GEGN.2/2019/55/CRP.55
18 March 2019

English

United Nations Group of Experts On Geographical Names
2019 session
New York, 29 April–3 May 2019
Item 15 of the provisional agenda*
Toponymic data files and gazetteers (data processing and tools, database management, data dissemination: products and services) (Working Group on Toponymic Data Files and Gazetteers)

Benefits of persistent unique identifiers of named places and place names

Submitted by Finland**

Summary:

Places have been given names to enable people to speak and write about them effectively. The purpose of a name is to identify a particular place in human communication. As long as the sender and recipient of a message use the same names for the same objects, the message serves its purpose and helps people to think and act correctly in each situation. In maps and gazetteers, names help places to be found.

For applications in information technology and geospatial information infrastructure, names or other attributes of places do not usually identify places precisely enough, for example, when different places have the same name. Therefore, the geographic names data models are often accompanied by unique identifiers for named places and/or place names in order to identify them unambiguously for technical purposes.

Persistent unique identifiers allow, for example:

- Integration of data resources; association of designated places and their standardized names with respective objects in other registers and geospatial datasets
- Solid information on the life cycle of places and place names, based on the database object versions generated whenever there is a change in an object
- Information on the life cycle of places and place names, based on change -only updates that include additions, changes and deletions of the objects during a given time period
- Dataset status at a given moment, based on information on th e life cycle
- Provision of information on a specific place or place name as linked data in both browsable and machine-readable formats.

The report provides information on the management of unique identifiers and on the life cycle of places and place names contained in the Geographic Names Register of the National Land Survey of Finland. The report also contains information on related benefits, applications and possibilities regarding the development of information services and the national geospatial information infra structure.

*GEGN.2/2019/1
**** Prepared by Teemu Leskinen (The National Land Survey of Finland)

1. The Geographic Names Register of the National Land Survey of Finland

The Geographic Names Register (GNR) of the National Land Survey (NLS) is the authoritative geographic names data repository of Finland and one of the elements included in the national and international geospatial data infrastructure. The GNR comprises the primary Place Name Register (PNR), containing information on more than 800 000 named places and their standardised names, and the Map Name Register (MNR), an operational dataset holding the cartographic attributes (placement, typography) related to the PNR names selected to be presented in the NLS cartographic products in nine scales ranging from the Basic map 1:25 000 to the General map 1:8 million.

1.1 GNR data model and unique identifiers

The GNR data model is feature-oriented and includes the objects Place, Place name and Map name. All objects are interconnected and provided with external persistent unique identifiers. A named feature (Place) has at least one name (Place name) – in Finland, commonly several names in different languages – that may have zero, one or multiple occurrences (Map names) in different cartographic products (Figures 1 and 2).

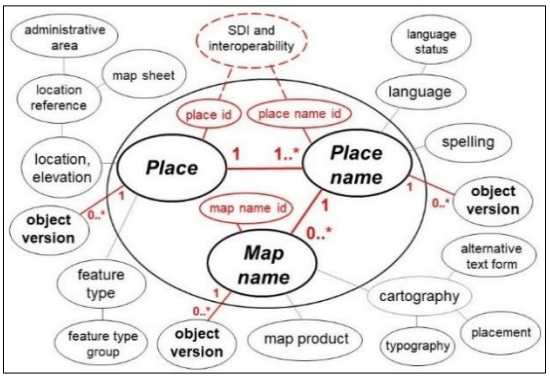

Figure 1. GNR data model.

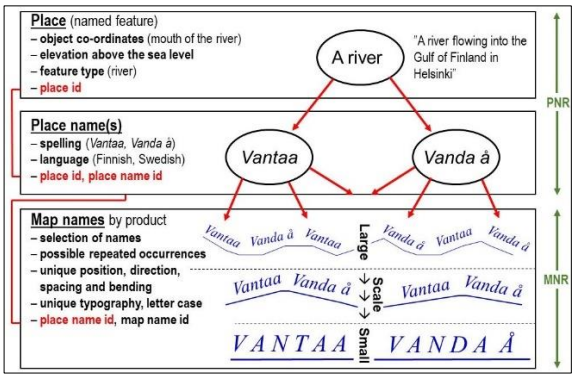

Figure 2. Realisation of the GNR data model.

1.2 Version management and life cycle information of GNR objects

In the GNR, a complete version management for Places, Place names and Map names is implemented, based on their identifiers, timestamps for the creation, modification and deletion of the versions and separate database tables for the modified object versions (Figure 3). Before any modification of a Place, Place name or Map name, the unaltered object version is inserted into the respective version table, i.e. Place version, Place name version or Map name version, and the version is identified by a sequential version number.

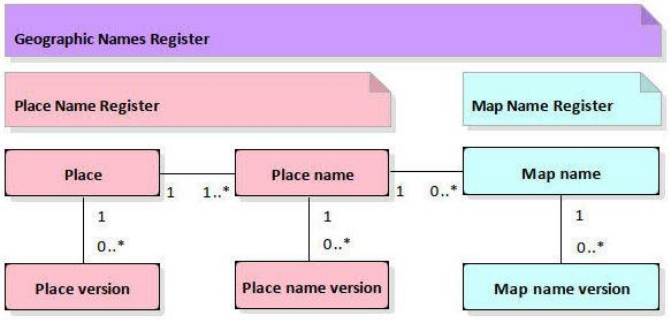

Figure 3. Version management of the GNR objects.

2. Dissemination of GNR data

The GNR datasets are available – globally free of charge – through the NLS Web Feature Service (WFS 1.1 and WFS 3.0 Beta), and as Geographic Markup Language (GML) data file products, the schemas of which correspond to the WFS schemas. In addition to the national schemas for the PNR and MNR, the WFS schema *Geographical names* sanctioned by the EU INSPIRE directive (Infrastructure for Spatial Information in Europe), as well as the corresponding European Location Framework (ELF) schema profile introduced by EuroGeographics, an organisation representing Europe's National Mapping Authorities, have been implemented for international use.

At present, the national GNR WFS 1.1 and GML data file products include two schemas for the PNR, *Places* and *Place names* – with equal data contents but different data structure, serving different use cases – and one schema for the MNR, *Map names*. In GNR WFS 3.0 (Beta), the PNR products are complemented by the product *Place names, simple*, an alternative for *Place names* for applications without a need for possible parallel names of a feature. The products are accompanied by explanatory code lists in Finnish, Swedish and English (Figure 4).

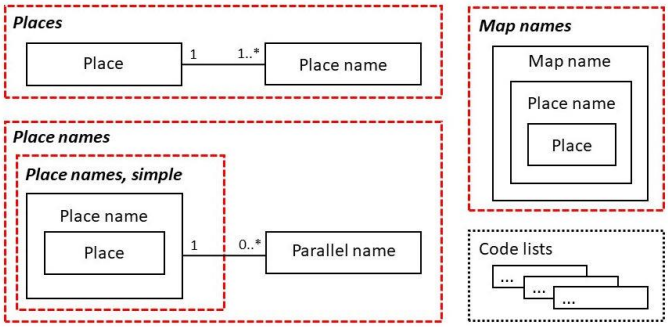

Figure 4. GNR Web Feature Service (WFS) and GML data file schema structures.

3. Applications on GNR unique identifiers and life cycle information

3.1 Plans for the integration of GRN with other data resources

In 2018, the Ministry of Agriculture and Forestry of Finland published a Report of spatial data policy, with the aim to oblige all actors in the public sector to see to the interoperability of geospatial data and access to them in a way that they can be used by the public authorities and companies as efficiently as possible. This will enable to improve the efficiency of administrative processes, reduce overlapping work, find cost savings in public administration and create conditions for innovation.

In the ongoing National Topographic Database (NTDB) reform programme, persistent unique identifiers of geospatial objects, together with standard web feature services, are declared as essential interoperability components in the integration of data resources. For ensuring the appropriate maintenance and proper use of standardised geographic names within the infrastructure, the GNR named places and place names are recommended to be linked with the NTDB when possible.

In the NTDB conceptual model, any object in any NTDB data theme (terrain, land use, hydrography, administrative units, transportation, buildings etc.) may have a GNR place id as an attribute, providing a direct link to a corresponding object in the GNR. Consequently, the names of these NTDB objects are maintained in the GNR only and available to them from GNR web feature services (a place id query) when needed. Similarly, the geometry and other attributes of the interconnected NTDB objects are available to corresponding GNR objects through NTDB web feature services whenever needed.

Alternatively, for NTDB objects with no corresponding Place stored in the GNR, a datatype Name is introduced in the NTDB conceptual model as an association for any NTDB object (Unified Modelling Language (UML) model, Figure 5).

Figure 5. Two alternative ways to model names for NTDB objects.

3.2 Plans for providing GNR change-only-updates and dataset status at a given moment

GNR persistent unique identifiers and life cycle information allow change-only-updates (COU) of the dataset. They also allow supplies that contain the dataset status at a given moment in the past. A COU supply may contain GNR data elements that have changed since the last refresh of the customer dataset or product, i.e. any additions, modifications or deletions of the elements during a given period.

Since the GNR objects Place, Place name and Map name have independent life cycles and version management, a change in an object does not affect or create a new version of other, related objects. However, GNR products described in section 3 combine information on different objects, and therefore an auxiliary, product-schema-element-based (e.g. GML Feature Member) life cycle management is needed. For example, the product *Places* combines data on both Places and Places names (Figure 4), and whenever there is a change in either of these objects, a product Feature Member event is recorded in the database. Initially, the events will be generated automatically for the entire dataset. In the WFS and GML data file product schemas these events are modelled as two attributes:

- Product event type (values: Creation, Modification or Deletion of the Feature Member)
- Product event time.

By combining these attributes in queries customers can update their datasets in different ways. In addition, the attributes allow to supply the dataset status at a given moment. In spring 2019 the management of product Feature Member events is under development and the event attributes are not yet implemented in GNR products.

3.3 Plans for making GNR places and place names available as linked data

Persistent unique object identifiers are necessary for making a geospatial dataset's contents available as linked (open) data (LOD). The NLS Finland has not yet published a GNR linked data service. However, in 2015, a prototype service was implemented to gain experiences of the technology (Eero Hietanen et al., XIII ISPRS Congress, 2016, Prague, Czechia) and plans exist to develop and publish an operational GNR linked data URI-service in the near future.