

Minutes

Title	Workshop on user/business requirements
Date	January 15 th , 2007, January 16 th 2007
Location	Utrecht University
Subject	Defining user/business requirements for EuroGeoNames (EGN)

1 Executive Summary

The main points to note from the meeting are:

- Introduction to the Gazetteer data collected and used by the Dutch and Belgian NGI
- Information to case studies of the needs and wants of user groups of geographical names.
- Information about user requirement surveys and market analysis.
- Information about (potential) commercial applications.
- Providing an information model that can be used as a basis for the data model, the EGN web service development, the design of the applications and the business model.
- Determine and compile use-cases based on the RISE methodology.
- Provide a final report on the user and business requirements.

2 Participants

Nr	Name	Institution
1	Ferjan Ormeling	Utrecht University
2	Nicoline McCarthy	Utrecht University
3	Martin Amersfoort	booking.com
4	Rob van Essen	Teleatlas
5	Tjeerd Tichelaar	Wolters-Noordhoff Atlasproducties
6	Pier-Giorgio Zaccheddu	BKG Deutschland
7	Lise Just	EuroGeographics
8	Andreas Hadjiraftis	Department of Lands and Surveys, Cyprus
9	Mara Sterna	LGIA, Latvia
10	Uldis Gitendorfs	LGIA, Latvia
11	Roman Stani-Fertl	University Vienna
12	Johnny Andersen	Statens Kartverk, Norway
13	Kathleen van Doren	National Geographic Institute Belgium
14	Maarten Storm	TDKadaster
15	Nico Bakker	TDKadaster
16	James Reid	EDINA

17	Markus Ulrich	GeoTask
18	Bernhard Schneider	GeoTask
19	Arendjan van der Neut	Geodan IT b.v.
20	Annita Vijverberg	Geodan IT b.v.
21	Peter J. Meijer	Geodan IT b.v.
22	Georg Börner	ESRI Germany
23	Axel Schaefer	ESRI Germany

3 Attachments

4 Contents

TOP	Topic title
1	1 st workshop user/business requirements – lectures and discussions
2	Summary of workshop, discussion points
3	Internal meeting on Tuesday, January 16 th 2007
4	Next meetings

5 Minutes and results of the discussion

Minutes	Action
<p>1. 1st workshop user/business requirements – lectures and discussions. Most presentations are available at the SharePoint. Please point to the “mydms” section and there at EuroGeoNames → Meetings_Workshops_Minutes → 2nd-WS_User/business requirements.</p> <p>Session 1: Ferjan Ormeling - Workshop Objectives (Utrecht University): The presentation is available at the SharePoint. Ferjan Ormeling pointed out the objectives of the EGN project and the specific work-package 2. He introduced the participants and the programme as well as the objectives of this workshop and its envisaged results.</p> <p>Maarten Storm - Names collecting and processing procedures in the Netherlands, nature and sources of recent name changes (TDKadaster): The presentation is available at the SharePoint. Main discussion points are the language-specific problems in the Dutch, Frisian, German and Belgian locations, including dialect-names. Discussions in the Netherlands might lead to a more official status of dialect names, which might have to become part of the gazetteer-database.</p> <p>Kathleen van Doren - Belgian toponymic data collected by the National Geographic Institute: current and future characteristics (NGI): The presentation is available at the SharePoint. As you know, the official languages in Belgium are French, Dutch and German. Belgium is divided into four linguistic regions:</p> <ul style="list-style-type: none"> • the Flemish-speaking region • the French-speaking region • the German-speaking region • the bilingual (Flemish-French) Brussels-Capital region <p>In accordance with the linguistic legislation to which it is submitted, the National Geographic Institut only uses the endonyms. For instance Antwerpen and not Anvers, as it is a toponym situated in the Flemish-speaking region. Liège and not Luik, as it is a toponym situated in the French-speaking region. But Bruxelles and Brussel for the bilingual Brussels-Capital region. Some municipalities have a special status assigning linguistic facilities to their inhabitants:</p> <ul style="list-style-type: none"> • To the French-speaking inhabitants of the Flemish-speaking region • To the Flemish-speaking inhabitants of the French-speaking region • To the German-speaking inhabitants of the French-speaking region • To the French-speaking inhabitants of the German-speaking region <p>In those municipalities, only the translations of the names of the municipalities into the language of the protected minority are given.</p>	

2nd session: Session 2: Identification of needs and wants of user groups of geographical names: case studies

Annita Vijverberg: Uses of geo-information by public institutions and private companies (GeoDan): The presentation is available at the SharePoint. The presentation provides a division into processes and functions as well as a matrix related to the needs/wants/nice to have statements. The main discussion points are related to that matrix: *Axel Schaefer* mentioned that the distribution process requires a bigger need for the place name translation than for the place name normalization function which leads to the needs of a fuzzy search rather than the definition of normalized names. In many processes the functions are based on the address level rather than on place names. *Bernhard Schaeffer* remarked that also the reverse of these functions would be applicable. *Rob van Essen* mentioned that the system integration may be an important part for solving concrete issues regarding the functions and processes; he proposed to take object coordinates as an ID. *Ferjan Ormeling* added that emergency systems may be considered as an additional process and should be considered in the EGN context. *Rob van Essen* mentioned that the place name enrichment may be seen as part of the attribute data, so this point may be deleted concerning to optimizing project-specific object information. As an example he declared the traffic-jam systems, where system integration is based on the names of the features. *James Reid* noted that the official status of the information in the EGN project may be more important and that this has to be developed in concrete use-cases.

Martin Amersfoort: Requirements of geographic name databases for online hotel reservation business: The document is available in the SharePoint. This presentation showed the solution of the hotel reservation business available at <http://www.booking.com>. Variant names are already used in this use-case. An interesting topic for booking.com would be the frequency of using these variant names to optimize the search function. Boundary decisions of the place-names are based on common sense, missing features are digitalized pragmatically. A customized service answering to the needs of booking.com would be a “simple” database with most fundamental data, like names, variants and coordinates (including area polygons) (in latitude/longitude). The procedure for incorporating a new language into the database is to hire some people for translating. It was asked how a new country, like Latvia, would be dealt with. *Uldis Gitendorfs* indicated a list of exonyms had been prepared.

Tjeerd Tichelaar: Requirements of cartographic editors: The presentation is available at the SharePoint. *Tjeerd Tichelaar* presented the issues important for the atlas producer Wolters-Noordhoff. Up-to-date information from EGN would be used but have to fit into the in-house workflows, although naming decisions in atlas production must be taken with responsibility. This comes with several unique issues according to name translation, de facto standards, resistance to change, relevance, political issues, etc. Old and obsolete names are used for historic issues and therefore kept in the database.

James Reid: Requirements of Search Engines and of document management system providers: The presentation is available at the SharePoint. James Reid from EDINA presented the GeoCrossWalk as an example of an integration of a Gazetteer into other systems. The second part presented “use cases” from the Santa Barbara Digital Gazetteer and Practice Workshop with the extended use cases of Harvesting, look-up, reverse lookup, geo-parsing and ontology. The future steps are building up a historical gazetteer for saving changes in place names. The data model takes care of historical content. Also good metadata is required to optimize the description of the

results.

Session 3: Inventory of data and functionality needed: user requirements surveys and market analyses

Markus Ulrich: Review of existing studies on user requirements and market analysis for utilization of Geo-information and geographical names:

The presentation is available at the SharePoint. The presentation concerns specific categories for user/business requirements regarding the data, the user interface, the search and gazetteer-specific issues. Common requirements are listed on the presentation. Questions occur regarding the definition of “high data quality”: meant as usability in conjunction with other data and the integration into external systems. *Ferjan Ormeling* noted that complete coverage is hard to define, for example the complete coverage of mountains in the dataset. *James Reid* mentioned that EGN has to be understood as a mid-scale product that cannot assure location based services. In further discussion it was pointed out that the user-audience of existing gazetteers is still not clear for the EGN project, and that the awareness of existing data-sources need to be clarified for the users. Marketing issues may be a starting point for the relevance of gazetteer services.

Georg Börner/Axel Schaefer: Potential commercial applications: The presentation is available at the SharePoint. The presentation was divided into two parts, first the preliminary analysis of commercial applications using gazetteers and second the presentation of ArcWeb Services for showing the needs for distributing spatial data, services and/or applications in a commercial context. *Börner* pointed out that the consortium members may add known commercial application using gazetteers into the distributed Excel table list. The focus is kept not on commercial gazetteers but on commercial applications using gazetteers, which is another focus than a list of available gazetteer-services. Also the question of completeness occurred which here means completeness according to the application context. Security issues have to be taken into account for the architecture of EGN and the division into free access and commercial access. *Johnny Andersen* mentioned that the user and business needs and wants stand against what is possible in such a service. *Rob van Essen* asked for the difference between Gazetteer services and Feature Services, which here means the service providing the connection between names and coordinates and the feature services which grants access to customer-specific attribute data and that it can be seen as a two step approach where the gazetteer data can be used in different context using different underlying feature services.

Rob van Essen: Tele Atlas and Geographic Names: The presentation is available at the SharePoint. Rob van Essen pointed out the need for geographic names in the context of Tele Atlas as a data and content provider. He pointed out the need for the EGN service to adjust and update the Tele Atlas data with the official names and locations of the EGN service. *Tjeerd Tichelaar* asked about the handling of homonyms: they are distinguished by the different feature types (e.g. hotels, streets) so that the number of results may be adequate for the user. *Axel Schaefer* asked if the handling of gazetteer and address data is handled differently: An address-finder is more complex than a gazetteer but that is depending on the client interface. Van Essen pointed out Tele Atlas needed extensive linguistic name attributes, like language, type and subtype, name components and character sets.

Nico Bakker: Value-adding resales of TDKadaster names files: *Nico Bakker* presented the value-added resellers of geographical name datasets in the Netherlands. Geographic names are used in maps, printed gazetteers, web

gazetteers, historical name research and web applications. It takes place in civil products, name registers and military products. It can be bought for one time or as a user-license for one year. Due to military requirements different codes are added to the names. Name indexes for example are incorporated in printed topographical atlases, in digital web-applications and in several GPS applications.

Rob van Essen asked whether address numbers were also available, and was answered that these could be added indirectly through a link with another database (the street address database of the NMCA).

The discussion following that presentation leads to the formulation of the requirements of the EGN-service. *Rob van Essen* pointed out that the data-model of EGN should make provisions for adding all possible data in the future.

Ferjan Ormeling: Information model: Please see the next topic of this protocol.

2 Summary of workshop, discussion points

The final discussion starts with the presentation by *Ferjan Ormeling*: Information model. The presentation is available at the SharePoint. The presentation was intended to open the discussion. It soon became clear that the group did not opt for a lowest common denominator approach. *Rob van Essen* pointed out his view of the quality needs, where scales are the wrong approach because the countries have a different number of names on the map (and so in the database). He would prefer predominant features in a specific scale, like e.g. city-areas. Therefore he would prefer a feature-specific, scale independent approach, like all settlement. As a counter-argument for that approach *Ferjan Ormeling* mentioned the different definitions for “settlements”, but was answered that we should use what the NMCA’s delivered. *Van Essens* argumentation leads to the thesis that a “complete” subset of categories would be better than one or more incomplete subsets so that a 100% completeness would be a valuable product and sales argument.

Axel Schaefer pointed out that in the customer-driven solution like Tele Atlas the size of the data matters and that they normally have an approach different from map production solutions. Like van Essen has mentioned in his presentation the customer feedback is an important point for the acceptance of the data and in this case the update mechanism for the data.

Andreas Hadjiraftis leads to the point, that the scale has to be used as quality-attribute regarding to a scaleless approach. He explained that delivery of digital cartographic data to the public has to have good quality control criteria. The contents and the accuracy of the data have to be defined for each dataset and has to be known to the public. When the scale is used as a quality control criterion, automatically the content, the accuracy and the density of information is regulated by applying simple cartographic rules. These parameters are easily implemented in a GIS environment.

Pier-Giorgio Zacchedu mentioned that the first months would make clear what data is available and that one country could provide data that other countries could not.

James Reid pointed out to the need of use-cases to focus and break down the high level discussion. He referred to the RISE methodology that EuroGeographic has provided and the value of use-cases created on that methodology.

Ferjan Ormeling thanked all participants for their presentations and participation in the discussions and closed the workshop.

3 Internal meeting on Tuesday, January 16th 2007

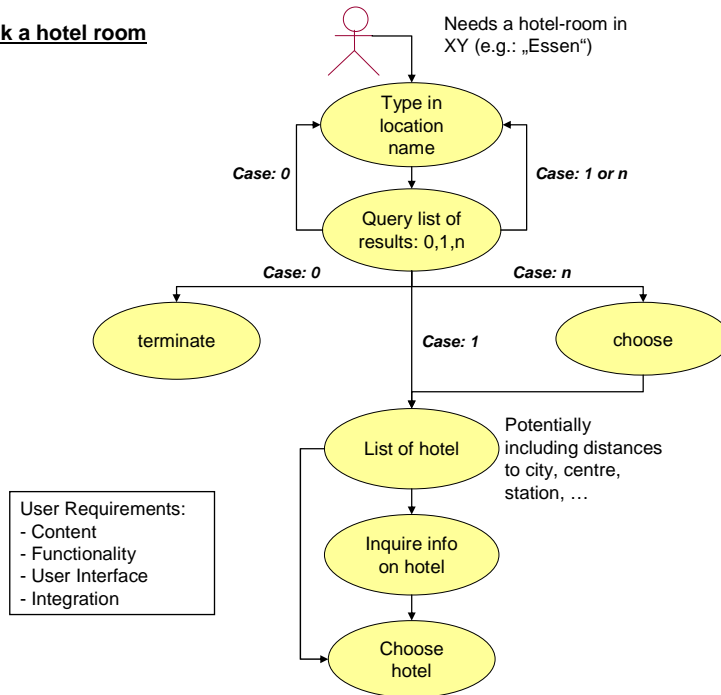
Present: Peter J. Meijer (Geodan) as well as Ferjan Ormeling, James Reid, Arendjan van der Neut, Markus Ulrich, Bernhard Schneider, Georg Börner and Axel Schaefer who had already attended the workshop on the previous day. This internal meeting summarized and carried on the discussion of the workshop.

The meeting started with a sketch of a simplified architecture of the EGN-system to point out the differences between the client application, the EGN service and the content. The discussion following that base differences the “end user” using the client application and the “direct user” using the EGN service which behaves more like a service integrator. The pragmatic way would be to provide as much content and functionality as possible to make a product for the main customer. The project may switch the focus from the end-user application to the service; it should be applicable to provide only the service and let the clients decide how to use it. Two ways in terms of the use-cases were distinguished: the end-user at his computer and the machine-to-machine communication. The outline of the RISE methodology was presented in order to start from these use-cases, compile them and put them on a higher abstraction level on which the requirements would be based.

James Reid shared a small paper describing the RISE methodology example (see RISE Methodology & Guidelines on Use Case & Schema Development, Version 1.1; page 41f. for reference), explaining that requirements can not be defined if the use-cases are not validated. The RISE document should be seen as a synoptic document and summary of use-cases.

Based on this methodology an example of a rise methodology use-case was created for the “Book a hotel room” example. The graphic below is a modified report of that draft.

- The user needs a hotel room near a place XY. He types in the location name.
- The response may be of three possible results: the location is found directly, exactly one location is found and more than one location is found.
- Depending on the number of results, different alternatives are possible: Terminate the workflow, try it again or choose one result.
- The list of the hotel may include information of distances to city or other information.
- The user may inquire further information of the hotel and finally choose one hotel.

Book a hotel room

The ensuing discussion was about the distinction between a use-case and a business model containing several use-cases. It was mentioned that this elaboration of use-cases may be finished if a point of iteration is finished and that the abstraction may follow after that. The use-case has to describe what is done and what the user does.

It was made clear that the RISE document doesn't note other user-requirements. Questions occur when to decide if a use-case is finished and can be called a use case and that it can only be an approach as good as possible. Some use-cases may also include several actors, like a common user and other specific actors. The influence of the data was pointed out in this use-case: Features have to be provided to start the geo-processing on that feature (distance-measuring). Other attribute-data must not be provided and other use-cases may demand other attribute data.

Bernhard Schneider pointed out the user-requirements to be focussed:

- content and data,
- functionality,
- user interface and
- Integration containing the integration into the database (e.g. cities or POIs) and under aspects of accuracy and completeness and scale and resolutions.

Further on, *Ferjan Ormeling* summarized the requirements in the **data** section:

- Name categories
- Relevant completeness, minimum content
- Coverage
- Currency
- Spatial accuracy
- Additional information: Variant names, coordinates, pronunciation.

The user-requirement **functionality** may be seen as the connection between receiving input and receiving output.

<p>The user-requirements to the user-interface is a more complex user-requirement which is specific to the work-packages referring to the reference application.</p> <p>The user-requirements of the integration concern the web-service itself and are summarized as:</p> <ul style="list-style-type: none"> • the protocol, • the API and • the availability. <p>A list of use-cases was developed which have to be described and compiled by the several participants:</p> <ul style="list-style-type: none"> • Hotel-booking (Geodan) • Geoname database update (Utrecht University) • Web map application with search functionality (reference application BKG) (GeoTask) • Geoparsing use-case, digital libraries (EDINA) • Historical research (EDINA) • Metadata information systems, OGC catalogue service (ESRI) • Spelling and pronunciation service (Utrecht University) • Real estate agencies (Utrecht University) • Interpreter and translation service (Geodan) • Searching documents that are spatially references (not allocated) <p><i>Ferjan Ormeling</i> proposed to collect information and use-cases in the next days (end of week 4/2007) to compile them and to send it around for corrections in week 5/2007.</p> <p><i>James Reid</i> was going to provide the RISE article into the Sharepoint. <i>Bernhard Schneider</i> was going to provide a template for the Sharepoint. <i>Ferjan Ormeling</i> provides to compile the use-cases and ask for comments from partners for the executive summary.</p> <p>Ferjan thanked the WP2 group for staying over and discussing and determining the procedure how to come up with the requirements.</p>	<p>Geodan, GeoTask, UU</p>
<p>4 Next meetings</p> <ul style="list-style-type: none"> • February 6th/7th: Workpackage 3/4/6 meeting (Amsterdam) • March 5th: Workshop preparation meeting (Madrid). • March 6th: 2nd Workshop Consortium & Reference Group - jointly with UNGEGN WG on Top Gaz Files (Madrid) • March 7th: 3rd consortium meeting (Madrid). 	
<p>5</p>	

Signature: Axel Schaefer