity analysis for cases 1 and 6
User interface: ease of use is crucial. Purpose of UI elements must be immediately clear.
 Multilanguage
Integration: available 24/7, protocol agnostic but ADL or OGC preferred. Ability to perform multiple iterative searches
Data quality: names data should be annually updated. Spatial accuracy from 10 (case 4) up to 100m (case 3).

Box 1: EGN geonames service requirements phase 1

Market analysis and inventory of current commercial applications

A study of the geospatial information marketing possibilities identified unexpected market players such as insurance companies and health research. Furthermore, commercial users wanted to use wide-spread software technology, they wanted straightforward incorporation of the gazetteer service in their existing infrastructure and they wanted an easy exchange of information, both through import/export and through open service capabilities.

In the inventory, it was registered whether operators were from the public or private sector, their business model and platform was indicated, service providers and search categories used (place names (9), administrative units (7), addresses (4), postal codes (3), motorways (2), other areas (2)).

Most common search categories in current commercial applications were complete address information and administrative area searches. The accuracy required was point coordinates for address information and bounding boxes for all other name categories. Annual updating of information seemed to be the norm.

Current offer of geonames webservers and comparison with EGN proposal

A survey of current free commercial geonames servers on the web shows 7 providers (Geonames, GONet Names Server (GNS), Getty thesaurus of geographical names, the Fuzzy Gazetteer, the ADL Gazetteer service, Earth Search and World gazetteer. Apart from FuzzyG all of them use the names taken from the GNS, maintained by the US National Geospatial Intelligence Agency (NGA) and the US Board on Geographical Names. The Fuzzy Gazetteer was maintained by the European Commission and the European Joint Research Centre. Attribute data all these servers had in common are names, coordinates, feature codes, country codes and functionality they all shared was the location of the named objects on the map. Half of them worked with unique name or object IDs.

The unique selling points (usp) of the EGN names service for Europe are, that the names data provided are from a primary source that is continuously updated and more detailed than GNS data. It is closer to the experts that collect the names, there is better quality control through official cooperation and it is based on European standards. The current geonames webservers reviewed score as follows in comparison with EGN regarding these usp's: (see table 2)

EGN usp's	EGN	current geoname servers	current route planners
primary data	yes	no	?
official data	yes	?	partly
high quality data**	yes	?	yes
up-to-date data	yes	no	yes
complete coverage	not yet*	no	depends on category
according to European standards	yes	?	yes
including street level data	no	no	yes
coverage	Europe	whole world	Europe +

Table 2, comparison between EGN geonames service, current name servers and route planners. *) complete coverage is envisaged. **) comparable to 1:50 000, coordinates to 10m

Information model

Based on the requirements in box 1, taken from the use cases, 5 main name categories are discerned, and pronunciation is added as a name attribute. Based on the user interface requirements in box 1, language should be an attribute, as well as the date a name was last changed. Street addresses and postal codes came up as a requirement in 2 use cases and in the market analysis and in the commercial applications. In view of the fact that they were beyond the scope of EGN, and also because all of the competing geoname servers did not provide street addresses and postal codes either, it was decided not to require them for the time being. It should be discussed whether to leave room for extending the EGN functionality with street address matching capabilities in future, so that EGN would be able to compete with this address matching possibility in Route planners. In view of the fact that after street address matching the second most frequently used search category was administrative area search, it was decided to include hierarchical objects in the information model.

The most extensive competing geoname servers all had ID's, either for the topographical objects named or for the geographical names. They also had information on the status of the names and the dates this changed. As the official character of EGN is stressed, information on this aspect of the names should be made available. As it was decided not to go into data enrichment and leave that to data-added resellers, the metadata box is rather empty.

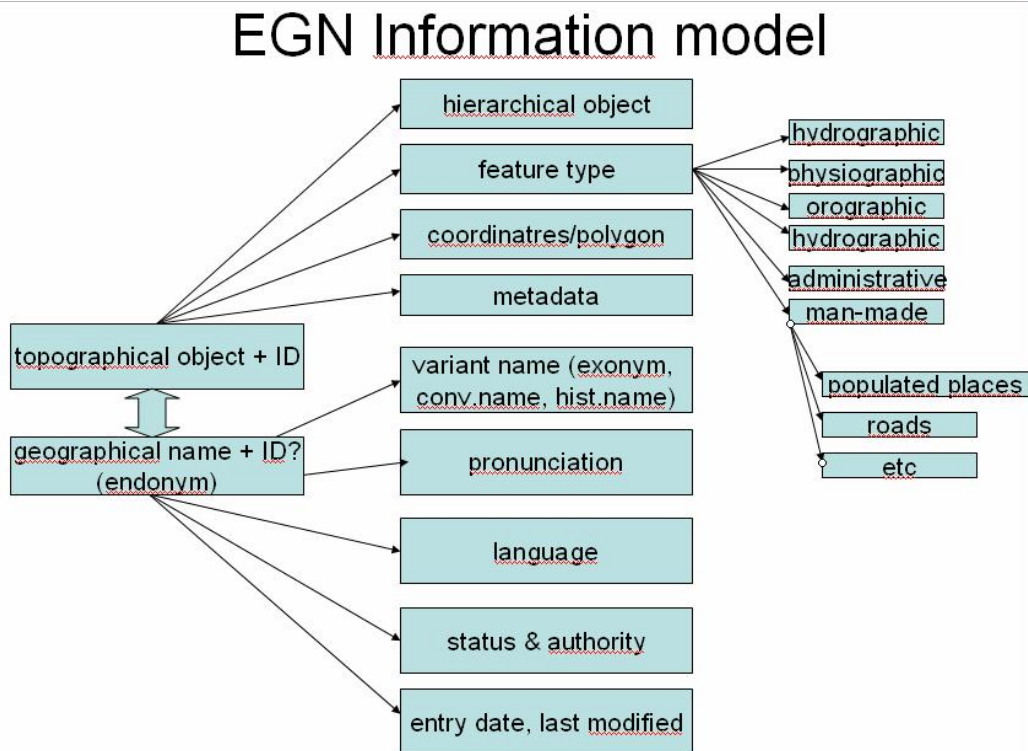


Figure 1 Proposed EGN Information model, based on use cases and market analysis