A GEOGRAPHIC NAMES REGISTER AS A PART OF THE SPATIAL DATA INFRASTRUCTURE

Teemu Leskinen
National Land Survey of Finland
Development Centre
P.O.Box 84 (Opastinsilta 12 C)
FI-00521 Helsinki
Finland
teemu.leskinen@nls.fi

Introduction

Geographic names are rather difficult to model in geographic information systems (GIS) and as a part of the spatial data infrastructure (SDI). In some cases names can be applied as attributes of appropriately modelled spatial objects but often the identity, definition, classification, geometry and other attributes of these objects do not correspond with respective named features in the real world. Besides, commonly named features such as elevations, depressions, islands, coastal land formations, stretches of water bodies or even populated places are hardly ever modelled as spatial objects in GISs while they are important places referred to by their names in any type of communication.

Partly due to the modelling problems, partly because geographic names are often subject to (national) standardisation, thus maintained and/or supervised centrally by e.g. a national names authority, the names data are typically stored and maintained in separate and/or integrated names repositories and databases rather than directly attached as attributes to spatial objects modelled in e.g. topographic datasets.

The *Geographic Names Register of Finland* (GNR) developed and maintained by the *National Land Survey* (NLS) is one of the elements included in the national and international geographic names and spatial data infrastructure. As the national geographic names repository and database, the GNR serves as the reference dataset for the standardisation of Finland's geographic names as well as the official names data source for interoperability within the SDI (persistent GNR named feature identifiers as the interoperability component), and for different kinds of information services and applications including the varied spatial data and map production in the NLS.

Finland – a multilingual country

Finland is a multilingual country. Finnish and Swedish are the official state languages, and three indigenous Saami languages – Inari Saami, North Saami and Skolt Saami – have an official minority language status in the four northernmost municipalities of the Finnish Lapland. Finnish, Swedish and the Saami languages are written in variants of the Latin/Roman alphabet and are all well established in Finland's nomenclature.

The multilingual areas are illustrated in Figure 1. The basic administrative and linguistic unit in Finland is a municipality. According to the Language Act, a municipality is considered bilingual (Finnish–Swedish) if the speakers of the minority language constitute at least 8% of its population, or, alternatively, if they amount to more than 3 000.

In 2009, there were 348 municipalities of which 291 were monolingually Finnish, 19 monolingually Swedish, 19 bilingual with a Finnish majority and a Swedish minority, and 15 bilingual with a Swedish majority and a Finnish minority. In the municipalities of *Enontekiö*, *Sodankylä* and *Utsjoki* North Saami and Finnish are spoken, and in *Inari* municipality all the three Saami languages are used, together with Finnish.

The *Toponymic guidelines for map editors and other editors, Finland* [Toponymic guidelines, 2009] provides more information on the languages and their status in different parts of Finland.
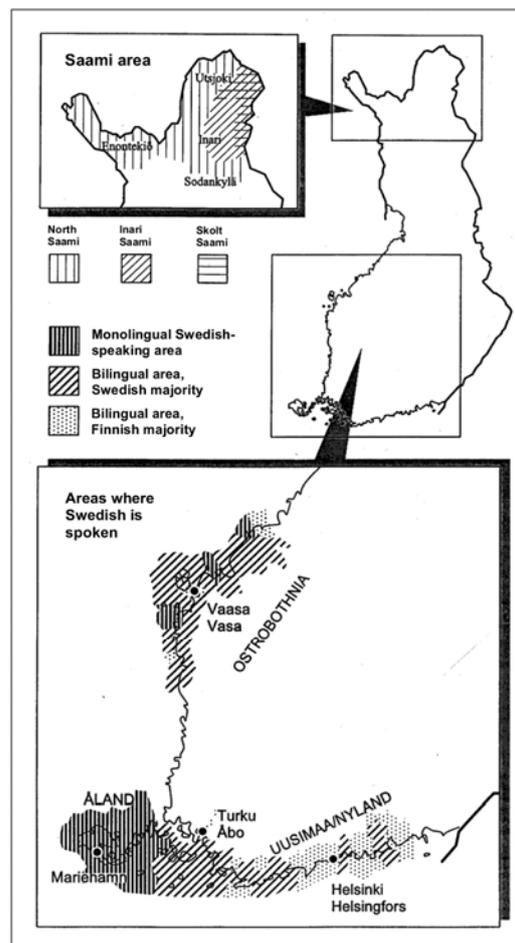

Figure 1. Multilingual areas in Finland.
© NLS Finland

Principles for presenting names in national topographic maps

The National Land Survey produces two national topographic map series based on the NLS *Topographic Database* (TDB): the *Basic Map 1:25 000* and the *Topographic Map 1:50 000*. In addition, national small scale map databases – generalised from the TDB data – are being produced at scales 1:100 000, 1:250 000, 1:500 000, 1:1 million, 1:2 million and 1:4.5 million.

Only names in actual use are presented in the national topographic maps. Names in bilingual (Finnish–Swedish) municipalities occur in both languages, the name in the majority language mentioned first (Figures 2 and 3). In Lapland, all existing Saami names of a given place are presented, in a standard language order, together with the (possible) Finnish name (Figure 4).

Figure 2. *Helsinki* is a bilingual city with a Finnish-speaking majority (Finnish names above Swedish names). © NLS Finland
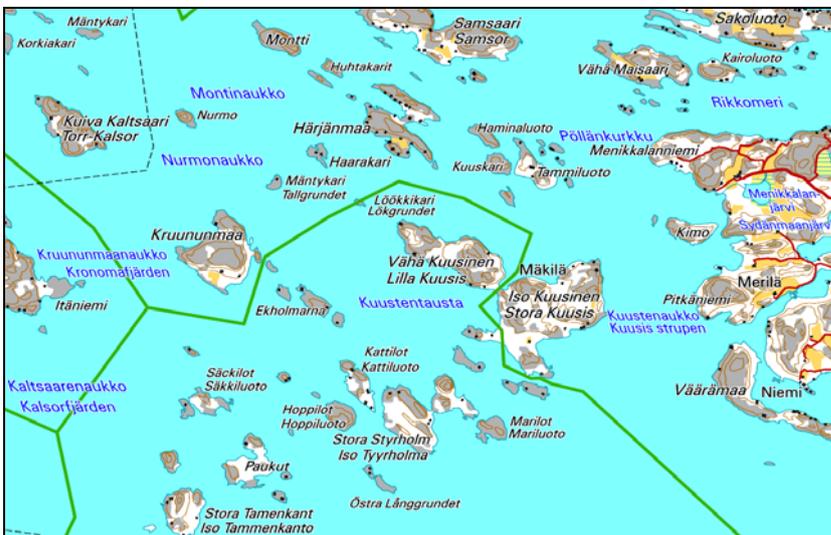
Figure 3. *Rymättylä* (above) is a monolingual Finnish-speaking municipality (Finnish names above Swedish names) and *Nagu* (below) a bilingual municipality with a Swedish-speaking majority (Swedish names above Finnish names). © NLS Finland
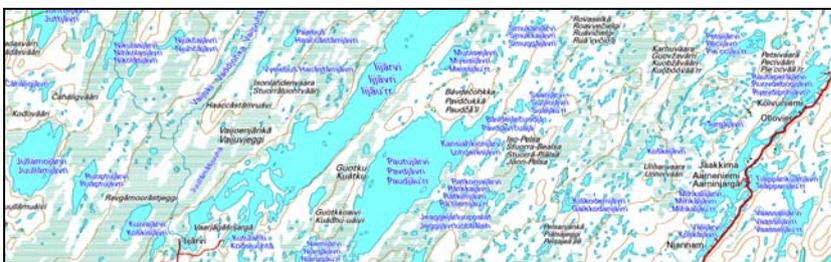
Figure 4. In *Inari* municipality places may have up to four names, in Finnish, North Saami, Inari Saami and Skolt Saami. © NLS Finland.

Names in the Topographic Database

The data source of the national topographic maps is the National Land Survey's Topographic Database. The database includes, among other detailed geographic information, all the natural and cultural feature names presented in the Basic Map 1:25 000. The spelling of the names have been checked and approved by the Finnish, Swedish and Saami names experts of the *Research Institute for the Languages of Finland*.

The Topographic Database includes about 800 000 names of natural and cultural/man-made features (Table 1), which probably comprise about one half of all the names in use. Among the names there are about 300 000 different name forms (spellings) of which 230 000 – almost 30% of all names – are unique.

About 42% of the toponyms are names of terrain features, 18% names of hydrographic features, 34% names of settlements and houses, and the remaining 6% are names of other cultural and man-made features. Roughly, 89% are Finnish, 9% Swedish, and 2% Saami names.

In the Topographic Database the names are not attached to the GI-objects as attributes but stored as separate map names. A TDB map name record includes all the data related to the named feature, the name itself, and the cartographic occurrence of the name. Parallel names are not id-connected in the TDB, nor repeated names (of e.g. rivers).

The cartographic background of the TDB is one of the reasons why it was necessary to develop a new, structured model for the geographic names data collected in the TDB.

NAMES BY LANGUAGE	
Finnish	722 161
Swedish	74 882
North Saami	6 248
Inari Saami	4 622
Skolt Saami	654
NAMES BY FEATURE TYPE	
Natural features	
Terrain features	
Elevated areas	97 054
Depressed areas	3 482
Wetlands	68 750
Forests	66 376
Islands	57 142
Capes	39 110
Other terrain features	15 304
Total of terrain features	*347 218*
Hydrographic features	
Sea and lake areas, ponds and parts of them	109 347
Rivers, streams and parts of them	37 008
Rapids	3 975
Other hydrographic features	424
Total of hydrographic features	*150 754*
Total of natural features	**497 972**
Cultural features	
Populated places	
Municipalities/Towns (in different languages)	451
Districts, villages and neighbourhoods	26 185
Houses/Farms	245 787
Other populated places	161
Total of populated places	*272 584*
Other cultural features	
Cultivated fields	35 196
Others	2 815
Total of other cultural features	*38 011*
Total of cultural features	**310 595**
TOTAL OF NAMES	**808 567**

Table 1. Names in the Basic Map 1:25 000 by language and feature type.

The Geographic Names Register

The multilingual and multi-names circumstances in Finland as well as the multi-product and multi-scale production environment of the National Land Survey have guided the design and development of the GNR. In Finland a specific geographic feature may have different parallel names in one or several languages. In bilingual (Finnish–Swedish) areas, a feature typically has two names, and in Lapland's Saami language areas, where Finnish, Inari Saami, North Saami and Skolt Saami are spoken, a feature may have up to four names, each of which being provided by national geographic names services and represented in the national topographic maps.

The Geographic Names Register comprises the *Place Name Register* (PNR) and the *Map Name Register* (MNR) integrated into a single database, in which every item of information is stored only once. The PNR is not based on any map scale and includes no cartographic information but contains data on the type and location of the named features and the approved spelling (unabbreviated, correct capitalisation) and the language and language status information of the names. The MNR is organised by map series and includes the product- and name-wise cartographic representation parameters (text placement, direction, bending, capitalisation, typeface, font, size etc.) for the selected names in 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, a flag stored in the MNR record tells 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 (Figures 5 and 6) is feature oriented. It includes objects such as *Place*, representing a named feature, *Place name*, representing a geographic name, and *Map name*, representing a cartographic occurrence of a geographic name. Places and Place names build up the PNR. The PNR is integrated as the core of the Map Name Register: the PNR together with Map names form the MNR. A Place has one or more Place names that for one may have zero, one or several Map name occurrences in different cartographic products. As an example, the longest river in Finland, *Kemijoki*, has one Place entry and one Place name entry in the database. As a Map name, Kemijoki occurs in every map scale and cartographic product, and typically several times in one product (over 30 times in the Basic Map 1:25 000 series), while a little stream, say, *Myllyoja*, has merely one Map name occurrence, in the Basic Map.

Places, Place names and Map names are interconnected and are provided with external persistent unique identifiers to facilitate efficient data management and to aim at well-functioning organisational, national and international interoperability including data exchange by change-only updates.
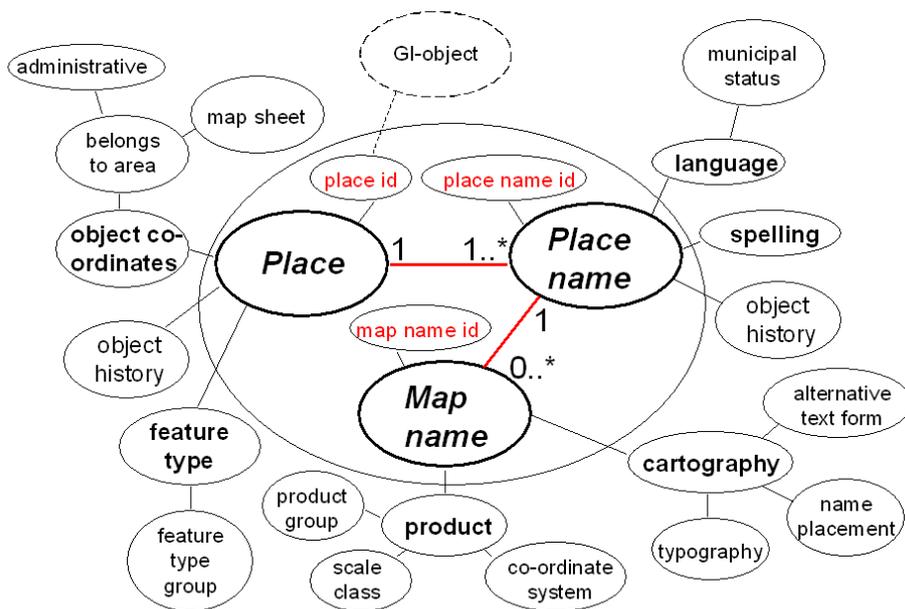
Figure 5. Conceptual model of the Geographic Names Register.
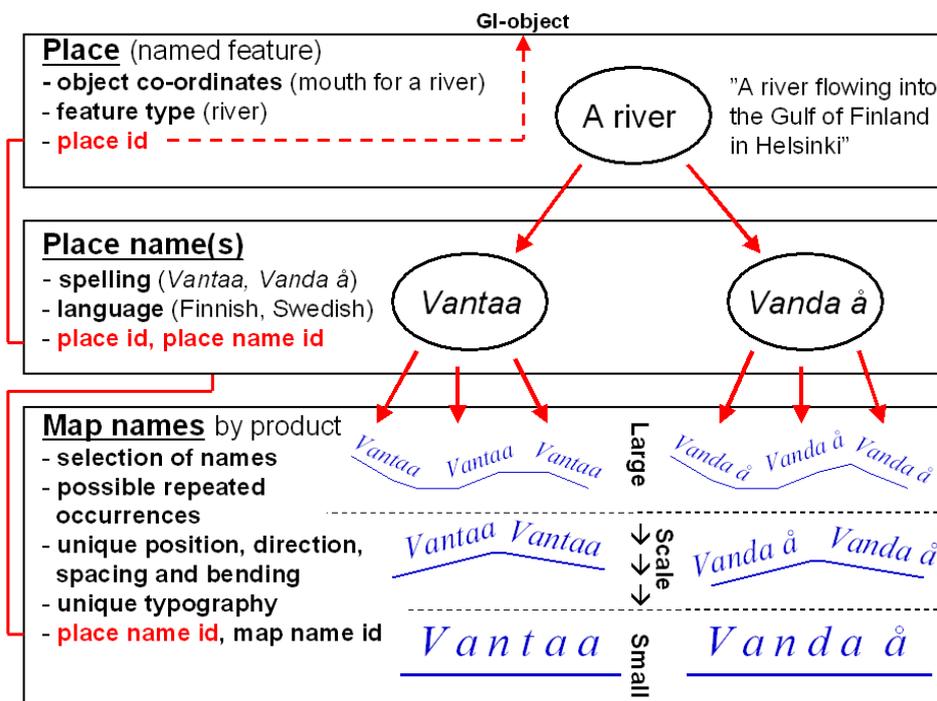
Figure 6. Realisation of the GNR model, river *Vantaa*/*Vanda å*.

Maintenance of the GNR

The maintenance of the Place Name Register and the Map Name Register 1:25 000 is integrated with the NLS Topographic Database and Basic Map production while the smaller scale Map Name Registers (at present 1:100 000 and 1:250 000) are integrated with the NLS spatial and map data production of respective scales. The integration is based on persistent identifiers and automatic and semi-automatic change-only updates.

The Place Name Register and Map Name Register 1:25 000 both include approximately 800 000, the Map Name Register 1:100 000 about 200 000 and the Map Name Register 1:250 000 some 65 000 records.

GNR Web Feature Service (WFS)

The GNR data are disseminated e.g. through NLS standard *Web Feature Service* (WFS) interfaces [GNR WFS, 2009]. The names data WFS products include two GML profile schemas for the Place Name Register – serving a little different use case purposes – and one schema for the Map Name Register. The Place Name Register GML profiles are a feature oriented schema (Figure 7, Place is the root and Place names are attributes of Place) and a name oriented schema (Figure 8, Place name is the root and Place and possible parallel names are "attributes" of Place name). The Map Name Register GML profile (Figure 9) has Map name as the root and (selected) corresponding Place and Place name information as "attributes". The schemas also include the enumerations and explanations for the codes used in the service.
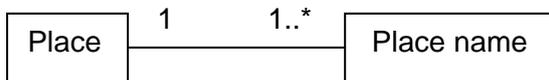
Figure 7. Conceptual model, the <u>feature oriented</u> Place Name Register WFS profile.
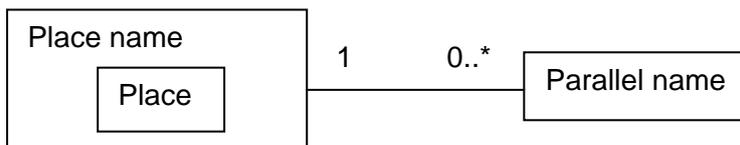
Figure 8. Conceptual model, the <u>name oriented</u> Place Name Register WFS profile.
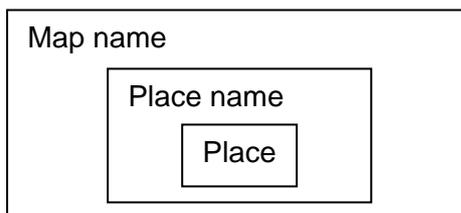
Figure 9. Conceptual model, the Map Name Register WFS profile.

GNR Web Feature Service UML and XML schemas

The following links to the NLS web documentation provide up-to-date UML and XML schemas of the GNR Web Feature Service products (for English translations, see Appendix).

The <u>feature oriented</u> Place Name Register WFS:

UML schema:
http://xml.nls.fi/Nimisto/Asiakasdokumentaatio/WFS/UMLmallit/Paikat/PaikannimirekisterinPaikat.jpg
XML schema:
http://xml.nls.fi/Nimisto/Asiakasdokumentaatio/WFS/Kuvaus/describeFeatureType_paikka_vastaus.xml

The <u>name oriented</u> Place Name Register WFS:

UML schema:
http://xml.nls.fi/Nimisto/Asiakasdokumentaatio/WFS/UMLmallit/Paikannimet/PaikannimirekisterinPaikannimet.jpg
XML schema:
http://xml.nls.fi/Nimisto/Asiakasdokumentaatio/WFS/Kuvaus/describeFeatureType_paikannimi_vastaus.xml

The Map Name Register WFS:

UML schema:
http://xml.nls.fi/Nimisto/Asiakasdokumentaatio/WFS/UMLmallit/Karttanimet/KarttanimirekisterinKarttanimet.jpg
XML schema:
http://xml.nls.fi/Nimisto/Asiakasdokumentaatio/WFS/Kuvaus/describeFeatureType_karttanimi_vastaus.xml

GNR Web Feature Service queries

A Place Name Register WFS query can be filtered e.g. by:
- Place and Place name identifiers;
- Location (co-ordinate based, or indirect by administrative unit or map sheet);
- Feature type or feature type group;
- Spelling;
- Language and municipal language status information;
- Objects' database lifespan information (timestamps for creation, modification and retirement) supporting change-only updates;
- 'Relevance at Scale' indicator depicting the size/importance of the named feature by means of the smallest map scale in which the name(s) of the feature have been rendered as Map name(s) in national topographic maps.

A Map Name Register WFS query can be filtered e.g. by:
- Product (e.g. topographic map 1:25 000, 1:100 000 or 1:250 000);
- Location (e.g. a bounding box or a polygon).

Fields of application of GNR data and services

Examples of fields of application for the GNR data and Web Feature Services include:
- Standardisation; clear and consistent use of accurate geographic names in any type of communication;
- Interoperability (typically based on identifier links); national and international names and spatial data infrastructures and related projects (e.g. the Finnish SDI, *EuroGeoNames* and *INSPIRE*);
- Search; finding named places and geographic names by using their attributes (location, feature type, spelling, language) as search criteria; geocoding; gazetteer services; map browsing applications; automatic positioning and navigation;
- Visualisation; map production; geographic names as information layer in viewing services;
- Research of different kind (onomastics, historical research, natural science, archaeology, genealogy etc.);
- Semantic web applications; ontology;
- Cultural heritage promotion; safeguarding of the cultural heritage related to (traditional) geographic names and their carriers.

Besides the NLS Web Feature Service and pre-defined customer datasets GNR data are available from EuroGeoNames [EGN, 2009] Central Service (WFS), and will be available from an NLS INSPIRE Web Feature Service after the Data Product Specification/Implementing Rule for the INSPIRE Directive [INSPIRE, 2007] Annex I Theme "Geographical Names" has been finalised.

References

[EGN, 2009] EuroGeographics (2009). *EuroGeoNames*;
http://www.eurogeonames.com/.

[GNR WFS, 2009] National Land Survey of Finland (2009). *Geographical Names Register Web Feature Service (WFS) user documentation*. In Finnish;
http://xml.nls.fi/Nimisto/Asiakasdokumentaatio/WFS/.

[INSPIRE, 2007] European Union (2007). *Directive 2007/2/EC of the European Parliament and the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*;
http://inspire.jrc.ec.europa.eu/.

[Toponymic guidelines, 2009] Research Institute for the Languages of Finland & National Land Survey of Finland (2009). *Toponymic guidelines for map editors and other editors, Finland;*
http://unstats.un.org/unsd/geoinfo/guidelineslink.htm > Finland.

Appendix

Geographic Names Register WFS – English translations for the UML and XML schemas

<u>Paikka – Place</u>
paikkaID – Place id
paikkatyyppiKoodi – Feature type, code
paikkatyyppiryhmaKoodi – Feature type group (e.g. natural features, cultural features), code
paikkatyyppialaryhmaKoodi – Feature type subgroup (e.g. terrain features, hydrographic features, administrative units, populated places), code
paikkaSijainti – Place location (point co-ordinates)
tm35Fin7Koodi – Map sheet, ETRS-TM35FIN index system, code
ylj7Koodi – Map sheet, national KKJ index system, code
pp6Koodi – National rescue grid square, code
kuntaKoodi – Municipality, code
maakuntaKoodi – Region, code
laaniKoodi – Province, code
seutukuntaKoodi – Statistical region, code
suuralueKoodi – Statistical province, code
mittakaavarelevanssiKoodi – 'Relevance at scale', code
paikkaLuontiAika – Timestamp for the creation of database object Place
paikkaMuutosAika – Timestamp for the last modification of database object Place
paikkaPoistoAika – Timestamp for the retirement of database object Place

<u>Paikannimi – Place name, Rinnakkaisnimi – Parallel name</u>
paikannimiID – Place name id
kirjoitusasu – Spelling
kieliKoodi – Language, code
kieliVirallisuusKoodi – Municipal status of the language (official/unofficial language), code
kieliEnemmistoKoodi – Municipal status of the language (majority/minority language), code
paikannimiLuontiAika – Timestamp for the creation of database object Place name
paikannimiMuutosAika – Timestamp for the last modification of database object Place name
paikannimiPoistoAika – Timestamp for the retirement of database object Place name

<u>Karttanimi – Map name (place name in map</u>)
karttanimiID – Map name id
karttanimiTeksti – Map name, text in map
karttanimiSijainti – Map name location (point co-ordinates)
karttanimiXP – Map name northing in product co-ordinate system
karttanimiYI – Map name easting in product co-ordinate system
karttanimiSuuntaDXP – Text direction, relative northing
karttanimiSuuntaDYI – Text direction, relative easting
karttanimiTaivutus – Text bending, pairs of relative co-ordinates conducting curved texts
kirjasintyyppiKoodi – Text typeface/font, code
kirjasinkoko – Text size (graphic size, in mm/100)
kirasinvariKoodi – Text colour, code
kirjasinKallistuskulma – Letter tilt angle (in degrees)
harvennusLippu – Spacing flag (whether the text direction parameters (dx,dy) also define the length of the text)
mtkLadontaKoodi – 'Typesetting code' in the Topographic Database (original data source)
paikkaID – Place id
paikkatyyppiKoodi – Feature type, code
paikkaSijainti – Place location (point co-ordinates)
paikannimiID – Place name id
kirjoitusasu – Spelling
kieliKoodi – Language, code
kieliVirallisuusKoodi – Municipal status of the language (official/unofficial language), code
kieliEnemmistoKoodi – Municipal status of the language (majority/minority language), code
karttatuoteTunnus – Map series, code
mittakaavaluokkaKoodi – Map scale (e.g. 1:25 000, 1:100 000 or 1:250 000), code
koordinaattijarjestelmaKoodi – Coordinate system for interpreting karttanimiXP and karttanimiYI, code