7 June 2002

English only

Eighth United Nations Conference on the Standardization of Geographical Names
Berlin, 27 August-5 September 2002
Item 12 (d) of the provisional agenda*
Toponymic data files: automated data-processing (ADP) systems

The Geographic Names Register of the National Land Survey of Finland

Submitted by Finland**

* E/CONF.94/1.
** Prepared by Teemu Leskinen, National Land Survey of Finland, Development Centre.

02-43519 (E)
0243519

1. Introduction

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 even pointing out or outlining these objects is difficult in the real world. Nevertheless, these names must be included in the database, at least for 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. This is the case with the Finnish National Topographic Database, too. However, the place names with object co-ordinates and a selected set of attributes form a good source of information for separate or integrated service databases providing different fields of application and products.

2. Background

The National Land Survey (NLS) launched the Topographic Data System (TDS) and the Topographic Database (TDB) in 1992. The TDB is a centralised database comprising the most detailed and up-to-date general topographic data of the whole country. The TDB production is decentralised in 13 regional NLS offices around Finland. The TDB includes, among other geographic features, about 800,000 names for physical features, populated places and administrative areas presented on the Basic Map 1:20,000. The place names are divided in 7 feature groups and further classified in 47 feature types.

In 1996 the NLS decided to establish a National Geographic Names Register. The time was right since the primary data for the names register would be gathered as a part of the TDS data collection process with no extra costs. Finland has had no national databases or gazetteers of standardised geographic names. The only proper nation-wide collection of accepted place names has been the Archive of Names (manual, about 2.5 million toponyms) of the Research Institute for the Languages of Finland. In addition, the geographic names presented on the Basic Map have been approved by the Institute.

3. Database

The National Geographic Names Register is a relational service database consisting of the National Place Name Register (PNR) and the National Map Name Register (MNR). The PNR is scale-less and includes no cartographic or product-related information. The MNR includes the unique product-dependent cartographic information for the place names in the PNR. The data source for the PNR and the MNR 1:20,000 is the TDB, from which the data are loaded in and processed further. The three elementary objects in the database are *place*, *place name* and *map name*. *Places* and *place names* build up the PNR. The PNR is integrated as a part of the MNR; the PNR together with *map names* form the MNR. The map names are arranged as products. Every map name instance is related to exactly one product.

In addition to the three basic tables there are some 30 background tables in the database enabling queries with various spatial, attribute and product search criteria. For the PNR they include tables for managing the object history, the area divisions hierarchy, the feature type groups and the official status of different languages by municipality. For the MNR the additional tables handle the product information and introduce the codes used in the database. The GNR is an Oracle database running on Unix.

3.1 Data Model

The data model of the GNR is illustrated in Figure 1. The PNR and the MNR 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 would probably appear in all cartographic products in every map scale, and possibly several times in one product. On the other hand, a little pond, Patalampi, might have a *map name* appearance just once, in the Basic Map 1:20,000.

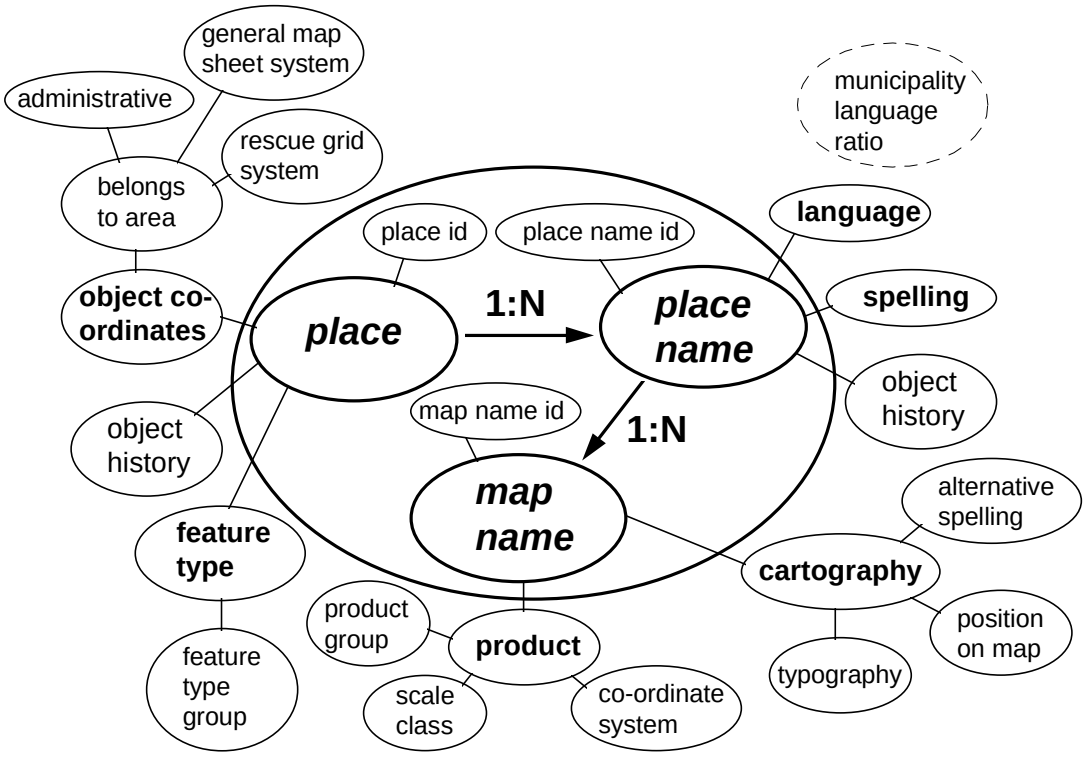

Figure 1. The data model of the National Geographic Names Register

3.2 Place Name Register

The PNR objects *place* and *place name* consist of the 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); centre point (mouth for a river)
- Information on the location of the object by municipality, General Map Sheet System and National Rescue Grid System
    - Generated automatically using point-in-polygon algorithm
- 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 Research Institute for the Languages of Finland, unabbreviated, upper and lower case)
    - North European 8-bit character set ISO 8859-10, altered for the extra letters needed in Skolt Saami
- Language of the name (Finnish, Swedish, North Saami, Inari Saami, Skolt Saami)
    - The official status of different languages in Finland's 448 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. However, in Lapland, Finnish names are always presented above and Saami names below.
- 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.*

3.3 Map Name Register

The MNR specific database 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 a common 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) also indicate the length of the text box)
- Bending (up to 32 pairs of co-ordinates for curved texts; in product co-ordinate system)

A map name is connected to a product in 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

4. Maintenance

The National Geographic Names Register is 'self-sufficient': all data for both the PNR and the MNR are stored in a single database and nowhere else. However, the register needs processes and tools to maintain the data. The processes can be divided into two distinct 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.

4.1 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 an application developed on an object oriented GIS, Smallworld. 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 automatic and semi-automatic processes. If need be, the PNR and MNR 1:20,000 can also be edited directly, using an application software developed on ArcInfo.

4.2 Small Scale Map Names

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

4.2.1 Generalisation Process

Besides the scale 1:20,000 the MNR will include nation-wide small scale map name presentations in 1:100,000, 1:250,000, 1:500,000, 1:1 million, 1:2 million and 1:4.5 million. They are cartographically compiled with other small scale vector GIS 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 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 ArcInfo work coverage using appropriate spatial criteria. In ArcInfo 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 left out. The middle level is the workspace consisting of the names that still need processing. The user is able 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 feature type and the text size. Naturally, in the following interactive selection phase several other criteria must be taken into consideration. One of these 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 and paper.

4.2.2 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 careful editing a parameterised ‘batch’ process sets the default typographic parameters for each name. As in generalisation this process is directed by the combination of 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 ArcInfo to the Oracle database.

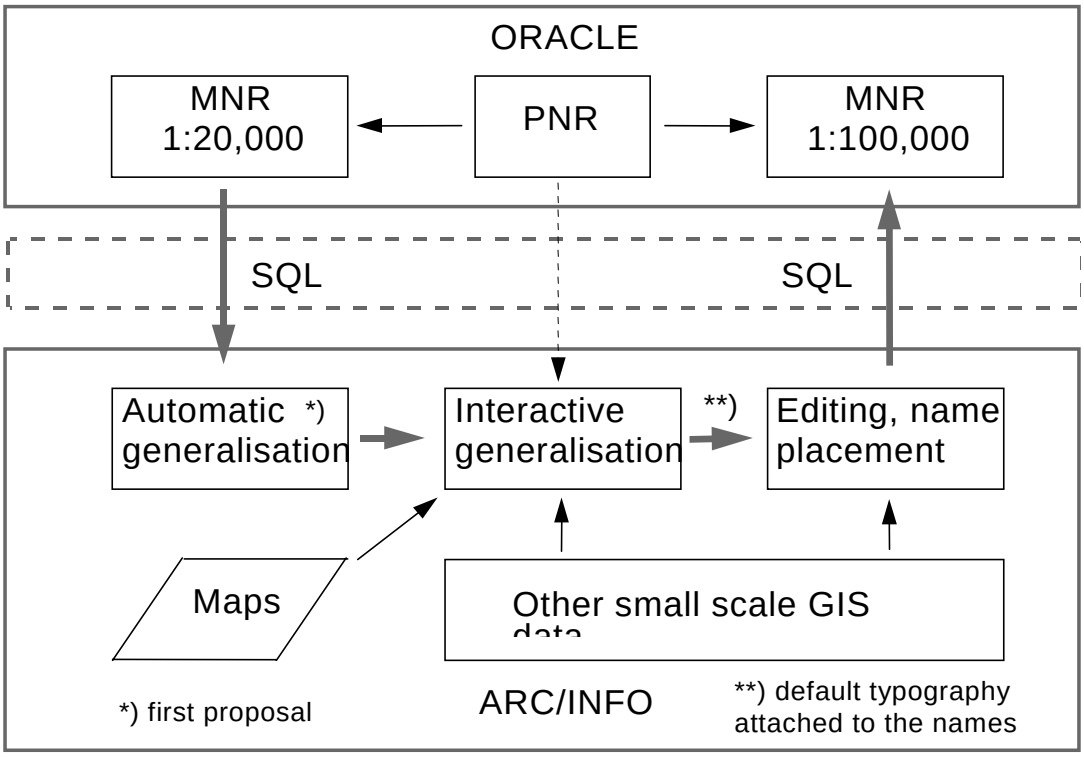

Figure 2. The small scale map name generalisation and editing process 1:20,000 -> 1:100,000 (1:100,000 -> 1:250,000 etc. respectively)

5. Queries

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. The user making a query defines the register (PNR or MNR), 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 illustrated in Figure 3. Plans exist to apply the functionality of the developed application in a public web service that would enable queries for the PNR and present the results as lists and on maps on different scales.

5.1 Selecting Register

The user interface is similar for both the PNR and MNR queries. In MNR queries the *product* becomes a search key among others.

5.2 Selecting Names

The user may apply combinations of various terms when searching the names. The *spatial* conditions are set by choosing municipalities, General Map Sheets and 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. Parts of spellings can be expressed using wild characters (* and ?).

5.3 Selecting Data Set

In this context the data set 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. Almost any data in the database can be included in one data set: the data in the three primary tables, or any background table data that can be derived from the primary table data fields. Once defined, a data set can be stored for future use.

5.4 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.).

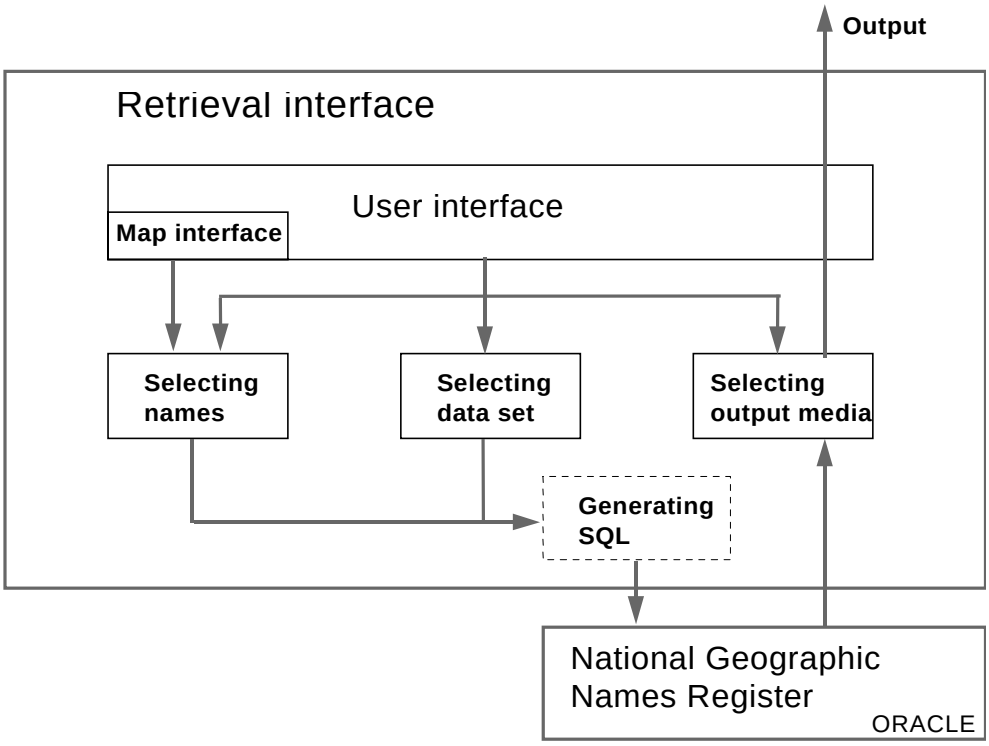

Figure 3. A register query

6. Names in the register

At present the Place Name Register and the Map Name Register 1:20,000 cover the whole country. The coverage of the MNR 1:100,000 is about 90 %, while the production of smaller scale MNRs has not yet commenced.

The Place Name Register includes 802,085 geographic names.

Names in the Place Name Register *by language*:

Finnish 718,834
Swedish 74,752
North Saami 4,564
Inari Saami 3,788
Skolt Saami 147

Names in the Place Name Register *by feature type group*:

Natural features 487,171
- terrain features 339,976
  - elevated areas 95,909
  - depressed areas 3,257
  - swamps 68,731
  - forests 63,417
  - islands 56,749
  - capes 38,704
  - other terrain features 13,209
- water features 147,195
  - seas, lakes, ponds and parts of them 107,215
  - rivers, ditches and parts of them 35,982
  - rapids 3,657
  - other water features 341

Cultural features 314,914
- habited places 274,636
  - municipalities 568
  - villages, neighbourhoods and districts within cities 25,848
  - houses 248,177
  - other habited places 43
- other cultural features 40,278
  - fields 37,729
  - others 2,549

7. Fields of Application

Fields of application related to the National Geographic Names Register are, for example:

National and international standardisation of geographic names
The spelling of the names in the register has been checked by the Research Institute for the Languages of Finland.

Rationalisation of the NLS map (database) production
Organised data management, uniform principles and specialised tools for geographic names in 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, national, regional and international gazetteers. On-line gazetteers in the Internet.

Internet services
Karttapaikka (MapSite, http://www.kartta.nls.fi/) is a service providing national topographic maps (e.g. Basic Map 1:20,000) for public on-line use. At present a database of some 30,000 toponyms is implemented for place searches. Plans exist to integrate the incrementally generalised GNR data sets as a flexible data source for navigating and zooming. The service would also enable queries for all the 800,000 names using different combinations of spatial, attribute and spelling terms and present the results as a lists and distribution maps.

Automatic positioning and navigation

For example, services by communications operators for mobile devices provided with GPS.

International projects and databases

Contribution of approved geographic names for international toponymic and GIS databases. The ids and object history records facilitate updates.

Place name planning

Research

A large, nation-wide and homogenous database of accepted place names with feature types and object co-ordinates gives a new source of information for researchers such as onomasticians, historians and natural scientists. By combining different spatial, attribute and spelling terms they can find answers to existing problems and derive new topics of study.