United Nations
Group of Experts on
Geographical Names

Working Paper
No. 1

Twentieth Session
New York, 17 – 28 January 2000

Item 9
of the Provisional Agenda

MEETING OF THE WORKING GROUP ON TRAINING DATA FILES GAZETTEERS

Finnish National Land Survey's Geographic Names Register

Submitted by Teemu Leskinen, National Land Survey of Finland, Development Centre, Helsinki, Finland

Finnish National Land Survey's Geographic Names Register

Teemu Leskinen
National Land Survey of Finland
Development Centre
Opastinsilta 12 C, P.O. Box 84
FIN-00521 Helsinki, Finland
E-mail teemu.leskinen@nls.fi

Abstract

Place names are often hard to model in a Geographic Information System (GIS). Sometimes they can be understood as well-defined attributes of well-defined geographic objects, but often these objects are difficult to give structure to, or even to point out, or mark the boundaries for, in the real world. Anyhow, these names must be included in the database for, e.g., cartographic reasons. In most GIS implementations the place names seem to form a layer of their own, with no tighter connection to the (possible) actual geographic objects in the database than location and classification. So is the case in the Finnish National Topographic Database too. However, the place names with object co-ordinates and a carefully selected set of attributes form a versatile source of information for separate or integrated service databases providing multiple different fields of application and products. This paper describes some of these possibilities found when developing the National Geographic Names Register in the National Land Survey of Finland.

1. Introduction

According to the United Nations Group of Experts on Geographical Names (UNGEGN meeting, June 1994) place names "*identify our landscape, express our national identity and cultural heritage, create our framework of orientation and keys to the information age and promote our awareness of the world around us. Geographic names are elements of basic information needed to refer to places in the world. The spelling and application of geographic names must be clear, accurate, current and unambiguous. Expressed in their standardised forms the place names support effective national and international communication and are essential to our socio-economic development in such fields as trade and commerce, population census and national statistics, property rights and cadastre, urban and regional planning, environmental conservation, natural disaster and emergency preparedness, security strategy, search and rescue operations, automatic navigation and tourism, map and atlas production.*"

Keeping in mind these benefits, but primarily to support and rationalise the national map and map database production, the National Land Survey (NLS) decided in 1995 to create a National Geographic Names Register. The time was right, since the NLS had launched the Topographic Data System (TDS) in 1992, and the primary data for the names register would be gathered as part of the TDS data collection process with no extra costs. In Finland standardised geographic names databases or gazetteers have been lacking this far. The only proper nationwide collections of accepted place names have been the archives (mainly manual) of the Research Institute for the Languages of Finland (RILF), and the Basic Maps 1:20,000 by the NLS, the geographic names of which have been accepted by the RILF.

1.1 National Topographic Database

The source of the National Geographic Names Register is the National Topographic Database (TDB). The TDB is a part of the National Topographic Data System and consists of the most detailed and up-to-date general topographic data of the whole of the country. The TDB production is decentralised in our 13 regional offices around Finland.

The TDB includes among (or as attributes of) other geographic features about one million names for physical features, populated places and administrative areas presented on Finnish Basic Maps 1:20,000. The place names in the TDB are divided in 7 feature groups and further classified in 47 feature types. In addition to place names the TDB includes other text objects, like explanatory texts and numeric values.

2. National Geographic Names Register

The National Geographic Names Register is a service database consisting of the National Place Name Register (PNR) and the National Map Name Register (MNR). The data source for the PNR and the MNR 1:20,000 is the TDB, from which the data are loaded in them. The three elementary objects in the database are *place*, *place name* and *map name*. *Places* and *place names* build up the PNR; the PNR and the *map names* form the MNR.

In addition to these three basic tables there are some 30 background tables in the database. For the PNR they include tables for maintaining the object history, the area divisions hierarchy, the feature groups and the official statuses of different languages by municipality. For the MNR there are additional tables for maintaining the product information and introducing the codes used in the database. These background data enable queries with various spatial and attribute search criteria.

2.1 National Place Name Register

The PNR is scaleless including no cartographic information. The PNR objects *place* and *place name* consist of following data:

Table place
- Unique place id as a possible link to external GIS data
- Geographic object co-ordinates (Gauss-Krüger metric X,Y;Z); the centre point (the mouth for a river)
- Information about the location of the object by municipality, General Map Index and National Rescue Grid
- Feature type
  - In the background tables feature types can be aggregated as feature groups freely

Table place name
- Unique place name id
- Place id as a link to the *place*
- Proper spelling of the place name (accepted by the RILF, unabbreviated, upper and lower case)
  - North European 8-bit character set ISO 8859-10, extended with the letters needed in Skolt Sami
- Language of the name (Finnish, Swedish, Northern Sami, Inari Sami, Skolt Sami)
  - The official status of different languages in Finland's 450 municipalities is given in background tables. On Finnish maps, in bilingual areas, both names are presented, e.g., Helsinki, Helsingfors. The order (above, below) depends on which language group has the majority in the municipality
- Source of the place name

The object history for places and place names is stored in tables *place history* and *place name history*:

Tables place (place name) history
- Place id (place name id)
- Event code; addition, deletion; change of location or feature type (change of spelling or language)
- Event time
- User id

When 'deleted' the place or place name record is not actually removed, but moved to other tables called *old place* and *old place name*.

2.2 National Map Name Register

The MNR includes the unique product-dependent cartographic information for the place names. The PNR is integrated as a part of the MRN. The MNR specific object *map name* consists of following information:

Table map name
- Unique map name id
- Place name id as a link to the *place name* (and to the *place*)
- Position (X,Y) in the product co-ordinate system
- Text box 'handle' (1..9; which point the text position is referring to in the rectangle imagined around the text)
- Alternative spelling (product-dependent spelling which differs from the spelling in table *place name*; for example, a name divided in two lines on a map forms two entries in this table, both having the same place name id)
- Text font code
- Text size (graphic size, mm/100)
- Text colour code
- Letter tilt angle

- Capitals flag (whether the name is written upper case in this product)
- Text direction (expressed as relative co-ordinates (dx,dy))
- Spacing flag (whether the text direction parameters (dx,dy) are also used for indicating the length of the text box)
- Bending (up to 32 pairs of co-ordinates for curved texts, in product co-ordinate system)

The map names are arranged as products. Every map name instance is related to one and only one product. The map names are connected to a product in the table *map name product*. The products are introduced in table *product* and grouped in additional background tables.

Table map name product
- Map name id
- Product code

Table product
- Product code
- Product group code
- Co-ordinate system code
- Map scale class code
- Product name

2.3 Integration, Data Model

The data model of the National Geographic Names Register is illustrated in Figure 1. The National Place Name Register and the National Map Name Register are integrated as one single consistent database where every piece of information is stored only once. A *place* may have one or more *place names* (e.g., in different languages) and a *place name* may have 0-N cartographic *map name* appearances in one or several products. For example, the largest lake in Finland, Saimaa, has one *place* entry and one *place name* entry in the database. The *map name* 'Saimaa' (or often 'SAIMAA', see 'capitals flag' in the table *map name*) would probably appear in all cartographic products in all map scales, and possibly several times in one product. On the other hand a little pond named Patalampi might have a *map name* appearance just once, in the Basic Map 1:20,000.

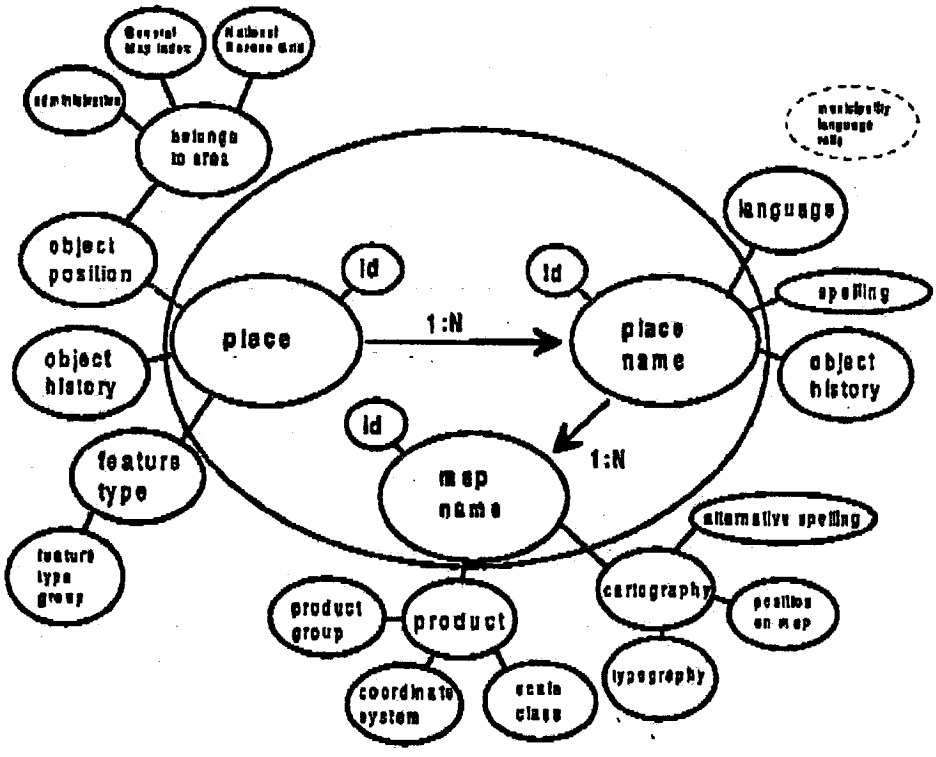

Figure 1. A 'mind map' presentation of the National Geographic Names Register

2.4 Implementation of Database

The National Geographic Names Register is implemented as a relational database (Oracle) running on Unix.

3. Maintaining National Geographic Names Register

The National Geographic Names Register is 'self-sufficient': all data both for the PNR and the MNR are stored in a single database and nowhere else. However, the nationwide register needs processes and tools to maintain the data. The processes can clearly be divided in two parts: maintaining the place names and the map names 1:20,000, and, creating and maintaining the small scale map names from 1:100,000 down to 1:4.5 million.

3.1 Maintaining Place Names and 1:20,000 Map Names

The Topographic Data System and the Topographic Database are the origins for the place names and the map names 1:20,000, and the data are maintained as a part of normal TDB updating and Basic Map 1:20,000 compiling processes. The TDB is processed using the GIS software developed in the NLS, and eventually using an object oriented GIS being currently evaluated. The PNR and the MNR 1:20,000 are not in on-line connection with the TDB. They can be updated with the changes made since the previous update when needed, by using an automatic process. If need be, the PNR and MNR 1:20,000 can also be edited directly using an application software developed on Arc/Info.

3.2 Creating and Maintaining Small Scale Map Names

The processes and tools for maintaining the small scale map names are developed on Arc/Info running on Unix. Arc/Info is also the production and storage system for other small scale GIS data. The main issues in processing are generalisation, name placement and other cartographic editing. The process workflow is illustrated in Figure 2.

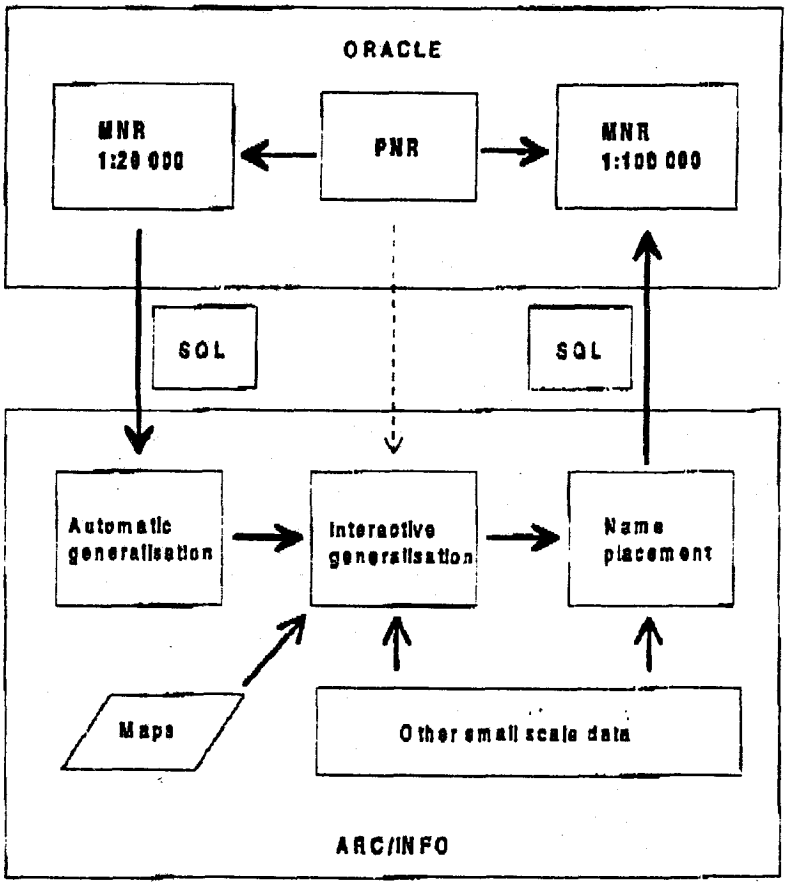

Figure 2. The small scale map name generalisation and editing process

Generalisation Process

Besides the scale 1:20,000 the MNR includes nationwide small scale map name presentations in 1:100,000, 1:250,000, 1:500,000, 1:1 million and 1:4.5 million. They are cartographically compiled with other small scale vector data in the respective scales. The generalisation process is incremental; the map names in 1:100,000 are selected from the map names in 1:20,000, the source data for the 1:250,000 map names are the map names in 1:100,000, and so on.

The map names in larger scale are fetched from the Oracle database to an Arc/Info work coverage using appropriate spatial criteria. In Arc/Info the names are handled on three levels: the upper, the middle and the lower level. The upper level consists of names that are selected to the target product. The lower level has the names that have been rejected. The

middle level is the workspace, i.e. it includes the names that still need to be taken a stance on. The user has possibilities to change the names from any one level to another interactively, one by one, or using group operations of different kinds.

For each combination of a source product and a target product a parameterised "batch" generalisation process has been implemented to give a starting point for interactive name selection. The process places each source product name in one of the three levels on grounds of the combination of the source product name feature type and text size. Naturally, in the following interactive selection phase several other criteria must be taken into consideration. One of them is all the other vector data in the target area. The names' generalisation cannot be started unless all other data are compiled and presentable as background information on screen or paper.

Name placement and editing
After the names have been selected in the upper level they need cartographic processing. To give a starting point for interactive and more enhanced editing a parameterised 'batch' process sets the default typographic parameters for each name. As in generalisation this process is directed by the combination of the source product name feature type and text size. This process doesn't affect name placement.

In the interactive editing phase the user may edit the name placement and other cartographic parameters freely, yet within a product-dependent framework given for every feature type. During this phase the user is able to bring all the necessary and helpful vector and raster data on the screen as background information. After completing the interactive phase the user stores the selected and edited map names from Arc/Info to the Oracle database.

4. Retrieval Interface

The graphic end-user retrieval application for the National Geographic Names Register's Oracle database was developed on NT using Visual Basic and TCP/IP/SQL*Net/ODBC tools. When making a query the user defines the register, the names he/she wants, the data set he/she is interested in, and how to output the results. The flow of a query is presented in Figure 3.

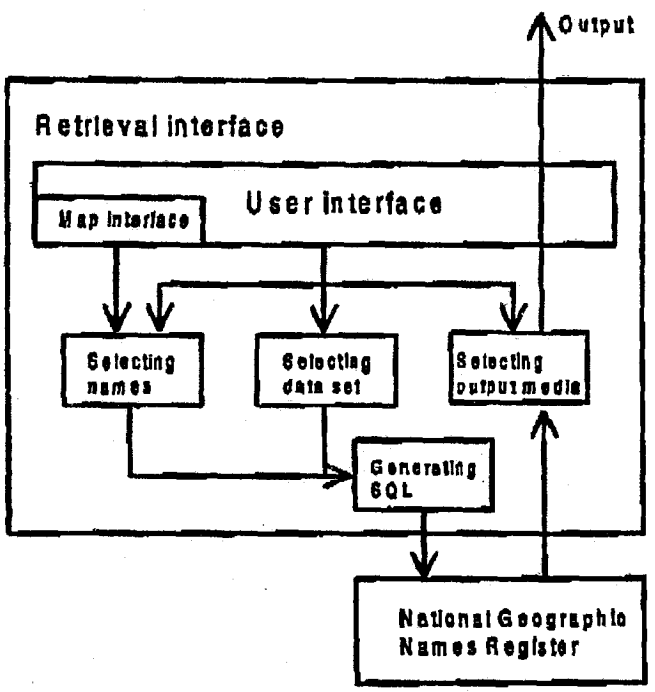

Figure 3. A register query

4.1 Selecting Register
The user interface is very similar to both the PNR and MNR queries. In map name queries the *product* becomes a search key among others.

4.2 Selecting Names

The user may apply combinations of various criteria when searching the names. The *spatial* conditions are set by choosing municipalities, the General Map Index sheets and the National Rescue Grid squares. By nature, they all form hierarchical area divisions, and their aggregates, like counties, can be used as well. A rectangle given by two co-ordinate pairs is one alternative. *Feature type* or a group of feature types is a common search key as well as the *language* and *spelling* of names.

4.3 Selecting Data Set

The data set in this context means the data model for the query results. Though the data in the database is divided in several tables, the data set model is one table. The data set contains data and almost any data in the database can be included in one data set. Every item of data is equal either to some data field in the three primary tables, or any background table data field that can be derived from the primary table data fields; the structure of the database does not restrict the data sets. Once defined, a data set can be stored for future use.

4.4 Selecting Output Media

The user can output the query results to the screen, to a text file, or as a database table (Access, Excel etc.).

5. Applications, Products

The Introduction presented a large variety of applications associated with proper place names proposed by the UNGEGN experts. Implemented and possible fields of application and products related to the National Geographic Names Register are, for example:

- *Rationalisation of the NLS map (database) production.* Organised data management, uniform principles and specialised tools for geographic names in full integration with other map database production.
- *Place name and map name data products.* Standard as well as tailored data sets for GIS developers and map makers.
- *Gazetteers.* Product related, regional, national and international gazetteers. On-line gazetteers in the Internet.
- *Internet services.* Karttapaikka (MapSite, http://www.kartta.nls.fi/) is a service that provides the national topographic map (Basic Map 1:20,000) for on-line public use. At present a database of about 30,000 names is implemented for place searches. Plans exist to integrate the incrementally generalised names data sets of the MNR to the service as a flexible and 'intelligent' data source for navigating and zooming.
- *International projects and databases.* Contribution of approved geographic names for international GIS and place name databases. Thanks to the object history records, updates are easily arranged.
- *Research.* A large, nationwide and homogenous database of accepted place names with feature types and object co-ordinates forms a new source of information for researchers, such as onomasticians, historians and natural scientists. By combining different spatial, attribute and spelling criteria they can find answers to existing problems and derive new subjects of study.

References

Leskinen, T. (1999): National Place Name Register Integrated with Cartographic Names Register for Multiple Scales and Products. *Proceedings 19th International Conference of the ICA*, Ottawa.