

Harmonizing National Topographic Data Using an Automated Quality Validation Process

Nils Mesterton ^{a, *}, Mari Isomäki ^b, Antti Jakobsson ^c, Joonas Jokela ^d

^a National Land Survey of Finland, nils.mesterton@nls.fi

^b National Land Survey of Finland, mari.isomaki@nls.fi

^c National Land Survey of Finland, antti.jakobsson@nls.fi

^d National Land Survey of Finland, joonas.jokela@nls.fi

* Corresponding author

Keywords: Spatial Data Quality, Spatial Data Harmonization, Standards, FME, SDI

Abstract:

The Finnish National Topographic Database (NTDB) is currently developed by the National Land Survey of Finland (NLS) together with municipalities and other governmental agencies. It will be a harmonized database for topographic data in Finland provided by municipalities, the NLS and other agencies. The NTDB has been divided into several themes, of which the buildings theme was the focus in the first stage of development. Data collection for the NTDB is performed by different municipalities and governmental organizations. Having many supplying organizations can lead to inconsistencies in spatial data. Without a robust quality process this could lead to a chaos. Fortunately data quality can be controlled with an automated data quality evaluation process. Reaching a better degree of harmonization across the database is one of the main goals of NTDB in the future, besides reducing the amount of overlapping work and making national topographic data more accessible to all potential users.

The aim of the NTDB spatial data management system architecture is to have a modular architecture. Therefore, the Data Quality Module named as QualityGuard can also be utilized in the National Geospatial Platform which will be a key component in the future Spatial Data Infrastructure of Finland. The National Geospatial Platform will include the NTDB data themes but also addresses, detailed plans and other land use information. FME was chosen as the implementation platform of the QualityGuard because it is robust and highly adaptable, allowing development of even the most complicated ETL workflows and spatial applications. This approach allows effortless communication with different applications via various types of interfaces, thus efficiently enabling the modularity requirement in all stages of development and integration.

The QualityGuard works in two modes: a) as a part of the import process to NTDB, and b) independently. Users can validate their data using the independent QualityGuard to find possible errors in their data and fix them. Once validated and the data is fixed, data producers can import their data using the import option. The users receive a data quality report containing statistics and a quality error dataset regarding their imported data, which can be inspected in any GIS software, e.g. overlaid on original features. Geographical locations of quality errors are displayed as points. Each error finding produces a row in the error dataset, containing information about the type and cause of the error as short descriptions.

Data quality evaluation is based on validating the conformance against data product specifications specified as quality rules. Three different ISO 19157 quality elements are utilized: format consistency, domain consistency and topological consistency. The quality rules have been defined in a co-operation with specialists in the field and the technical developing team. The definition work is based on the concept developed in the ESDIN project, quality specifications of INSPIRE, national topographic database quality specifications, national and international quality recommendations and standards, quality rules developed in European Location Framework (ELF) project and interviews of experts from National Land Survey of Finland and municipalities. In fact the NLS was one of the first agencies in the world who published a quality model for the digital topographic data in 1995.

Quality rules are currently documented in spreadsheet documents representing each theme. Each quality rule has been defined using RuleSpeak, a structured notation for expressing business rules. RuleSpeak provides a consistent structure for each definition. The rules are divided in general rules and feature-specific rules. General rules are relevant for all feature types of a specific theme, although exceptions can be defined.

A nation-wide, centralized automated spatial data quality process is one of the key elements in an effort towards achieving better harmonization of the NTDB. In principle, the greater aim is to achieve compliance with the auditing process described in ISO 19158. This process is meant to ensure that the supplying organizations are capable of delivering data of expected quality. However, implementing a nation-wide process is rather challenging because municipalities and other organizations might not have the capability or resources to repair the quality issues identified by the QualityGuard. Inconsistent data quality is not desirable, and data quality requirements will be less strict at first phases of implementation. Some of the issues will be automatically repaired by the software once the process has been established, but the organizations will still receive a notification about data quality issues in any conflicting features.

The Finnish NTDB is in a continuous state of development and currently effort is made towards reaching automation, improved data quality and less overlapping work in co-operation with municipalities and other data producers. The QualityGuard has enabled an automated spatial data quality validation process for incoming data and it is currently being evaluated in practice. The results have already been well received by the users. Automating data quality validation is no longer a work of fiction. As indicated earlier we believe this will be a common practice with all SDI datasets in Finland.