United Nations E/CONF.98/133/Add.1

Economic and Social Council

Distr.: General

2 July 2007

Original: English

Ninth United Nations Conference on the Standardization of Geographical Names
New York, 21 - 30 August 2007
Item 12 of the provisional agenda*
Toponymic data files

Geographic Names Register of Finland

Submitted by Finland **

---

* E/CONF.98/1.

** Prepared by Teemu Leskinen, Finland.

Geographic Names Register of Finland

The Finnish National Land Survey's (NLS) *Geographic Names Register* (GNR) is a toponymic-cartographic database serving the needs of geographic names standardisation and information service together with national GI (Geographic Information) and map data production. The main objective in the GNR design was the management of multilingual geographic names in a multi-scale and multi-product GI-environment.

Topographic Database of the National Land Survey

The principal data source of the Geographic Names Register is the National Land Survey's *Topographic Database* (TDB). The database includes, among other geographic information, over 800,000 natural and cultural feature names presented in the Basic Map 1:20,000. The spelling and location of the names have been checked and approved by the Finnish, Swedish and Saami experts of the *Research Institute for the Languages of Finland* (RILF).

In the present Topographic Database the geographic names are not attached to the GI-objects as attributes, but stored as separate map names. A map name record includes all the data related to the named feature (object co-ordinates, feature type), the name itself (spelling, language) and the cartographic occurrence of the name (placement, direction, font, size etc.). Parallel names and repeated names of e.g. rivers and large areas are not id-connected in the TDB.

Names in the Basic Map 1:20,000 probably comprise about a half of all geographic names in use. Table 1 shows the amounts of names by language and feature type. Among the 800,000 names there are about 300,000 different name forms (spellings) of which as many as 230,000 – almost 30 % of all names – are unique.

NAMES BY LANGUAGE	
Finnish	722,058
Swedish	74,820
North Saami	5,822
Inari Saami	4,258
Skolt Saami	153
NAMES BY FEATURE TYPE	
Natural features	
Terrain features	
Elevated areas	96,764
Depressed areas	3,427
Swamps	68,977
Forests	65,264
Islands	57,035
Capes	38,995
Other terrain features	14,452
Total of terrain features	*344,914*
Water features	
Seas, lakes, ponds and parts of them	108,606
Rivers, streams and parts of them	36,802
Rapids	3,935
Other water features	379
Total of water features	*149,722*
Total of natural features	***494,636***
Cultural features	
Habited places	
Municipalities (incl. names in several languages)	532
Villages, neighbourhoods and districts within cities	25,993
Houses/Farms	246,816
Other habited places	99
Total of habited places	*273,440*
Other cultural features	
Fields	36,267
Others	2,768
Total of other cultural features	*39,035*
Total of cultural features	***312,475***
TOTAL OF NAMES	***807,111***

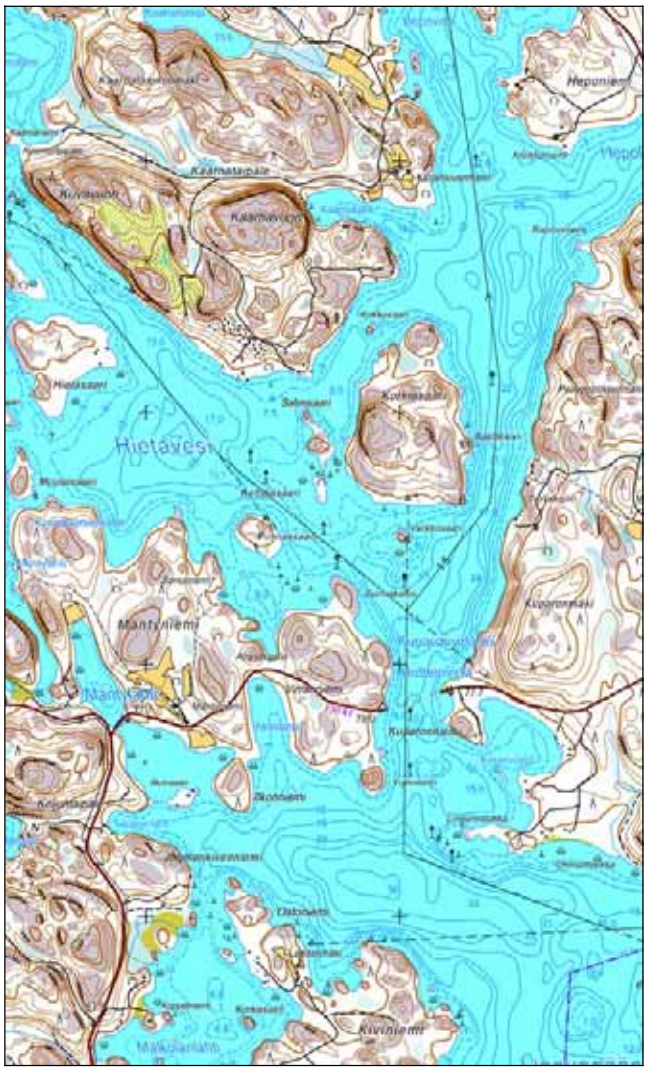

Table 1. Names in the Basic Map 1:20,000 by language and feature type.

Geographic Names Register

In the Geographic Names Register the names data are structured differently from the TDB. The GNR comprises the *Place Name Register* (PNR) and the *Map Name Register* (MNR) integrated as a single database. The PNR is scale-less, contains no cartographic data, but data on place names' feature type, feature location (centre point, mouth for a river), approved spelling (unabbreviated, small and capital letters used correctly) and language (Finnish, Swedish, North Saami, Inari Saami, Skolt Saami).

The MNR includes the product- and name-wise cartographic parameters (text placement, direction, bending, font and size etc.) for the selected names of the PNR. The names (character strings) themselves are not stored in the MNR unless they are rendered in a map differently from the name form in the PNR, e.g., divided in several lines. For a PNR name to be represented in capitals in a given product there is a flag stored in the MNR record indicating that the characters shall be turned to capitals during the plotting.

GNR data model: Places, Place names and Map names

The data model of the GNR is illustrated in Figures 1 and 2. The three elementary objects (tables) in the database are *Place*, *Place name* and *Map name*. Places and Place names build up the PNR. The PNR is integrated as the core of the MNR: the PNR together with Map names form the MNR. The Map names are arranged by *product* – every Map name instance belongs to exactly one product. Places, Place names and Map names are id-connected and all information on these objects is available for the other objects by id-link chains between them. A Place has one or more Place names and a Place name may have 0-N cartographic Map name occurrences in one or several products. For example, the longest river in Finland, *Kemijoki*, has one Place entry and one Place name entry in the database. On the other hand, as a Map name, Kemijoki occurs in every map scale and cartographic product, and typically several times in one product (32 times in the Basic Map series 1:20,000), while a little pond, say, *Kalattamalampi*, has a Map name occurrence just once, in the Basic Map.

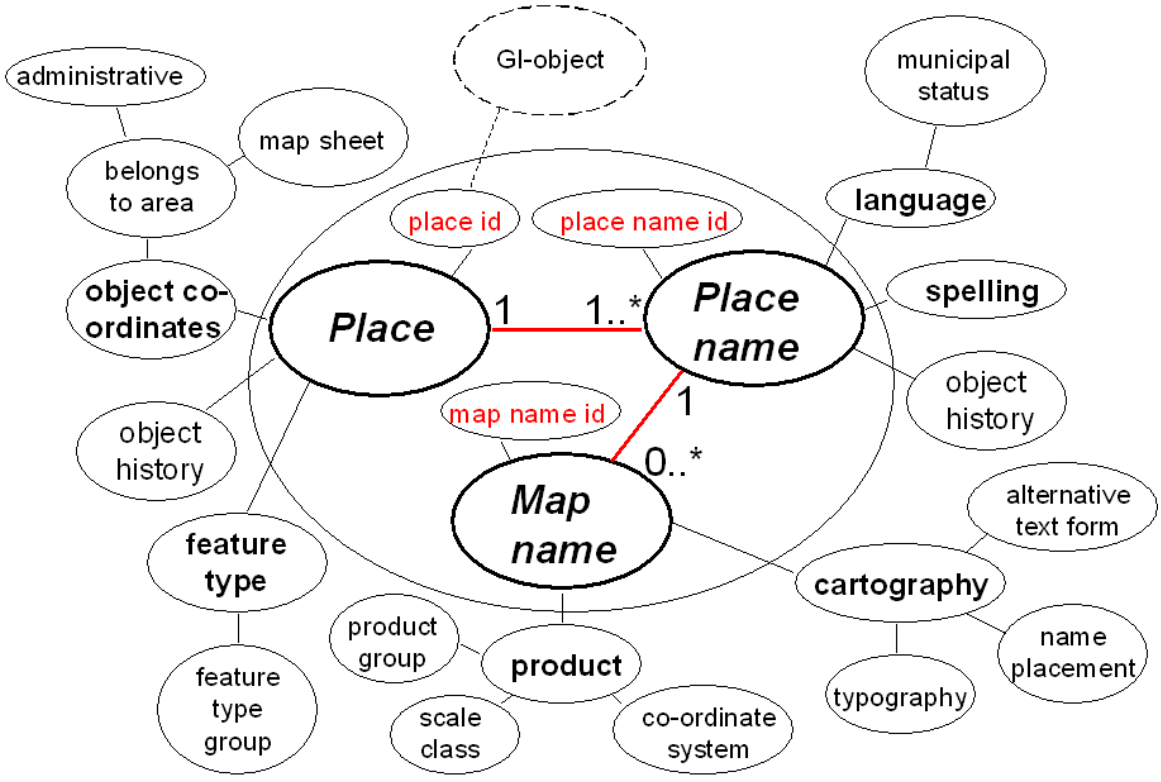

Figure 1. Data model of the Geographic Names Register. A place has one or more names that may occur one or several times in different cartographic products.

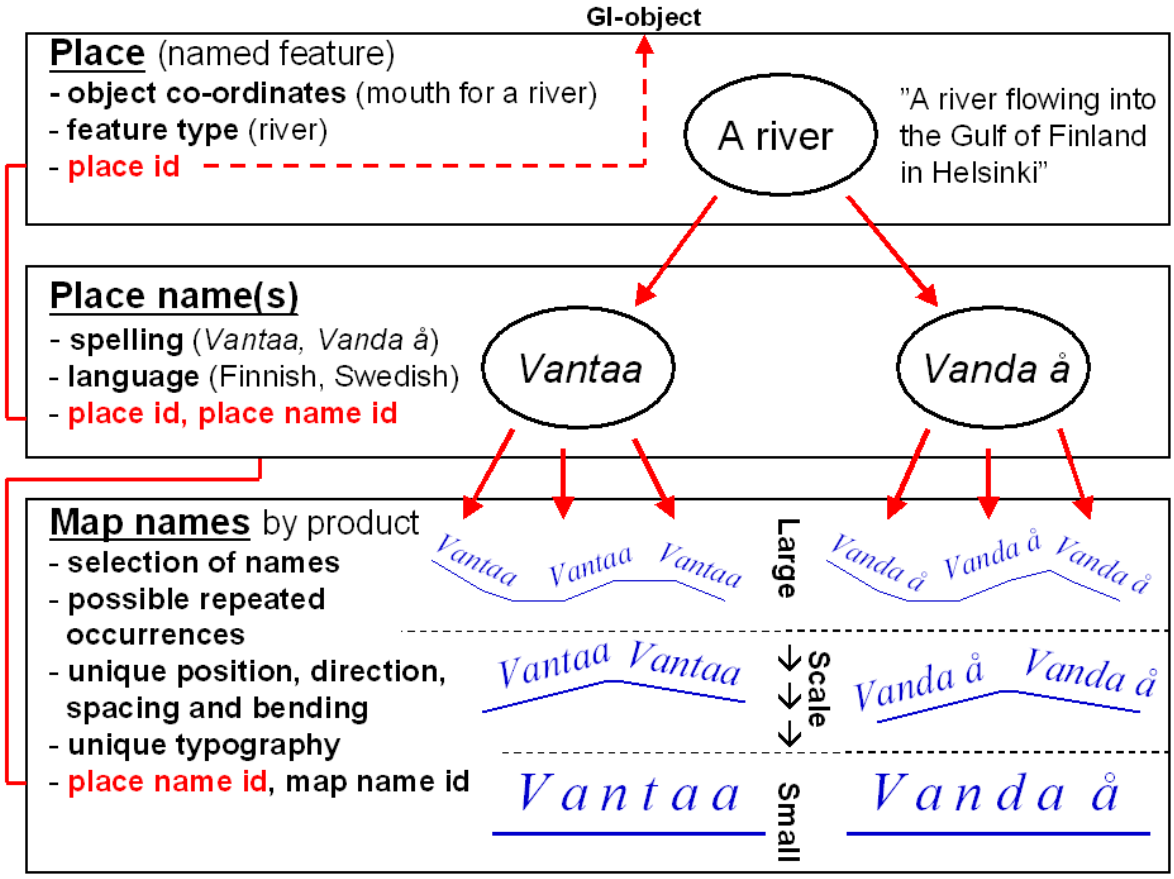

Figure 2. Example of the realisation of the data model, river *Vantaa*/*Vanda å*.

GNR data

The Place Name Register tables *Place* and *Place name* consist of the following data:

Table Place

- Unique external place id as a link to Place name(s) and as a possible link to external GI-object
- Geographic object co-ordinates; centre point (mouth for a river)
- Feature type
  - Feature type groups available via code tables
- The municipality in which the object belongs to
  - Information generated automatically
  - Larger hierarchical administrative units available via code tables
- Map sheet numbers the object belongs to (three different map sheet index systems)
  - Information generated automatically

Table Place name

- Unique external place name id
- Place id as a link to the Place
- Proper spelling of the name (accepted by the RILF, unabbreviated, upper and lower case)
  - North European 8-bit character set ISO 8859-10, modified with a few extra letters used in Skolt Saami
- Language of the name (Finnish, Swedish, North Saami, Inari Saami, Skolt Saami)
  - The status of each language in the municipalities (official/unofficial, majority/minority) is given in code tables; by using this information one can follow the recommendations given in the *Toponymic guidelines for map editors and other editors, Finland* when deciding which of the possible parallel names to prefer

The database history of objects Place and Place name is stored in tables *Place history* and *Place name history*:

Tables Place (Place name) history

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

When 'deleted', a Place or Place name record is not actually removed from the database but moved to tables *Old Place* and *Old Place name*.

The Map Name Register specific database table *Map name* consists of the following data:

Table Map name

- Unique external map name id
- Place name id as a link to the Place name (and to the Place)
- Position of the text (lower left corner) in the product coordinate system
- Alternative text (product-dependent spelling which differs from the spelling in table Place name; for example, a name divided in two lines on the map forms two records in this table, both sharing the same place name id)
- Text font code
- Text size (graphic size, mm/100)
- Text colour code
- Letter tilt angle
- Capitals flag (whether the PNR name is written in 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
- Bending (up to 32 pairs of relative co-ordinates conducting curved texts)

A Map Name is connected to a product in table *Map name product* and the products are introduced in table *Product*:

Table Map name product

- Map name id
- Product code

Table Product

- Product code
- Product name
- Co-ordinate system
- Map scale class

Beside the tables mentioned, there are several other tables in the database, mostly for managing and explaining the codes used in the database. An example of the GNR data contents (river *Ivalojoki*) is given in Table 2.

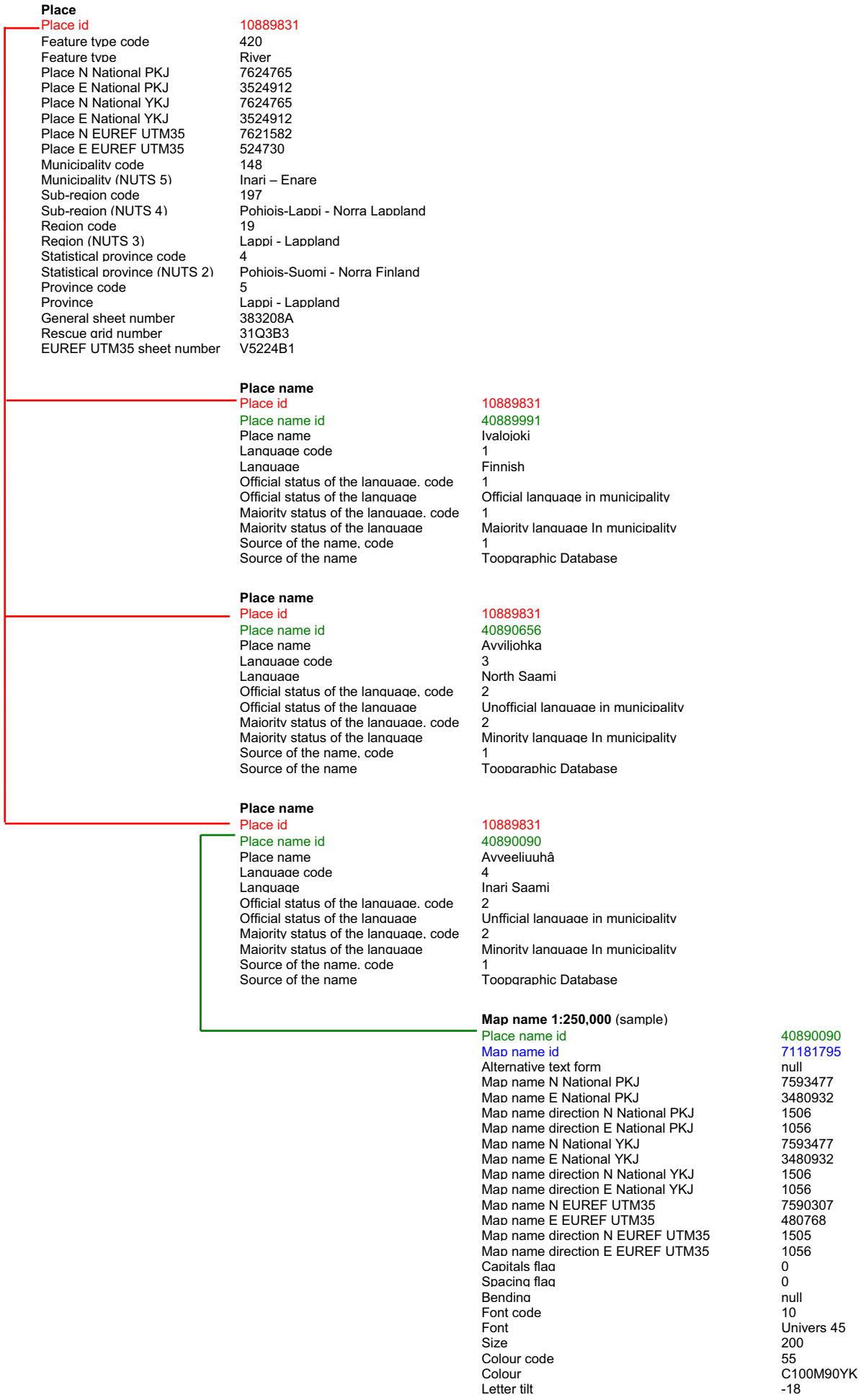

Table 2. Example of GNR data contents: river *Ivalojoki* (Finnish) = *Avviljohka* (North Saami) = *Avveeljuuhâ* (Inari Saami).

Maintenance of the GNR

At the moment the GNR data are maintained in two processes. The source data of the PNR and MNR 1:20,000 are managed as a part of normal Topographic Database updating and Basic Map compiling processes. The continuous updating of the GNR is based on transferring the TDB map names' change information to the GNR. Since the data model of the GNR is not (yet) implemented in the Topographic Database, an application software is necessary to maintain the GNR data model, i.e. to create and update the place id and place name id links between Places, Place names and Map names (Figure 3). In the future the GNR data model and database shall be maintained directly in the TDB production. Upon a decision to include names as attributes of selected TDB GI-objects, the names need not be directly attached to the objects: by using the place id links between the GNR Places and TDB GI-objects the names need be stored only once, in the GNR (Figures 1 and 2).

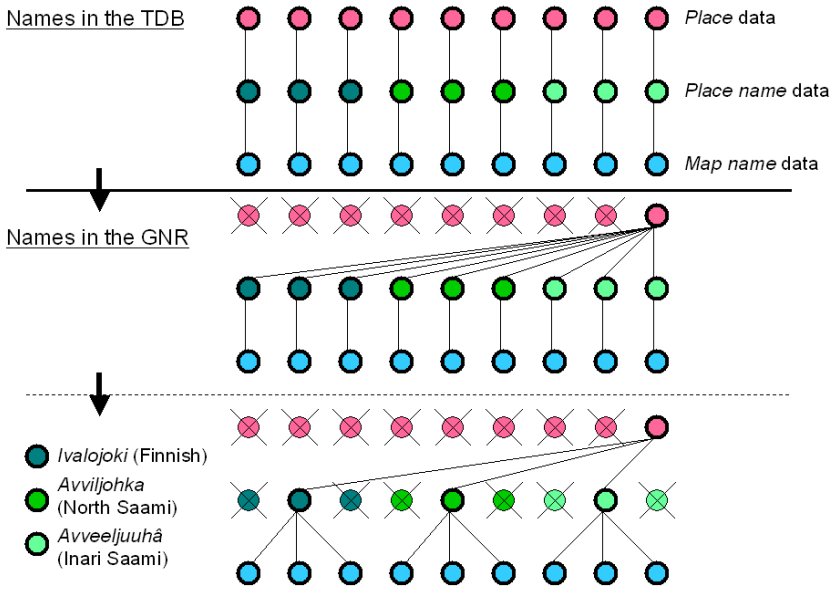

Figure 3. Creating and updating the id-links between Places, Place names and Map names.

The smaller scale Map names are compiled in integration with other GI-data production of respective scales. Besides the scale 1:20,000, the MNR includes nation-wide Map name representations in 1:100,000 and 1:250,000 (completed, Figure 4) as well as 1:500,000, 1:1 million, 1:2 million and 1:4.5 million (to come). The main tasks in processing are the selection of names, name placement and other cartographic editing. The selection is incremental and computer-aided. The Map names in 1:100,000 are selected from the Map names in 1:20,000, the source of the 1:250,000 Map names are the Map names in 1:100,000, and so on. In order to give a starting point for selection a parameterised automatic process is provided, based on the combination of source product names' feature type and text size. Naturally, in the following interactive selection phase, several other criteria must be taken into consideration. As a starting point for interactive cartographic editing default typography is generated automatically for the names, also based on the combination of names' feature type and text size in the source product.

Figure 4. Extract from the Map Name Register 1:250,000 – places with Finnish, North Saami, Inari Saami and Skolt Saami names in Inari municipality.

GNR products and fields of application of names data

The present Place Name Register and Map Name Register standard customer data sets include

- Place Name Register (PNR, about 800,000 names)
- Place Name Register 1:100,000 (PNR100, about 200,000 PNR names, which are selected in the map scale 1:100,000)
- Place Name Register 1:250,000 (PNR250, about 65,000 PNR names, which are selected in the map scale 1:250,000)

- Map Name Register 1:20,000 (MNR20, about 800,000 names)
- Map Name Register 1:100,000 (MNR100, about 200,000 names)
- Map Name Register 1:250,000 (MNR250, about 65,000 names)

Examples of fields of application regarding the GNR and the names data are

- National and international names standardisation and information service
- Rationalisation of the NLS map (database) production
- Place name and map name customer data sets
- Digital and traditional gazetteers
- Internet place name and map services (like the NLS' MapSite, http://www.karttapaikka.fi → English)
- Automatic positioning and navigation
- International GI-projects and databases
- Place name planning
- Ontology
- Different kind of research.

In research, a large, nation-wide database of accepted names with feature types and object co-ordinates has given a new source of information for researchers, as onomasticians, historians and archaeologists. By analysing names' meanings and combining their spatial, feature type, language and spelling characteristics they may find answers to existing questions and derive new hypotheses and objects of study. Since 2006, two doctor's theses, both based on the analysis of the PNR data, have been defended at the University of Helsinki.