**ECP 2005 GEO 038026 EGN**

**EGN**

# Analysis of the “state of the art”: Issues for implementing the EGN infrastructure

| | |
|---|---|
| **Deliverable number** | *D11.1* |
| **Dissemination level** | *Confidential* |
| **Delivery date** | *31 March 2008* |
| **Status** | *1.2* |
| **Author(s)** | *BKG , Geodan, ESRI* |

***e*Content*plus***

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

# Analysis of the “State of the Art”: issues for implementing the EGN infrastructure

## Table of Contents

# ~~1~~ 1 Change summary

| Version Revison | Date | Author | Notes |
|---|---|---|---|
| 0.1 | 25/06/07 | FK, AB | 1st draft |
| 0.2 | 24/07/07 | FK, PGZ, RM, JSpr | |
| 0.3 | 27/07/07 | JSR, JSpr, PGZ, RM, AvD | |
| 0.4 | 02/08/07 | JSR, JSpr, PGZ, RM, JS | |
| 0.5 | 16/08/07 | FK, AB, PGZ, RM | |
| 0.6 | 17/08/07 | FK, AB, PGZ, JSpr, RM, JS | |
| 0.7 | 11/09/07 | GB, JSR, JS, JSpr | Comments of 5th CM (Vienna) added |
| 0.8 | 12/09/07 | JS, JSpr | Chapter on architecture added |
| 0.9 | 13/09/07 | JS, RM, SA, FK | Comments of FK re V0.8 considered |
| 0.9.1 | 13/09/07 | JS, BS | List of tasks beyond project lifetime, chap 13.2 |
| 1.0 | 14/0707 | JS, FJO, RM | Final version |
| 1.1 | 30/1107 | GBO | Chapters 8.1 (user scenario) and 8.2 (server load) added |
| 1.2 | 2008-03-19 | FK | Added more issues to chapter 12 |

# ~~2~~ 2 Introduction

## ~~2.1~~ 2.1 Project aim

The eContentplus EuroGeoNames (EGN) Project is currently developing a European Infrastructure for the exploitation and management of geographical names. In popular terminology, the project is building up a federation of distributed digital gazetteers maintained by the respective official national authorities, hence referred to as the National Mapping and Cadastral Agencies (NMCAs).

## ~~2.2~~ 2.2 Purpose of this document

This document is designed to address Task 11.1 as defined in the original Description of Work. Its primary purpose is to provide an overview and summary of technical progress in developing the EGN Infrastructure and, in a discursive fashion, to sketch out issues and options available to the Consortia for the purposes of progressing the overall objectives of EGN. This document therefore represents an interim attempt of a gap analysis combined with some aspects of a SWOT analysis. Its principal purpose is to be an informational focal point around which alternate views and issues can be expressed and articulated.

The work described in this document is a work in progress and therefore is only a snapshot of the issues at this time - it will change as issues arise and get addressed. In short it is a 'living' document that will evolve as the project does.

It is envisaged that a European gazetteer infrastructure will support a variety of different geospatial application domains a range of formal Use Cases that have been identified as part of the project. A more comprehensive background on EGN and other deliverables are available from the EGN website at:

http://www.eurogeonames.com

## ~~2.3~~ 2.3 Summary

This document consists of a list of diverse issues that need to be resolved in the process of developing the EGN infrastructure. The issues are placed in no particular order. Some issues are related to other issues. When this is the case, this is indicated by a remark at the end of each issue.

The document is intended to be used as a common focal point for discussions and decision-making for all partners.

# ~~3~~ 3 Glossary

## ~~3.1~~ 3.1 Abbreviations

Abbreviations used in this document are listed and explained below, in alphabetical order.

**BKG:** The German federal agency for cartography and geodesy

**DoW**: Description of Work (Annex I ECP_2005_GEO_038026 EGN to the Grant Agreement signed by EC and BKG on 19 October 2006 ). A description of the goals and the workplan for the EGN project

**EGN:** EuroGeoNames

**ETL**: Extract, Transform and Load. A common process in IT

**INSPIRE:** Infrastructure for Spatial Information in Europe

**ISO:** International Standards Organization

**LoD:** Level of Detail

**NMCA**: National Mapping and Cadastral Agency

**OGC:** The Open Geospatial Consortium

**SWOT:** Strengths Weaknesses Opportunities and Threats

**WFS:** Web Feature Server

**WFS-G:** A Web Feature Server used as gazetteer

## ~~3.2~~ 3.2 EGN infrastructure components

In diagram 1 the basic components of the EGN infrastructure are depicted. The components are briefly explained below.

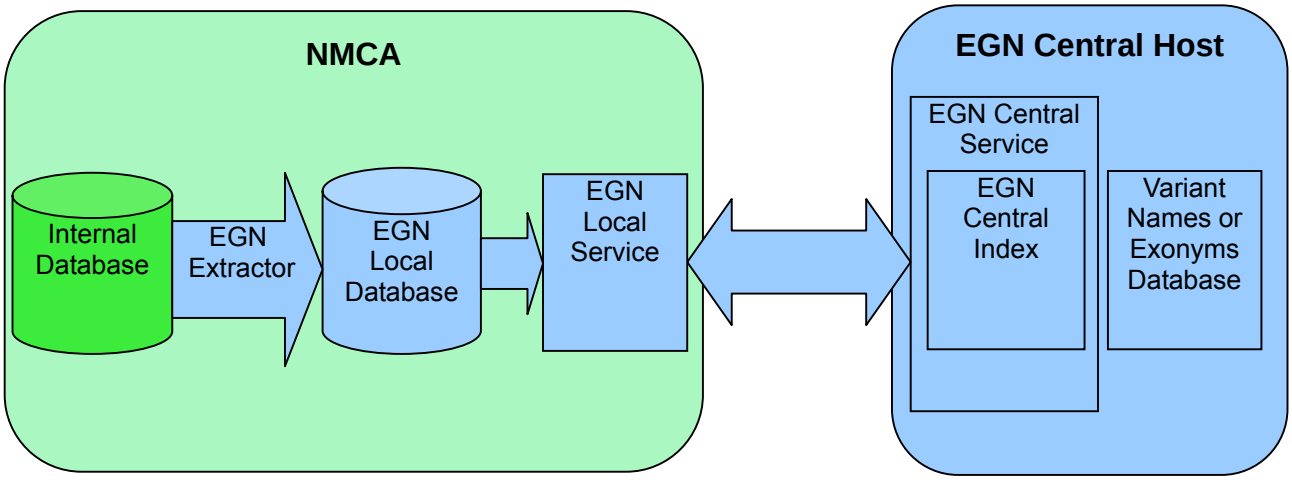



*Diagram 1: EGN infrastructure components*

**Internal Database:** This is the database of the NMCA containing the geographical data that will be used for the EuroGeoNames service. It is entirely in the authority of the NMCA, the EGN Consortia will however request reading access for extracting and converting the geographical names information.

**EGN Extractor:** Reads data from the Internal Database, converts it to the EGN data model and writes it to the EGN Local Database.

**EGN Local Database:** Contains data needed for the EGN Local Service, using the EGN data model.

**EGN Local Service:** The EGN Local Service is a Web Service (probably a WFS) accepting queries for geographical name information issued by the EGN Central Service.

**EGN Central Service:** The EGN Central Service is the core component of the EGN infrastructure. It will accept requests for name searches from the clients, it will perform the search in the EGN Variant Names or Exonyms Database and in the EGN Central Index, it will forward the search to the appropriate EGN Local Services and send a response to the client.

**EGN Central Index:** This index will help the EGN Central Service in the search for the data necessary to create a response to a request.

**Variant Names or Exonyms Database:** This is a database containing variant names or exonyms which are, in general, not contained in NMCAs' names databases containing the official national names forms (endonyms). It is created and maintained by the EGN consortium with help from language communities.

Remark: The **Variant Names or Exonyms Database** itself will not be part of the **EGN Central Index**, only an extract of it will be transferred to and used in the **EGN Central Index**.

# ~~4~~ 4 EGN infrastructure architecture

There is an ongoing discussion in the Consortium on the concept of the architecture of the EGN infrastructure. The architecture proposal described here, therefore, is a snapshot of an advanced discussion as of 20th August 2007. Currently the following prerequisites have to be observed:

- The EGN infrastructure to be developed shall be based on the INSPIRE principle to access decentralized data kept and maintained in the responsibility of the data provider (NMCA);

- Indexes - containing the geographical name (endonym, exonym or variant names), point location or bounding box, EGN feature unique ID, feature class - in the central RDBMS are necessary to directly assign a geographical name inquiry to the respective NMCA WFS (EGN Local Service);

- A larger number of NMCAs does not allow to let their full data being cached in a central database of the EGN Central Service as a secondary data set;

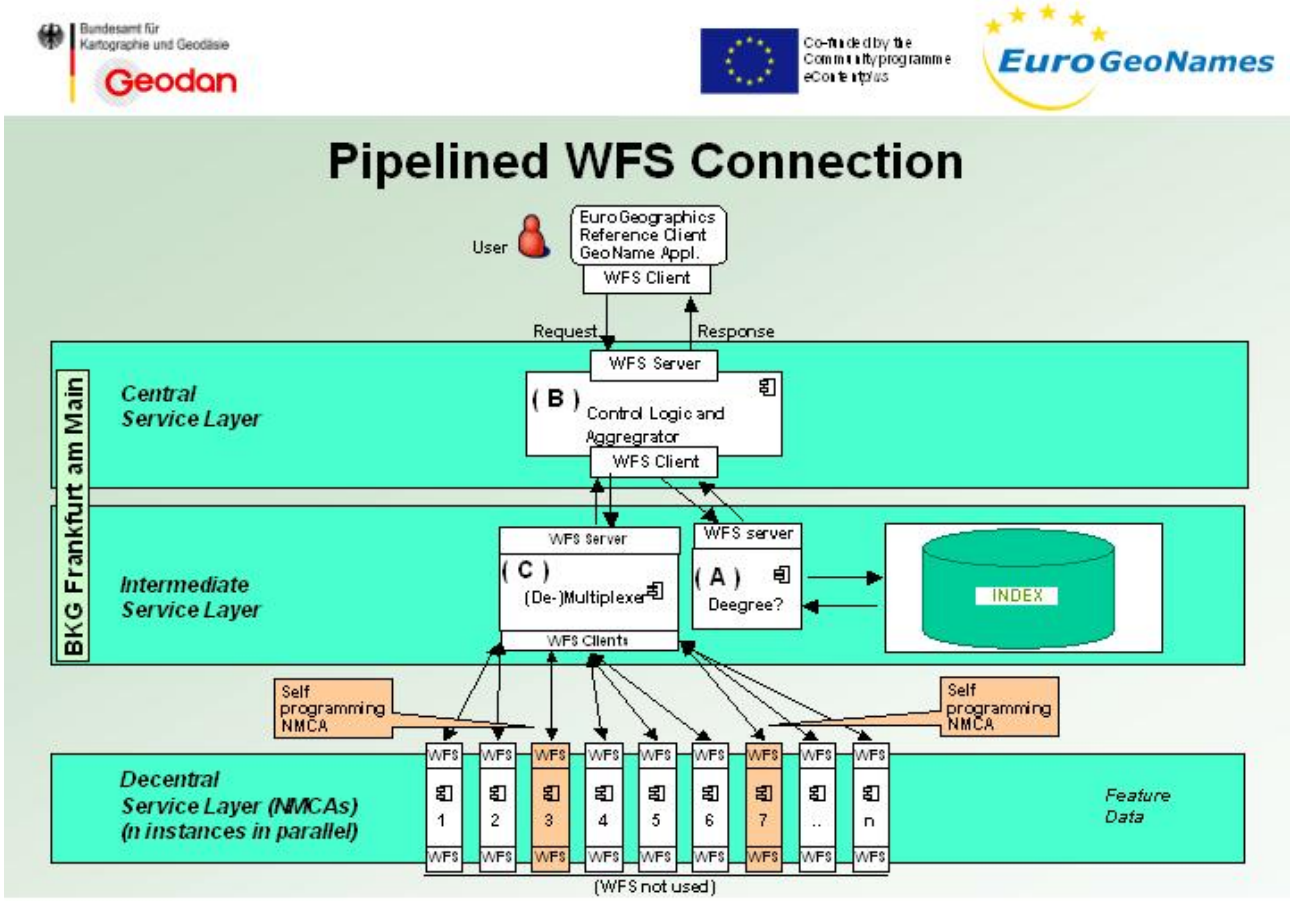
- The host of the EGN Central Service (BKG) is not willing to take over responsibility for keeping and maintaining any NMCA's data in the Central Service.

## ~~4.1~~ 4.1 *Architecture scheme – Status of discussion: 20-08-2007*

The EGN infrastructure shall fulfil the following conditions:

1. Use existing powerful WFS and RDBMS tools (A)
   Use RDBMS update and indexing functionality for fast lookups and individual NMCA index updates.
2. Extend the software by adding two components:
   - Add a preprocessing tool (as cascading WFS) that handles WFS requests from clients (B). This preprocessing tool uses the standard WFS tool (A) for index / exonym lookups, simple box queries and simple results;
   - Add (de-)multiplexer tool (as cascading WFS) that queries NMCA nodes for information or complex spatial queries (C)

### ~~4.1.1~~ 4.1.1 Diagram of the architecture schema




*EuroGeoNames IT Architecture Issues, Frankfurt am Main, August 20th, 2007*

**(A)** Existing powerful WFS and RDBMS tools;
**(B)** Preprocessing tool (as cascading WFS) to handle WFS client requests;
**(C)** Multiplexer tool (as cascading WFS) to query NMCA nodes;
**Index** contains geographical names (endonyms, exonyms or name variants), point locations or bounding boxes, EGN IDs, feature classes (see "EGN Central Index" chapter 2.2);
**1 .. n** NMCA names data conform to EGN data model to serve WFS requests.

*(As mentioned, this architecture proposal is a snapshot stating the discussion as of 20th August. Meanwhile Geodan has come up with an alternative architecture version on 13th September that has to be discussed in detail. The following text refers to the 20th August proposal only)*

### ~~4.1.2~~ 4.1.2 Sequence of a client query

1. Reference client queries use the index in the RDBMS at the intermediate layer (via WFS from (B) to (A)).
2. If a result from the index requires more information from the NMCA, the NMCA node is queried. This will be done via Software (B), which will not be able to use a WFS interface

between (B) and (C). The information which NMCA WFS is to be queried by the multiplexer is not part of the WFS specification.

3. The NMCA has its own RDBMS. This (may be) is used to store all data including geometry data according to the EGN data model. This RDBMS can do complex spatial queries as a standard feature.
4. The NMCA does not need cascading, aggregation or multiplexing and can therefore just use the standard WFS software (A) configured for EGN.
5. After all results are aggregated by software (C), the response is WFS encoded and returned to the client.

### ~~4.1.3~~ 4.1.3 NMCA updates of EGN Central Index

1. The indexes in the central RDBMS are updated using a separate software (D) (not shown in the diagram 3.1.1) that connects to the NMCA WFS servers.
2. The software (D) can be triggered by the NMCA administrator. To keep data and index aligned, NMCAs must send updates every time the NMCA’s EGN data is updated.
3. On request there can be a monitoring tool that monitors the availability of NMCA nodes. If a node is not available, the corresponding index entries in the intermediate layer RDBMS are disabled.

## ~~4.2~~ 4.2 *Final remarks to the EGN infrastructure architecture*

1. The suggested EGN infrastructure seems to make use of only one WFS interface definition, as agreed upon in May 2007.
2. A standard WFS software (A) can be used at the NMCA.
3. The same standard WFS software (A) will work on the EGN Central Index which by definition will follow the same data model as the NMCA's WFS.
4. For security reasons it makes sense to update the indexes in the central RDBMS using a separate software (D) that connects to the NMCA's WFS servers.
5. WFS software (B) by definition will follow the same data model as the NMCA's WFS.
6. This solution makes use of four software modules (A, B, C, D) instead of one. Two of these software modules (B, C) contain both WFS server and WFS client interfaces. At least software modules (B), (C) and (D) must be maintained by the consortium or its successors.
7. There is no possibility for further cascading at the NMCA side.
8. Advantages to the previous suggestion are not clear. There are now two different cascading WFS softwares (B), (C) instead of one.
9. Separation of software modules (A), (B) and (C) avoids the combination of general control logic and multiplexer control logic within one software.
10. Between software modules (C) and (B) no further WFS is possible.

11. Software module (B) works in phases:
    – Phase 1: Ask index
    – Phase 2: Receive index information
    – Phase 3: Build request for multiplexer
    – Phase 4: Ask multiplexer
    – Phase 5: Receive aggregated answer from multiplexer or aggregate answers from multiplexer
    – Phase 6: Give response back to user.

# ~~5~~ 5 EGN Extractor

For those NMCAs that are not going to develop and maintain the EGN Local Service themselves the EGN Local Database needs to be filled with data derived from local sources. This task is commonly known as ETL (Extract, Transform and Load). Source data is expected to be very diversely structured and make use of diverse storage systems (e.g. ESRI shapefiles, RDBMS, MS Excel, ..). Creating mappings between local data structures and the EGN data model and implementing an automated procedure that can execute data conversion based on these mappings is expected to take a lot of time.

## ~~5.1~~ 5.1 Risks

- There is a risk that we need to spend a lot of time on each individual case of data conversion.
- Data mapping software might be invalidated by changes made to local data structures.
- This task was not foreseen in the Description of Work. It was decided to share responsibility between work packages 4 and 6 and therefore between Geodan and EDINA.

## ~~5.2~~ 5.2 Solutions

**1)** Make use of standard ETL software. First evaluate free / Open Source solutions. This will hopefully provide us with a standardized way of making data model mappings and will save us from having to write a different program for each NMCA.

Steps:

- Ask NMCAs and EGN partners for advise. Maybe someone has experience with free ETL software.
- Compile a list of ETL software candidates and make a selection of software to evaluate. Selection criteria are:
  1. Price
  2. Reviews, experiences from other users
  3. Activity of developers
  4. Data mapping easy, clear, separate and portable
  5. Extensibility (it may be necessary to add our own transformation functions)
  6. A wide support for data source formats
  7. Support for PostgreSQL/PostGIS
- Test ETL software candidates with real NMCA data.

To counter the risk of EGN partners having to do extra work if an NMCA changes its data structure, NMCAs should be able to change the mapping for their data themselves (criterion 4).

**2)** Do not wait with development of the EGN Infrastructure, but try to start with NMCAs that will be able to provide EGN Local Services according to the interface specifications.

# ~~6~~ 6 Translations

The EGN infrastructure aims at being multi-lingual, which means that end-users should be able to read the attribute information related to the geographical name in a preferred language to be chosen from a list (an explicit list of all supported languages is provided in EGN Deliverable D5.1, chapter 3.3.3, table 2). This means that the standard terms used in the EGN services need to be translated in all supported languages. This applies to the following items (between brackets an estimate of the number of terms):

- status (2, "official" and "other")
- language (15, Czech, Dutch, English, Finnish, French, German, Greek, Hungarian, Latvian, Lithuanian, Slovak, Slovene, Spanish and Swedish which are the EU acknowledged official languages of the EGN Reference Group countries as well as Turkish)
- country (15 Reference Group member countries)
- feature classification (not yet decided)

The proposed way of doing the translations is that the EGN consortium will make the lists of terms in English. Participating NMCAs will then provide EGN with translations for their languages. As the list of terms is not very long, this will not be a big task for a single NMCA. The obligation of NMCAs to provide translations should be added to the official agreement (memorandum of understanding) between EGN and NMCA.

# ~~7~~ 7 Security

It is common understanding within the IT community in general and EuroGeoNames in specific that security is a critical issue. The EuroGeoNames Web Service, first, represents a service, and second, delivers goods, namely geonames information. Both the service and the goods are of great value and will not be available for free without limitations. As a result, access to both needs to be regulated. Unauthorised access needs to be prevented.

The EuroGeoNames service together with the geonames data reside on distributed IT infrastructures. These infrastructures need to be protected from intrusion and damage. Ultimately, the effort of integrating security in the architecture and development of the EuroGeoNames Web Service from the beginning is much smaller than the effort of dealing with damages during Web Service operation.

Within the context of EuroGeoNames, three aspects of security are distinguished:

- protection against intrusion of hackers and malware (viruses, trojan horses, etc.);
- protection of messages sent through the Internet;
- management of user access to the EuroGeoNames service.

While the last point is dealt with in the subsequent section, the former two are discussed in this section.

Appendix A gives an overview of security threats and countermeasures. The EuroGeoNames consortium will need to distinguish between the threats to be negated and the threats to be prepared for, considering the probability of the occurrence, the potential damage, and the costs of applying countermeasures. Appropriate countermeasures will have to be applied for the critical threats.

**Security questionnaire.** It is important to incorporate the existing strategies, tools, and expertise of the members of the reference group into the development of the EuroGeoNames service. Furthermore, it has been deemed reasonable to build from existing solutions that are frequently used by the members of the reference group and that have proven effective.

For these reasons, a questionnaire has been sent out to the members of the reference group and to a number of affiliated organisations asking for the security measures they apply. 15 forms have been sent back. 13 forms indicate that Web Services of some kind are in operation. Firewalls were mentioned as protective measure 13 times. Also, demilitarised zones (DMZs) and https are commonly included in the security strategy. Nine organisations mention Apache as a platform for Web Servers and Web Services. Security gateway and user authentication was indicated six times, while Web Authentication and Authorisation Services (WAAS) was taken into consideration by one organisation only.

**Protection against intrusion and malware.** The EuroGeoNames architecture distinguishes between the EuroGeoNames Central Service and the EuroGeoNames Local Service. While the central service is presumed to be installed and operated at the premises of the BKG, the local services are operated on the infrastructures of the connected NMCAs. There is agreement in the consortium and the reference group that the NMCAs are responsible to secure their instances of the EuroGeoNames service.

The architecture of the EuroGeoNames Local Services as presented by Geodan foresees a small number of components to be installed. This architecture offers to hold these components within a Demilitarised Zone (DMZ) and, in this way, separating them from all non-EuroGeoNames

components. Each NMCA decides for itself whether to follow such a strategy or to implement other security concepts. In any case, the connection of the local service to the Internet will be secured by means of a standard firewall.

For the EuroGeoNames Central Service, the security regulations of the hosting organisation will apply.

**Securing information transfer through the Internet.** In order to secure the Internet connection from the user to the EuroGeoNames Central Service and from the central to the local services, https will be used where critical information such as user names and passwords are transmitted. Some security can be provided for the communication between central and local services because the IP addresses of both are known and do not change without advance announcement.

**Miscellaneous.** For the discussions, the assumption was made that the EuroGeoNames service is only accessed through the central service. The local services will only respond to requests from the EuroGeoNames Central Service. Final decision on this issue is pending at the time of this writing.

Anyway, the EGN Central Service at BKG must be protected against malware. This means to have a firewall not only in direction of the users but also in direction of the NMCAs.

# ~~8~~ 8 Access control and user administration

Control of access to the EGN Central Service (authentication and authorization) can be required for two reasons:

a. A countermeasure against security threats

b. Providing different levels of access to different types of users (for example, paying versus non-paying customers)

A full-featured system for user administration and access control is not needed in the first three years of the project, but it might be a good idea to start with a simple system at the EGN Central Service that can distinguish between two three roles:

1. **Public** may request a certain fixed amount of features per day.
2. **Test** with extended rights in the context of prototyping.
3. **Superusers** have no usage limits.

If we have a simple but working and extendible access control system, we will have both a proof of concept and a smooth transition to a more complex access control and user administration system at the EGN Central Service.

The EuroGeoNames Consortium decided at its 5th Consortium meeting held in Vienna that the implementation of a simple but extendable access control system for the EGN Central Service is a fundamental functional requirement for the success of the project.

## ~~8.1~~ 8.1 *User scenario*

The securityManager module of the sdi.suite is the instrument with which access rights to the central EGN-Service are organised. Access will be restricted to authorised users, thus securing the recognition of user agreements and preventing unauthorised use.

Users and rights are not defined statically. The administration and management of users and rights can be performed in the securityManager by way of an intuitive user interface in a standard web browser. Thus users and rights can be added, edited or deleted whenever it is necessary in the EGN project.

From current view the following users are intended:

1) PUBLIC:

The user PUBLIC is a proxy for all anonymous users in the sense of "Single Inquiry" (e.g. Users of the EGN ArcGIS-Extension). The user PUBLIC can be uniquely administered with the user interface of the securityManager.

The securityGateway will provide a pre-configured access point with unrestricted lifetime (permanent or so-called persistent gateway) for public use, which does not require the user to log in.

In order to grant restricted access in the sense of "Single Inquiry" a new Interceptor (EGN-SI-Interceptor) will be realized within the EGN project period. The EGN-SI-Interceptor identifies anonymous users over the internet and restricts the number of inquiries to 50 per user and day.

2) TEST:

A test entrance can be made available e.g. to a potential VAR in order to test the behavior of the EGN service with extended rights in the context of prototyping. The administration of a test account presupposes always compellingly the conclusion of a contract.

To prevent abusive use for each test account an own user (username/password) has to be configured. The administration of concrete users (e.g. VAR1) and the individual rights assignment (e.g. spatial restriction, temporal restriction, feature-type restriction) take place thereby with the user interface of the securityManager.

For the monitoring of the use intensity in the context of test entrances a new Interceptor (EGN-UI-Interceptor) will be realized within the EGN project period. The EGN-UI-Interceptor counts the requests for each test user and stores the information (user; number of inquiries) in one log file for later evaluations.

3) SUPERUSER

An entrance as a superuser can be administrated for each entitled one (e.g. EGN-Consortium Partner, NMCA of the Reference Group).

Superusers receive unrestricted access to all functions and data of the EGN service. If necessary, individual rights can be assigned (e.g. spatial restriction to grant individiual access to the NMCAs).

To prevent abusive use for each superuser (e.g. BKG) an own user (username/password) has to be configurated. The administration and management of users can be performed with the user interface of the securityManager.

## ~~8.2~~ 8.2 Delimitation of the server load by user public

A large number of anonymous users (e.g. through strong demand for the EGN ArcGIS Extension) will produce high hit rates. As a function of the efficiency of the EGN central service this might be a possible performance risk and in particular reduce the serviceability for contractually bound partners.

If the components of the EGN web service infrastructure are not capable to handle high hit rates (e.g. divisibility of the components on dedicated servers, load balancing mechanisms) a possible fallback strategy would be to limit the contemporarily accessibility of the EGN central service to a certain amount of anonymous users.

This can be achieved for example by configuration of dedicated points of entrance for anonymous and contractually bound users. The number of simultaneous anonymous users can be limited to a reasonable number with functionalities of the assigned web server. If these points of entrance are to be made available over the components of the securityManager, it is sufficient to duplicate the components Web Authentication Service and security Gateway into dedicated web server environments.

## ~~8.3~~ 8.3 Access control within the EC-funded project period:

A simple access control system has to support identification of users via authentication (username / password) and authorisation to grant restricted access (by area, by request, single inquiries as defined in EGN-project) to individual users of the EGN Central Service.

The EGN access control system will be built upon existing software components. The sdi.suite securityManager employs the Web Authentication Service (WAS) and Web Security Service (WSS) developed by the OpenSource Software Initiative 52°North. WAS and WSS are based on an outcome of the spatial data infrastructure initiative of North Rhine-Westphalia (GDI NRW). They are tried and tested and implemented in various projects like INSPIRE@EC (Eurostat), LoG-IN (Interreg III Project) etc. They make use of web services standards such as “Security Assertion Markup Language” (SAML) and “Extensible Access Control Markup Language” (XACML). The management of users and user rights by the securityManager can be performed easily using a web browser. The securityManager license will be provided by con terra for free.

The EGN-specific adaptations of the used components (conceptual work, implementation, integration, documentation) have been estimated as an effort of 2 person-months. The adaptations do not contain efforts concerning the client development. There are at three options for developers of client applications to integrate access control into their clients:

- via access control Java API (provided with sdi.suite securityManager)

- via access control http Service API (provided with sdi.suite securityManager)

- via web security Gateway (provided with sdi.suite securityManager; Option without integration efforts)

## ~~8.4~~ 8.4 Access control beyond the EC-funded project period:

The access control system may be extended to fulfil the needs of the (not yet finalized) business model (implementation of new authorization features, e.g. specific filter rules), to ensure that contracts / user agreements of EGN are adhered to at all times. The further maintenance (e.g. ongoing management of users and user rights) will be a task of the service provider.

# ~~9~~ 9 Conformance with gazetteer standards

It is the aim of EGN to comply as much as possible to any applicable standards. Because of the nature of EGN, any standards for gazetteers warrant close inspection. At the time of writing, the only officially published standard for gazetteers is ISO-19112. Based on this standard, a gazetteer profile for WFS was proposed by the OGC (OGC 05-035r2). Also based on ISO-19112, the INSPIRE Data Specifications drafting team has made a draft proposal for a Gazetteer application schema that is not in total conformance with ISO-19112, as it suggests corrections to ISO-19112. This proposal, which is intended to become mandatory within INSPIRE, can be found in the draft for INSPIRE deliverable D2.5: Generic Conceptual Model, page 47, Figure 16.

## ~~9.1~~ 9.1 *ISO-19112*

ISO-19112 deals only with spatial referencing by geographic identifiers. This type of spatial reference is sometimes called "“indirect”. ISO-19112 defines the conceptual schema for spatial references based on geographic identifiers. Spatial referencing by coordinates is addressed in ISO-19111. However, a mechanism for recording complementary coordinate references is included in ISO-19112.

The standard ISO-19112

- establishes a general model for spatial referencing using geographic identifiers,

- defines the components of a spatial reference system and

- defines the essential components of a gazetteer.

### ~~9.1.1~~ 9.1.1 Definition of terms (ISO)

The following terms are used/defined in ISO-19112:

**feature**
abstraction of real world phenomena.
NOTE A feature may occur as a type or an instance. Feature type or instance shall be used when only one is meant. [ISO 19101: 2002]

**gazetteer**
directory of instances of a class or classes of features containing some information regarding position.
NOTE The positional information need not be coordinates, but could be descriptive.

**geographic identifier**
spatial reference in the form of a label or code that identifies a location.
EXAMPLE “Spain” is an example of a country name; “SW1P 3AD” is an example of a postcode.

**location**
identifiable geographic place.
EXAMPLE “Eiffel Tower”, “Madrid”, “California”

**spatial reference**
description of position in the real world.

### ~~9.1.2~~ 9.1.2 Unique identification

ISO-19112 states that a geographic identifier should be unique, which means it should uniquely identify a feature (paragraphs 6.1 and 7.1). In practice, this is not possible as the INSPIRE Data Specifications drafting team has stated, too. Other information that is needed for clear identification is the relation with other features and the feature class. For example, the name 'Rijswijk' can be used to identify four different features in the Netherlands:

| Feature name | Feature class | Province |
|---|---|---|
| Rijswijk | municipality | Zuid-Holland |
| Rijswijk | town | Zuid-Holland |
| Rijswijk | town | Gelderland |
| Rijswijk | town | Noord-Brabant |

In this case, to uniquely identify the feature we need to know the containing feature (province) and the feature class.

### ~~9.1.3~~ 9.1.3 Version numbering

For conformance of a gazetteer with ISO-19112, a version number of the gazetteer (data set) may be needed. On page 9 it says “*A new version of the gazetteer shall be created whenever any location instance is created or destroyed, or a new version of a location instance is created.*” It would not be much trouble to include a version number and it might be useful. Contrary to a spatial reference system, for which a similar demand is made (page 7), there is no obligation to publish the version number of a gazetteer.

In the EGN metadata profile (see D3.3) there is a version number, which is the version number of the data set. The data set is not entirely the same as the gazetteer, but in practice an update of the data set will result in a new gazetteer instance, so with regard to version numbering the two can be seen as equals. Because there is a link between gazetteers and the metadata profile (property metaDataURL of a Gazetteer object), there seems to be little reason to store separate version numbers for Gazetteers.

### ~~9.1.4~~ 9.1.4 Superfluous attributes

There are many mandatory attributes for a location type (feature class), see table 3 on page 8:

- name (character string)
- theme (character string)
- identification (character string)
- definition (character string)

It is hard to reconcile this with EGN. In the first place, free text fields in general are problematic in any multi-lingual system. Secondly, there does not seem to be a need for storing and publishing all of these attributes.

### ~~9.1.5~~ 9.1.5 Conformance tests

Annex A of ISO-19112 lists conformance tests. Conformance with ISO-19112 can be achieved on two levels: for a spatial reference system or for a gazetteer. Two tests can check whether a gazetteer is ISO-19112 compliant:

1. Check that the properties of the gazetteer are known
2. Check that all location instance attribute data are correct.

This test boils down to check whether all mandatory attributes are available.

## ~~9.2~~ 9.2 *OGC*

The OGC has published a proposal for a WFS gazetteer profile. The proposal is based on ISO-19112.

### ~~9.2.1~~ 9.2.1 IsGlobal

It is not clear what the purpose of SI_Gazetteer.IsGlobal is.

### ~~9.2.2~~ 9.2.2 Revision Working Group

The OGC Technical Committee recently announced the formation of a new Revision Working Group (RWG) for WFS Gazetteer Application Profile. At least one of the people involved with EGN should participate in this group, so that conflicts can be discussed and compliance with the definitive application profile can be achieved.

## ~~9.3~~ 9.3 *INSPIRE*

The proposal for a Gazetteer application schema in INSPIRE was published on the 5th of May, 2007, as part of the draft of D2.5: Generic Conceptual Model.

### ~~9.3.1~~ 9.3.1 Definition of terms (INSPIRE)

The following terms are used in INSPIRE D2.5_V2.0.
Numbers in **(..)** refer to the numbering in that document:

**(3) class**

description of a set of **objects** that share the same attributes, operations, methods, relationships, and semantics [ISO/IEC 19501]

**(16) external object identifier**

a **unique object identifier** which is published by the responsible body, which may be used by third parties to reference the spatial object

**(17) feature**

abstraction of real world phenomena [ISO 19101]

NOTE The term "(geographic) feature" as used in the ISO 19100 series of International Standards and in this document is synonymous with **spatial object** as used in this document.

Unfortunately "spatial object" is also used in the ISO 19100 series of International Standards, however with a different meaning: a spatial object in the ISO 19100 series is a spatial geometry or topology.

**(18) feature catalogue**

catalogue(s) containing definitions and descriptions of the **spatial object types**, their attributes and associated components occurring in one or more **spatial data sets**, together with any operations that may be applied [ISO 19110 – modified]

**(20) gazetteer**

directory of instances of a class or classes of features containing some information regarding position [ISO 19112]

NOTE A gazetteer can be considered as a geographical index or dictionary.

**(22) geographic identifier**

spatial reference in the form of a label or code that identifies a location [ISO 19112]

EXAMPLE 1 Place names: Paris, Rhine, Mont Blanc

EXAMPLE 2 Postal codes: 53115, 01009, SW1, IV19 1PZ

**(34) object**

*in this document used synonymous with* ***spatial object***

**(35) object identifier**

**unique object identifier** associated with a **spatial object**

**(42) spatial data**

data with a direct or indirect reference to a specific location or geographic area [INSPIRE Directive]

NOTE The use of the word "spatial" in INSPIRE is unfortunate as in the everyday language its meaning goes beyond the meaning of "geographic" – which is considered by the Drafting Team as the intended scope – and includes subjects such as medical images, molecules, or other planets to name a few. However, since the term is used as a synonym for geographic in the draft Directive, this document uses the term "spatial data" as a synonym for the term "geographic information" used by the ISO 19100 series of International Standards.

**(44) spatial object**

abstract representation of a real-world phenomenon related to a specific location or geographical area [INSPIRE Directive]

NOTE It should be noted that the term has a different meaning in the ISO 19100 series. It is also synonymous with " (geographic) feature" as used in the ISO 19100 series.

**(52) unique object identifier**

piece of data, usually in the form of printable characters, that unequivocally identifies a **spatial object**

### ~~9.3.2~~ 9.3.2 Feature Catalogue

The EGN conceptual schema allows for the use of different feature classifications. The EGN feature classification scheme will be satisfactory for the purposes to which it is required (essentially, query filtering). The EGN project uses the term ‘classification’ throughout to distinguish from the term ‘catalogue’ used in other contexts (such as ISO 19110) which implies a degree of formalism not required for the purposes of EGN. The purpose of the EGN classification is primarily to assist in query and discovery.

At time of writing several options have been explored and presented to the EGN Reference Group. However, given the difficulties alluded to above, it has not been possible to agree on which alternative feature classification is most expedient. Following a meeting of the Reference Group in New York on 23rd August 2007, the Reference Group members proposed another classification which the Consortia are now evaluating. At this juncture it is likely that this alternative scheme is likely to become the preferred and implemented scheme due mainly to its general acceptance by the Reference Group members (who will ultimately need to be happy with mapping their own national coding schemes into the EGN scheme).

### ~~9.3.3~~ 9.3.3 Features and feature names

A LocationInstance is considered to be a spatial object i.e. an abstract representation of a real world phenomenon related to a specific location or geographic area. In the proposed application scheme a geographic feature (LocationInstance) can only have one name (geographicIdentifier). In the EGN line of thinking, a feature can have many names which is in accordance with the amendments suggested by the INSPIRE drafting team within D2.5.

The attribute “alternativeGeographicIdentifier” is used within EGN for storing the “egnFeatureUID” which uniquely identifies the feature. It must be unique not only within its own gazetteer but across all the gazetteers.
According to the INSPIRE requirements, the egnFeatureUID will be composed of a namespace to identify the data source plus a local identifier, assigned by the data provider (this must be unique within the namespace). All namespaces will start with the two letter ISO 3166 code (e.g. DE, NL). “DE.BKG.GN” may be the namespace used by BKG Germany for spatial objects in their geographical names database. The local identifier is generated according to the BASE36-system (e.g. “2XH50000A”).
The result will be: “DE.BKG.GN.2XH50000A”

This proposal will be tested against actual data examples by Geodan.

### ~~9.3.4~~ 9.3.4 Features and geometries

In the EGN line of thinking, a feature can have many geometries, which is in accordance with the amendments suggested by the INSPIRE drafting team within D2.5. GM_Object makes it possible to store different geometries for one location. The concept of having multiple GM_Objects related to different scales is not part of the INSPIRE gazetteer application schema, but can be modelled as an extension of that schema.
In the EGN data model, the classes „EGNLocationInstance.EGNSpatialLocation->geometry“ refer to GM_Object as defined by ISO 19107. A GM_Object contains the GM_Point, GM_Envelope and

GM_Boundary classes for holding point coordinates, bounding box coordinates and polygonal boundaries. It also allows for the holding of multiple geometrical representations of an object.

This will be tested against actual data examples by Geodan.

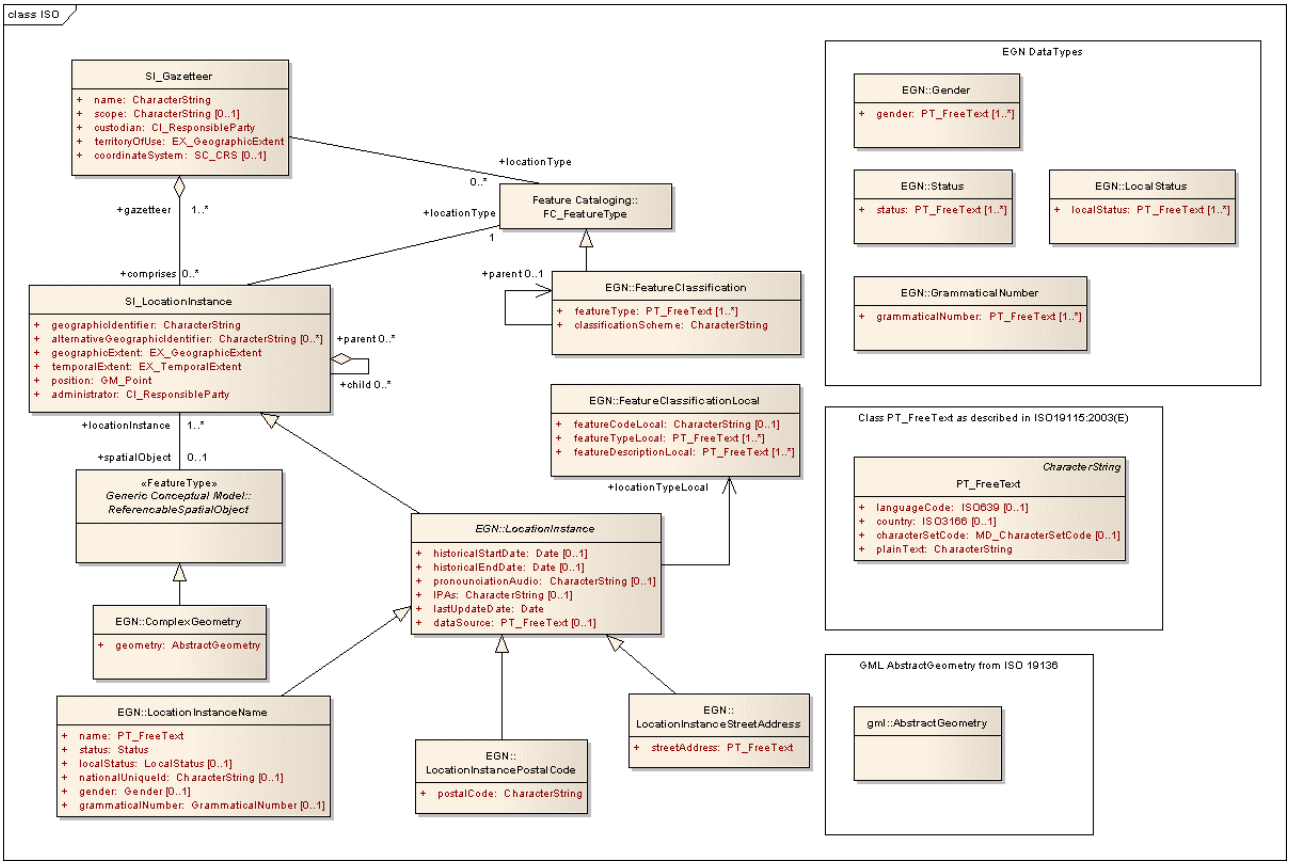
## ~~9.4~~ 9.4 ISO, OGC and INSPIRE

There are some differences between the EGN data model and elements all three (ISO, OGC and INSPIRE) specifications have in common. For the time being there seems to be no conflict between the EGN conceptual model and the INSPIRE generic conceptual model.

This will be tested against actual data examples by Geodan.

# ~~10~~ 10 Data model

## ~~10.1~~ 10.1 The UML representation of the EGN Conceptual model




An earlier version of the scheme dated 31st May 2007 and publicly published as EGN Deliverable D4.2 noted at that time that: “the Conceptual schema should be considered informative rather than definitive at time of writing as subsequent developments within INSPIRE and during implementation may necessitate model revision”. The UML data model rendered here reflects the latest and *final* (as of January 2008) revision of the Conceptual. We have also taken the opportunity to streamline the presentation of the model and more fully expanded on the explanations and attribute definitions.

The proposed data model will be tested against actual data examples by Geodan.

## ~~10.2~~ 10.2 10.2 Relationship to D4.3 and D6.2

The data model presented in D4.2 has been used to derive the official EGN Application Profile which is a collection of XML schemas documents intended to provide XML validators with the

template against which conformance to the official EGN data model can be gauged.

Due to software and implementation constraints the Application Profile presented in D4.3 has been modified slightly in the Interface Specification (D6.2). System implementers should be aware that this discontinuity exists and D4.3 viewed as the gold standard reference with D6.2 providing a pragmatic near approximation.

***In general, those interested in system implementation or in integrating with the EGN Infrastructure, should consult D6.2.***

In the longer term, with maturation of the EGN Infrastructure and supporting software, it is expected that the Interface Specification will fully support the Application Profile as presented in D4.3.

In summary:

- Document D4.3 contains the XML schema documents representing the official EGN Application Profile as generated from the UML model presented in D4.2.
- Document D6.2 presents the Interface Specification for the EGN services (generated by WFS).

In the interim until full support for D4.3 is readily implementable by existing software, D6.2 should be regarded as canonical for purposes of inter-operation with existing EGN Infrastructure environment.

# ~~11~~ 11 Software selection

This is an overview of the software that will be used for the EGN infrastructure and services. For all software components, first we try if there is free software available that will do what is required. If such software is not available, low cost alternatives are considered first.

## ~~11.1~~ 11.1 Operating System

There are two candidates: Linux and Windows. Linux can be nominally free, but it is not unimaginable that the real costs are higher than with Windows because more system administration time is needed. With Linux another problem is having to pick a particular distribution. MS Windows is used by the majority of NMCAs (13 use Windows, 2 use Linux).

Therefore, we are considering to use Windows as being the operating system to start with. Nevertheless, Linux will be considered in the test environment.

## ~~11.2~~ 11.2 Application server

We are considering to use Apache Tomcat 5.5 because it is required by the Deegree WFS.

## ~~11.3~~ 11.3 RDBMS

We have selected PostgreSQL 8.2 with PostGIS 1.2.1. Other open source databases with the ability to store spatial data exist, but the PostgreSQL/PostGIS combination has the most features. Also, it is a widely used RDBMS in geospatial open source communities.

## ~~11.4~~ 11.4 WFS

We have selected Deegree WFS 2.1. Another free and open source WFS is Geoserver, but that software does not support WFS 1.1.

## ~~11.5~~ 11.5 ETL

This software package still has to be selected (see EGN Extractor).

# ~~12~~ 12 WFS implementation

## ~~12.1~~ 12.1 Support of system requirements

In EGN D6.1 appendix A the system requirements for the EGN services are listed. No guarantee of support for those requirements was given. Based on current knowledge, the table of system requirements has been expanded in Appendix D of this document. This extended overview of system requirements shows the intended level of support for each requirement.

## ~~12.2~~ 12.2 Nested features

EGN features (named objects) are complex features: one EGN feature can have several WFS child features. Examples of these child features are the geometry (spatial location) and the translation.

Based on experience with the WFS prototype it does not seem possible to request only one child feature if there are many. Based on the GetFeature query a set of features is returned that comprises all child features. To illustrate this, a feature will have all geometries for all levels of detail (LoD) and all translations of the country name as children.

Negative effects of this WFS behaviour is that unneeded data may be transferred between server and client and that client software may need to do extra processing to filter out the desired information.

### ~~12.2.1~~ 12.2.1 Solutions

First we need to establish how big this problem is. If this behaviour of the EGN services does not cause severe problems at the client-side, nothing will be done. If this behaviour is considered problematic, Geodan has to investigate the possibilities of adapting the WFS in such a way that is will be possible to make selections in nested features.

## ~~12.3~~ 12.3 Unclear relationships between feature types

The GML application schema that is produced by the EGN Local Service prototype (see appendix B) by itself is not clear in the relationships between feature types. For example, the attribute *coordinateSystem* of the feature type *SI_Gazetteer* is specified as follows:

```
<xsd:element name="coordinateSystem" type="gml:FeaturePropertyType" minOccurs="0"/>
```

The specification *type="gml:FeaturePropertyType"* shows that this is a complex property: the property is another feature type. But it is not clear that this feature type is *egn:SC_CoordinateReferenceSystem*. This information can be derived from the UML class diagram, but the output from a DescribeFeatureType itself itself is insufficient.

The EGN Local Service prototype uses a Deegree 2.1 WFS. This issue could be regarded as a weak point of this particular WFS Server. However, in theory it is possible to use separate schemas for DescribeFeatureType and GetFeature. We could decide to make use of this possibility to let DescribeFeatureType return a response in which the relationships between feature types are clearer.

## ~~12.4~~ 12.4 Fuzzy and 'sounds like' searching

We want to support fuzzy and 'sounds like' searching for geonames. Fuzzy searching means that the spelling does not need to be correct. For example, a search for 'Parris' should still find return information for Paris if fuzzy searching is enabled. 'Sounds like' searching uses a special database index (a soundex index) to match search string with names that sound like the search string. The soundex algorithm is language-dependent, although a simple algorithm could suffice for most languages used in Europe.

These two methods of searching are not covered by the OGC WFS or Filter specifications. A way of supporting fuzzy and 'sounds like' searching within the limits of the specifications and current WFS software has not been fully researched yet. One possible way of doing this could be to define extra attributes of LocationInstanceName (e.g. *fuzzyName* and *soundexName*) derived from the *name* attribute.

If good methods of enabling fuzzy and 'sounds like' searching are found, the extra interface specifications will only be required for the EGN Central Service.

## ~~12.5~~ 12.5 XLinks

The way of using XLinks described in paragraph 2.2.3 has not been tested in practice yet, it is not certain that this method will work in practice. The first part of the problem is to configure the WFS server and its database in such a way that these types of XLinks can be produces. The matter is complicated by the following factors:

1. XLinks could replace objects with different data types. For example, elements of type 'number' or 'geometry' could both be replaced by an XLink, which is of type 'string' itself.

2. Whether or not an element should be replaced by an XLink depends on the policy of the data provider. This means that if there are some similar objects in one GML document, some objects might have XLinks in their properties while others have not.

## ~~12.6~~ 12.6 Pagination

Pagination is needed to make 'single inquiries' possible. It is also nice to have something like pagination to prevent GetFeature responses from becoming very big. However, pagination is no out-of-the-box functionality, it is a subject that is not touched upon by the WFS specification. The EGN Consortium therefore has to come up with its own way of supporting pagination. This has support for pagination has not been fully developed yet. The way in which this will affect the interface specification is not clear yet, but it will probably consist of an extra so called 'vendor specific' request parameter. This parameter could be called 'PAGENUMBER', for example. The following (GET) request would return the third page of the response:

```
http://egn.geodan.nl/deegree-
wfs/services?SERVICE=WFS&VERSION=1.1.0&REQUEST=GetFeature&TYPENAME=egn:LocationInstanceName&NAMESPAC
E=xmlns(egn=http://www.eurogeonames.eu/egn)&PAGENUMBER=3
```

Because the 'single inquiry' concept only affects the EGN Central Service and because do not want to limit NMCA's in their choice of WFS server software, the requirement to handle this extra parameter will only be made for the EGN Central Service.

# ~~13~~ 13 Plans for further development

## ~~13.1~~ 13.1 Task list for further development of the EGN web service during the EC-funded project lifetime

| Step | Description | Finish date | Participants | Remarks |
|---|---|---|---|---|
| 1 | Prototype EGN Local Service | | Geodan, EDINA, BKG | Once the prototype has been accepted, work can start on the reference client, the ArcGIS extension and the test plan. |
| 2 | Evaluation of ETL software | | Geodan, EDINA | |
| 3 | Identify security risks that need to be addressed and propose countermeasures | | GeoTask | |
| 4 | Installation of first working EGN Local Service (including data) | | EDINA, Geodan | Which NMCA? If this is a NMCA that makes its own EGN Local Service, this step is not dependant on ETL software. |
| 5 | Installation of second working EGN Local Service (including data) | | EDINA, Geodan | Which NMCA? If this is a NMCA that makes its own EGN Local Service, this step is not dependant on ETL software. |
| 6 | Design of EGN Central Service | | Geodan | Can start after there are two functional EGN Local Services. |
| 7 | Installation of EGN Central Index and EGN Central Service | | Geodan, BKG | To be hosted by BKG |
| 8 | Installation of Variant Names or Exonyms Database | | BKG | To be hosted by BKG |
| | | | | |

## ~~13.2~~ 13.2 Task list (preliminary) of the future EuroGeoNames institution or consortium

a) Host the EuroGeoNames Web Service, or sub-contract this hosting

b) Continuing support for NMCAs (in hosting EuroGeoNames Local Services)

c) Technical support for VARs (in integrating the EuroGeoNames Web Service in their applications or services)

d) Administrative support for VARs (making contracts, managing accounts for EuroGeoNames access, billing)

e) Continuous development of the EuroGeoNames Web Service (bug fixing, development of new functionality requested from the users/VARs, adopting to new technological developments)

f) An appropriate form of organizing these tasks has to be drafted before the end of the EC-funded project period.

# Appendix A: Security threats and countermeasures

The following table is considered as a snapshot of the issues at this time - it will change as issues arise and get addressed. In short it is a 'living' document that will evolve as the project does.

## *Table 1: Security threats*

| no. | threat | approximate cost if threat becomes reality | counter-measure in advance | approximate cost of counter-measure<br>- once<br>- permanent | side effects of counter measure | which resources are at risk, when counter-measure fails? | time & effort to restore the system | remaining risks & possible unwanted side effects | remarks |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Unauthorized access to **and use of** the system(s) through back-doors (enabled through trojan horses, viruses or unpatched security holes) | | | | | | | | unauthorized use: see also numbers 3 and 4. |
| 1.1 | attacking other systems inside or outside the organization | none, if an appropriate disclaimer is placed unknown | Monitoring site | - once<br>100 Euro (installation)<br>- permanent 200 Euro p.m. (staff) | none | attacked system(s) | allowed recovery time to be defined for every EGN site | legal prosecution, if disclaimer “does not work”. | not EGN specific; remaining problem: how easily can the attack be repeated? |
| 1.2 | setup a phishing site to collect (identity and financial) information from unaware people | none, if an appropriate disclaimer is placed | Monitoring site | (same as above, included) | none | any | allowed recovery time to be defined for every EGN site | legal prosecution, if disclaimer “does not work”. | not EGN specific; remaining problem: how easily can the attack be repeated? |

| no. | threat | approximate cost if threat becomes reality | counter-measure in advance | approximate cost of counter-measure - once - permanent | side effects of counter measure | which resources are at risk, when counter-measure fails? | time & effort to restore the system | remaining risks & possible unwanted side effects | remarks |
|---|---|---|---|---|---|---|---|---|---|
| 1.3 | distributing (illegal) content using disk space and network capacity of the host | none, if an appropriate disclaimer is placed | Monitoring site | (same as above, included) | none | attacked system(s) attacked network(s) | allowed recovery time to be defined for every EGN site | legal prosecution, if disclaimer "does not work". | not EGN specific; remaining problem: how easily can the attack be repeated? |
| 1.4 | altering system configuration to "irritate" the owners of the host. | depending on contracts and penalties for not being online | Monitoring site | (same as above, included) | none | attacked system(s) attacked network(s) | allowed recovery time to be defined for every EGN site | effort to restore system to proper state | not EGN specific; remaining problem: how easily can the attack be repeated? |
| 1.5 | Shutting down one EGN site | depending on contracts and penalties for not being online | "Hardening" System | - once: 5.000 for first site, 1.000 for the next ones - permanent: Zero | none | attacked system(s) attacked network(s) | recovery time may be long, allowed recovery time to be defined for every EGN site | effort to restore system to proper state | not EGN specific; remaining problem: how easily can the attack be repeated? |
| 1.6 | Shutting down more than one EGN site | depending on contracts and penalties for not being online | "Hardening" System | - once: 5.000 for first site, 1.000 for the next ones - permanent: Zero | none | attacked system(s) attacked network(s) | recovery time may be long, allowed recovery time to be defined for every EGN site | effort to restore system to proper state | not EGN specific; remaining problem: how easily can the attack be repeated? |
| 2 | Overloading | | | | | | | | |

| no. | threat | approximate cost if threat becomes reality | counter-measure in advance | approximate cost of counter-measure - once - permanent | side effects of counter measure | which resources are at risk, when counter-measure fails? | time & effort to restore the system | remaining risks & possible unwanted side effects | remarks |
|---|---|---|---|---|---|---|---|---|---|
| 2.1 | Overload through "normal" use (non-malicious) | depending on contracts and penalties for not being online | a) System design b) Fine-grained Access Control c) Monitoring EGN service(s) | To be investigated for cases a) b) and c) | some overhead, to be investigated | EGN may not be able to fulfil contracts. Consequences: penalties, litigation. | none; except for restart to clear buffers (we should look for a better solution than a restart!) | cost & effort to install bigger system | |
| 2.2 | Overload on purpose (DOS attack) | depending on contracts and penalties for not being online | Monitoring EGN service(s) | To be investigated | some overhead, to be investigated | EGN may not be able to fulfil contracts. Consequences: penalties, litigation. | none; except for restart to clear buffers (we should look for a better solution than a restart!) | None | |
| 3 | using the system for free or at a lower rate than they are supposed to | none, as long system is not severely slower than contracted | a) System design b) Fine-grained Access Control | To be investigated for cases a) and b) | some overhead, to be investigated | financial turnover goes down; economical risks | none, system runs on. Only effort for hardening the system (see above) | | |
| 4 | theft of data | | | | | | | | - |
| 4.1 | directly (source data copy) | similar to previous point 3. Cost should be specified by NMCAs | Hardening system: Shielding Data Base Server | to be investigated | to be investigated | financial turnover goes down; economical risks | to be investigated | | remaining problem: how easily can the attack be repeated? |
| 4.2 | indirectly (data mining) | similar to previous point 3. Cost should be specified by NMCAs | Access Control by volume (?) | to be investigated | to be investigated | financial turnover goes down; economical risks | to be investigated | | remaining problem: how easily can the attack be repeated? |
| 5 | Falsification of data | | | | | | | | |

| no. | threat | approximate cost if threat becomes reality | counter-measure in advance | approximate cost of counter-measure - once - permanent | side effects of counter measure | which resources are at risk, when counter-measure fails? | time & effort to restore the system | remaining risks & possible unwanted side effects | remarks |
|---|---|---|---|---|---|---|---|---|---|
| 5.1 | names data in data base | depending on contracts and penalties | restricted Write Access | nearly zero cost | none | credibility with customers. EGN may not be able to fulfil contracts. Consequences: penalties, litigation. | low, cost of restore run. | | remaining problem: how easily can the attack be repeated? |
| 5.2 | index data in data base | depending on contracts and penalties | restricted Write Access | nearly zero cost | none | credibility with customers. EGN may not be able to fulfil contracts. Consequences: penalties, litigation. | low, just cost of restore run.. | | problem: index update procedure |
| 5.3 | variant names or exonym data in data base | depending on contracts and penalties | restricted Write Access | nearly zero cost | none | credibility with customers. EGN may not be able to fulfil contracts. Consequences: penalties, litigation. | low, just cost of restore run. | | problem: variant names and exonym update procedure |
| 5.4 | during transmission | depending on contracts and penalties | HTTPS | to be investigated | transmission about 15% slower (1st estimation, to be validated) | credibility with customers. EGN may not be able to fulfil contracts. Consequen-ces: penalties, litigation. | may be high, because method of hack must be analysed | | |

## *Table 2: Countermeasures*

| number | countermeasure | against threats | responsibility | remarks |
|---|---|---|---|---|
| 1 | Put EGN computers in a DMZ | | | |
| 2 | Use HTTPS between WFS and clients | | | |
| 3 | Install anti-virus software and keep it up to date | | | |
| 4 | Automatic intrusion detection | | | |
| 5 | | | | |

# ~~14~~ 14 Appendix C: Overview of name status modes used in EGN countries

| Country | Status-Field in Countries | Status | | | Example |
|---|---|---|---|---|---|
| | | | *Status* | | |
| | | *id´s* | *german* | *english* | |
| Germany | STATUS_ID | 1 | amtlich | official | Alzenau i. Ufr. |
| | | 2 | ausgeschrieben | full | Alzenau in Unterfranken |
| | | 3 | historisch | historical | |
| | | 4 | gebräuchlich | in use | Alzenau |
| | | 5 | offen | open | |
| | | | | | |
| | | | *Language_Official_Status* | | |
| | | *id´s* | *finnish* | *english* | |
| Finland | Language_Official_Satus_Code | 1 | | Official language in this munipality | |
| | | 2 | | Unofficial language in this municipality | |
| | | *id´s* | *Language_Majority_Status* | | |
| | Language_Majority_Status_Code | 1 | | Majority language in this municipality | |
| | | 2 | | Minority language in this municipality | |
| | | | | | |
| | | | *Status* | | |
| Spain | | *id´s* | *spanish* | *english* | |
| | official | 1 | oficial | official | |
| | | 2 | normalizado | standardized | |

| | | | | | |
|---|---|---|---|---|---|
| | | 3 | no normalizado | no standardized | |
| | | 4 | no disponible | no available | |
| | | *id´s* | *Classname* | | |
| | | 1 | preferente | primary | Ría de Mundaka |
| | | 2 | alternativo | alternative | Mundakako itsasadarra |
| | | 3 | variante | variant | Ría de Gernika |
| | | 4 | histórico | historic | |
| | kind of toponyms (Type_classname) | 5 | anterior | previous | |
| | | 6 | sobrenombre | nickname | |
| | | 7 | exónimo | exonym | |
| | | 8 | no disponible | no available | |
| | | | | | |
| | | | *Status* | | |
| Norway | | *chars* | *norwegian* | *english* | |
| | SNSKRSTAT | G | Godkjent etter gammel ordning | Accepted according to former provisions | Halvarssteinen |
| | | F | Foreslått | Proposed | Halvfarssteinen |
| | | V | Vedtatt etter lov om stadnamn (vedtak på enkeltnavn) | Authorized by the provisions of the Place name Act (single names) | |
| | | S | Vedtatt etter lov om stadnamn (samlevedtak) | Authorized by the provisions of the Place name Act (frequent names) | |
| | | K | Vedtak påklaget | Decision appealed | |
| | | A | Avslått | Rejected | |

In Norway we are not allowed by law to differ between majority or minority placenames in a municipality. This has to be evaluated in the presentation om maps

| | | | | | |
|---|---|---|---|---|---|
| | | U | Uvurdert | Not treated | (this code will be replaced by the code "Proposed " as attribute) |
| | | P | Privat navn | Private place name | (names on hotels, turisthuts, villas,etc) |
| | | I | Internasjonalt område | International area | (all names outside Norway) |
| | | H | Historisk navn | Historical place name | Halvfarestensrøset |
| France | NO STATUS AVAILABLE | | | | |
| Czech Republic | NAZEV_001 (in XLS and DBF tables; field is an Attribute of an dataset)<br>NAZEV_TISK_001 (in XLS and DBF tables; field is an Attribute of an dataset)<br>NESTAND_001 (in XLS and DBF tables; field is an Attribute of an dataset) | name as string | *Status (in XLS and DBF tables)*<br>*czech* | *english*<br>standardized name<br>form of the standardized name shown on the map<br>variant names without standardization | |
| | NAME (in SHP-Files, field is an Attribute of an dataset)<br>Text (in SHP-Files; field is an Attribute of an dataset) | name as string | *Status (in SHP-Files)* | standardized name<br>form of the standardized name shown on the map | |
| Hungary | NO STATUS AVAILABLE; variant names are available by dataset | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Austria | NO STATUS AVAILABLE | | | | |
| Slovac Republic | NO STATUS AVAILABLE; historical names are available by dataset | | | | |
| Netherlands | Netherlands holds a status field by the Geographical feature not by the name | | | | |
| Slovenia | ID_STATUSA | *id´s* | *Status* *slovak* | *english* | |
| | | 1 | | Official geographical name | |
| | | 2 | | Official minority geographical name | |
| | | 3 | | Generally established geographical name | |
| | | 4 | | Unofficial minority geographical name | |
| | | 5 | | Traditional geographical name | |
| Lithunaia | Only official forms are used for cartography | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Turkey | NAM_STD_TYP | | *Status* | | |
| | | *chars (1-2)* | *turkish* | *english* | |
| | | Y | Yaygin Ad | Conventional | |
| | | C | Cografi Adlar Uzmanlar Kurulu (COGAD) | BGN Standard | |
| | | CR | COGAD (Roman-alfabesi disinda kodlanmis) | BGN Standard in non-Roman script | |
| | | D | Dogrulanmamis | Not verified | |
| | | DR | Dogrulanmamis (Roman-alfabesi disinda kodlanmis) | Not verified in non-Roman script | |
| | | T | Tarihi | Historic | |
| | | TR | Tarihi (Roman-alfabesi disinda kodlanmis) | Historic in non-Roman script | |
| | | O | Öneri halinde | Provisional | |
| | | OR | Öneri halinde (Roman-alfabesi disinda kodlanmis) | Provisional in non-Roman script | |
| | | A | Farkli kullanim ya da alternatif | Variant or alternate | |
| | | AD | Farkli kullanim ya da alternatif (Roman-alfabesi disinda kodlanmis) | Variant or alternate in non-Roman script | |

| | | | | | |
|---|---|---|---|---|---|
| Cyprus | All geographical names appearing on our maps and in our databases are the official names. | | For this reason, all geographical names that will be involved in the EGN database will be the official ones. | | |
| Latvia | galvenais | *boolean* | Status<br>*Is this the prefered name?*<br>*lavian* | *english* | |
| | | YES | | Preferred name | Rubini |
| | | NO | | Alternate name | Rubeni, Rubyni |
| | oficials | *boolean* | *Is this an official name?* | | |
| | | YES | | Official | Rubeni |
| | | NO | | Not official | Rubini, Rubyni |
| | v_nosaukumi | name as string | Type of the name (each type is referred to Name ID) inter alia: previous, historical, dialectal form, erroneous form, Lithuanian, Byelorussian, Russian, Estonian, upper reaches, lower reaches (for rivers) etc. | | latg. (dialectal form in Latgale) Rubyni |
| | | | | | kļūdaini (erroneous form) Rubini |

# Appendix D: Support for system requirements

The system requirements from D6.1 are listed here. The following columns have been added:

**Optimum:** The best possible support for this requirement. If it is little extra trouble, we should reach this level of support.

**Average:** The support considered feasible.

**Unacceptable:** This level of support is unacceptable.

## *From the perspective of the end user*

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 1 | The service should be easily accessible | Yes, based on HTTP standard | Yes, based on HTTP standard and a servlet engine | Own development |
| 2 | The service should always be accessible. | 365 (d) per year; 24(h)* 7 (d) per week, = 100% availability | 346,75 (d) per year, means 18,25 (d) or 438 (h) of unavailability<br>= 95,0 % availability | Less than 95% availability |
| 3 | The service should provide a fast response. | < 1 second<br><br>- max. 0.3 seconds until beginning of result transmission from a local NMCA to the central EGN service;<br>- plus max. 0.2 seconds (for each resulting ID/data string) until result transmission from a local NMCA to the EGN Central Service;<br>- max. transmission delay through central service: 0.2 seconds;<br>- Typical response time of EGN Central Service itself (without need to ask EGN Local Services): 0.3 seconds; | < 5 seconds | > 1 minute |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 4 | The service should provide understandable answers, in the language of the user. | (Requirement for the EGN Central Service or for the Client Software?)<br>Yes, as for the EGN Central Service in any languages based on defined ISO and OGC standards<br><br>Yes, as for the Client software in any languages based on defined ISO and OGC standards | (Requirement for the EGN Central Service or for the Client Software?)<br>Yes, as for the EGN Central Service in 14 languages (English, 13 Original) based on defined ISO and OGC standards.<br><br>Yes, as for the Client software (EGN reference Application) in 14 languages (English, 13 Original) based on defined ISO and OGC standards. | Own development |
| 5 | It should be possible to use historical names in a query. | Yes | Yes<br>(but, historical names are not mandatoryly maintained in the EGN infrastructure) | No |
| 6 | It should be possible to only query for features of a certain type (feature class, e.g. 'rivers' or 'provinces'). | Yes | Yes | No |
| 7 | The service should be able to provide historical names in the query result. | See #5 | See #5 | See #5 |
| 8 | The service should be able to provide variant names in the query result. | See #5 | See #5 | See #5 |
| 9 | The service should be able to return the centroid or a single label point of each found feature. | Yes | Yes | No |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 10 | The service should be able to return a geometry for each feature found (for example, a line string, polygon or multi polygon). | Yes | No, in the first phase of EGN the minimum geometry which will be returned for each feature will be<br>- a point with x and y coordinates<br>- a bounding geometry/box.<br><br>The provision of the full geometry depends on a solid business model and has to be agreed with the NMCAs. | No |
| 11 | The service should be able to return a MBR (minimal bounding rectangle or bounding box) for each feature found. | Yes, it is calculated dynamically at runtime | Yes, predefined bounding boxes will be calculated and stored either in the NMCA database only or additionally in the EGN Central Index (database).<br>The CRS used by EGN will be: ETRS89. | No MBR |
| 12 | It should be possible to use the results from a query to make a subsequent query (iterative searches). | Yes | Yes | No |
| 13 | The service should provide information on the pronunciation of geographical names. | Yes | Yes<br>(if the pronunciation of the name is maintained somewhere in the EGN infrastructure) | No |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 14 | It should be possible to search for names that do not exactly match the name entered by the user (fuzzy searching). | Yes, with elaborated algorithms. | No, a fuzzy search needs a logical algorithm, which cannot be developed within the EU-funded project duration (30 months).<br><br>(However, it will be considered and further evaluated either in the EGN Central Service software and/or in the EGN Reference Application software) | No |
| 15 | It should be possible to search for names that sound like the name the user has entered (soundex searching). | Yes, with elaborated algorithms. | Yes, it will be considered in the EGN Central Service software and/or in the EGN Reference Application software. | No |
| 16 | It should be possible to use wildcard characters (for example "*" or "?") in the search term. | Yes | Yes, but the response of the EGN infrastructure will limit somehow the number of features returned for one query | No |
| 17 | It should be possible to limit a search to features within a user-defined rectangle. | See #16 | See #16 | See #16 |
| 18 | It should be possible to search for features that have a certain spatial relationship (like 'overlaps', 'inside' or 'near') to a specified other feature. | Yes, many/complex spatial relationships calculated dynamically directly in the NMCA database. | Yes, simple spatial relationships will be facilitated in the first phase of EGN by using the bounding geometries/boxes provided by the NMCAs. | No spatial relationship search possible |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 19 | The service should provide information on the grammatical number of geographical names. | Yes | Yes<br>(if the grammatical number of the name is maintained somewhere in the EGN infrastructure) | No |
| 20 | The service should provide information on the gender of geographical names. | Yes | Yes<br>(if the gender of the name is maintained somewhere in the EGN infrastructure) | No |

### *From the perspective of the Value Added Resellers (VAR)*

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 1 | The service should be compatible with standard GIS components. | Yes, based on defined ISO and OGC standards | Yes, based on defined ISO and OGC standards | Own development |
| 2 | The service should be of a standard service type. | Yes, based on ISO, OGC and INSPIRE standards / specifications | Yes, based on ISO, OGC and INSPIRE standards / specifications | Own development or based on one standard only |
| 3 | The service should use standard interfaces for interoperability and communication. | Yes | Yes | No, own development |
| 4 | Non exclusive providing | Access for everyone, including providers | Access for everyone, including providers | Access for special groups or providers only |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 5 | The service should be well documented. | Yes, full documentation of interfaces and functionality based on ISO, OGC and INSPIRE standards / specifications | Yes, full documentation of interfaces and functionality based on ISO, OGC and INSPIRE standards / specifications | No documentation of interfaces and functionality |
| 6 | The service should have a high availability (99,5% uptime, 24 hours a day, every day of the year) | 365 (d) per year; 24(h)* 7 (d) per week, = 100% availability | 346,75 (d) per year, means 18,25 (d) or 438 (h) of unavailability<br>= 95,0 % availability | Less than 95% availability |
| 7 | The service should be fast (2,5 seconds per request). | < 1 second<br>(the same figures as in #3 of user perspective minus approx. 1 second for VAR service) | < 5 seconds | > 1 minute |
| 8 | The service should be monitored continuously to detect any problems. | Yes. monitoring data under continuous control | Yes, limited number of parameters monitored continuously (is service up); additional checking for load and average process time | No monitoring |
| 9 | Technical support should be available and respond within an hour after the report of a problem. | Yes | Yes, within 6 hours (office time 8 to 17 hrs) | Yes, after 72 hours |
| 10 | EGN has to respond quickly to changes in market (e.g. prices, user-requirements). | Yes | Yes, but parameters have to be stated more precisely; e.g. price changes or agreed change requests will have a fixed date for implementation.<br>A solid business and pricing model will be established and agreed with the NMCAs within the EU-funded project duration which will be valid for the time beyond. | No |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 11 | EGN has to provide a fine-grained user-management to distinguish the free services from the non-free services. | Yes, full access control and user management (to be defined more precisely) | Yes, simple access control with no explicit user management (guest role only)<br>A simple but working and extendible access control system will be established. | No access control = No user management. |
| 12 | Non-free services must have a clear benefit for the user (improved data quality or data quantity). | Yes<br>(This depends also on client functionality. EGN will be able to set up more powerful servers, if the services are paid by VARs) | Yes, but within the project duration only free services are planned. | No clear benefit for VARs (to be defined more precisely) |
| 13 | The EGN service should cover off-line- or in-house usage. | Yes<br>(This depends on a specific contract with local data providers (NMCAs), VARs are enabled to set up their own EGN services in-house. The EGN architecture has to be designed accordingly.) | Not within project duration. | Not at all. |
| 14 | The EGN service should have a well-known service interface. | Yes, based on ISO, OGC and INSPIRE standards / specifications | Yes, based on ISO, OGC and INSPIRE standards / specifications | No, own development |
| 15 | Access to the EGN service should be provided as soon as possible, so that clients can be tested. | Yes, permanent access for VARs to a test suite / test site | Yes, a test environment will be setup until 15.01.2008.<br>(Which clients are to be tested and how secure access is possible will be defined.) | No possibility for tests. |

### *From the perspective of the data provider*

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 1 | The service should not create any additional security threats to internal data and systems. | Yes<br>(but, this is nearly impossible to realize, as it would mean a 100,0 % risk free EGN infrastructure) | Yes, a security concept will be prepared identifying the threats, risks and countermeasures. For each countermeasure the party responsible will be assigned. | No, security issues are out of scope of the project. |
| 2 | The service should be usable for other purposes than the EGN gazetteer (for example: a custom gazetteer service). | Yes, but these other purposes will have to be defined precisely and contracted. | Yes, the service should be capable to provide all data according to INSPIRE specifications to the respective clients. However, these other purposes will have to be defined precisely and have to be contracted. The planning shows that there will be no additional resources within the project duration. | No |
| 3 | Already existing NMCA systems, services and procedures should be used as much as possible. | Yes | Yes | No |
| 4 | The service should keep the 'richness' of source data intact, as much as possible. | Yes | Yes<br>(EGN defines a common overall data model for the EGN infrastructure. This data model defines the maximum possible output. Whether all features and attributes are provided depends what the NMCAs maintain in their databases) | No consideration of the richness of source data. |
| 5 | The EGN software and data must be easy to maintain. | Yes<br><br>(It must be kept in mind, that special requests | Yes, the EGN Central Service as well as the EGN Local Service software will be easy to maintain. | No regards to maintainability. |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| | | for software functions normally will result in higher maintenance costs) | (For maintainability reasons, the software has to be lean, structured and well documented) | |
| | | | | |

### *From the perspective of the EC*

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 1 | Adhere to INSPIRE directives and recommendations. | Yes | Yes | No adherence |
| 2 | Develop a clear and re-usable architecture. | Yes, distributed clear architecture (see INSPIRE) | Yes, distributed clear architecture (see INSPIRE) | No, only a central database |
| 3 | Make the system extendible. | Yes, all EC member states can be connected. | Yes, up to 5 NMCAs can be connected during the project duration of 30 months. The architecture will be designed for phasing in about 30 member states. | No, only 5 countries are connected |
| 4 | Use (ISO/OGC/W3C) standards where possible. | Yes | Yes, and INSPIRE | No |
| 5 | The service must continue to work after the EGN project has finished. | Yes | Yes | No |

| # | Requirement | Optimum | Average | Unacceptable |
|---|---|---|---|---|
| 6 | Single searches for geographical names within the EGN infrastructure will be free of charge. | Yes | Yes | No |