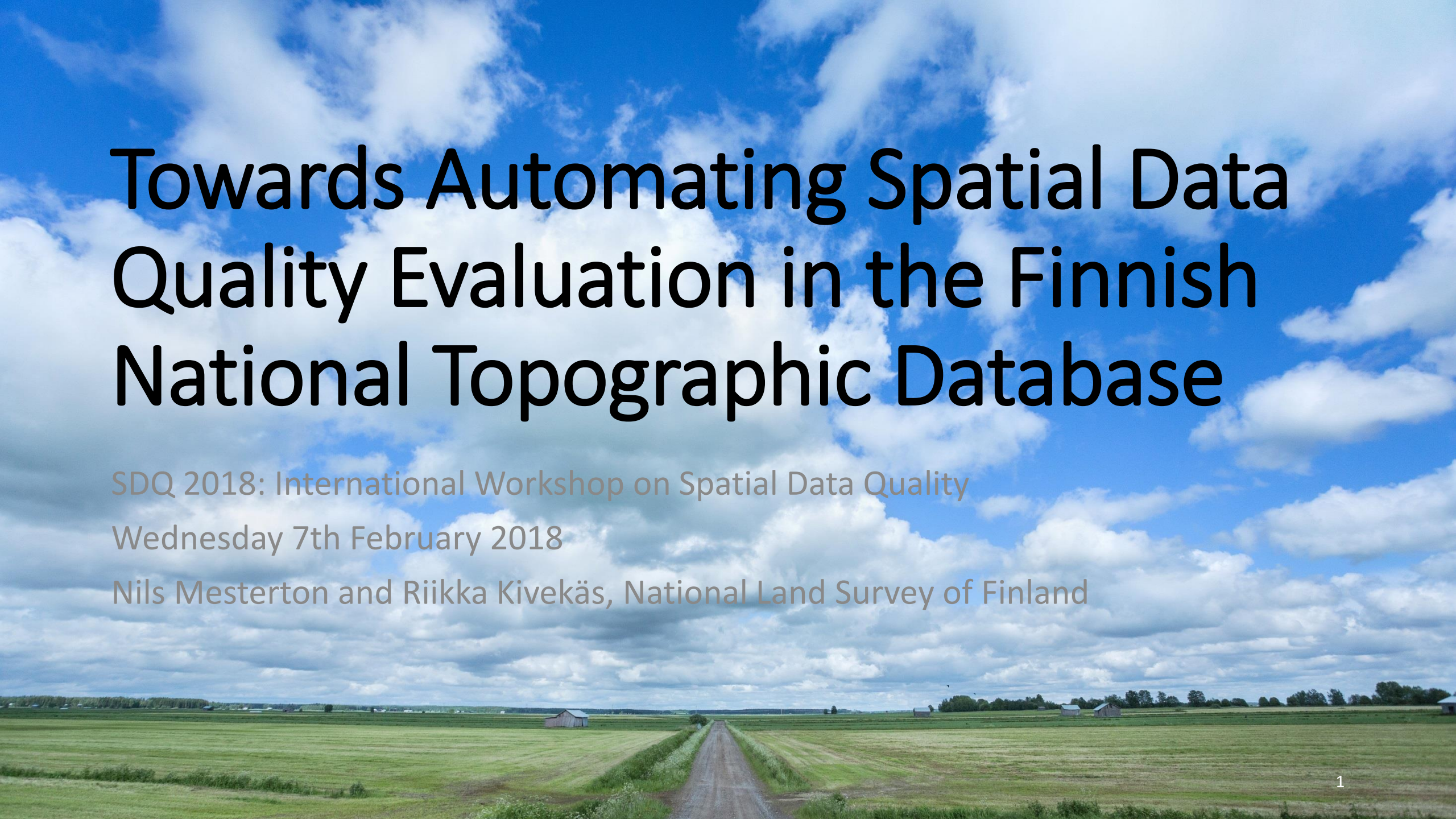


Towards Automating Spatial Data Quality Evaluation in the Finnish National Topographic Database

SDQ 2018: International Workshop on Spatial Data Quality

Wednesday 7th February 2018

Nils Mesterton and Riikka Kivekäs, National Land Survey of Finland

Content

- Introduction to National Topographic Database (NTDB) in Finland
- Quality rules
- Technical architecture and implementation
- Case Tampere: Buildings from city to NTDB
- Any questions?

KMTK
Kansallinen
Maastotietokanta

Towards National Topographic Database in Finland

Brief introduction to the National Topographic Database

Present State

- National Land Survey of Finland maintains the topographic database.
- All ~300 municipalities have their own topographic data.

➢ Topographic data is stored and maintained in silos by numerous organisations.

➢Overlapping work: Updates must be made in several places.

KMTK
Kansallinen
Maastotietokanta

What Are We Aiming At?

- Basic topographic data in one place
  - Buildings, transport network, hydrography, land cover, elevation
- Distributed data collection and maintenance
  - Reduced amount of overlapping work
  - Updates in real time
- Quality check before publication
  - Reduced amount of quality issues
  - Harmonization through synchronized requirements

KMTK
Kansallinen
Maastotietokanta

What Makes The New NTDB Better?

- Open data
  - Analysis of spatial data
  - Utilization in decision making
- Reduced costs
  - Data is collected only once
  - … And maintained in one place
- History management
  - Persistent IDs
  - Versions
- Moving to an entirely new dimension: 2D -> 3D

KMTK
Kansallinen
Maastotietokanta

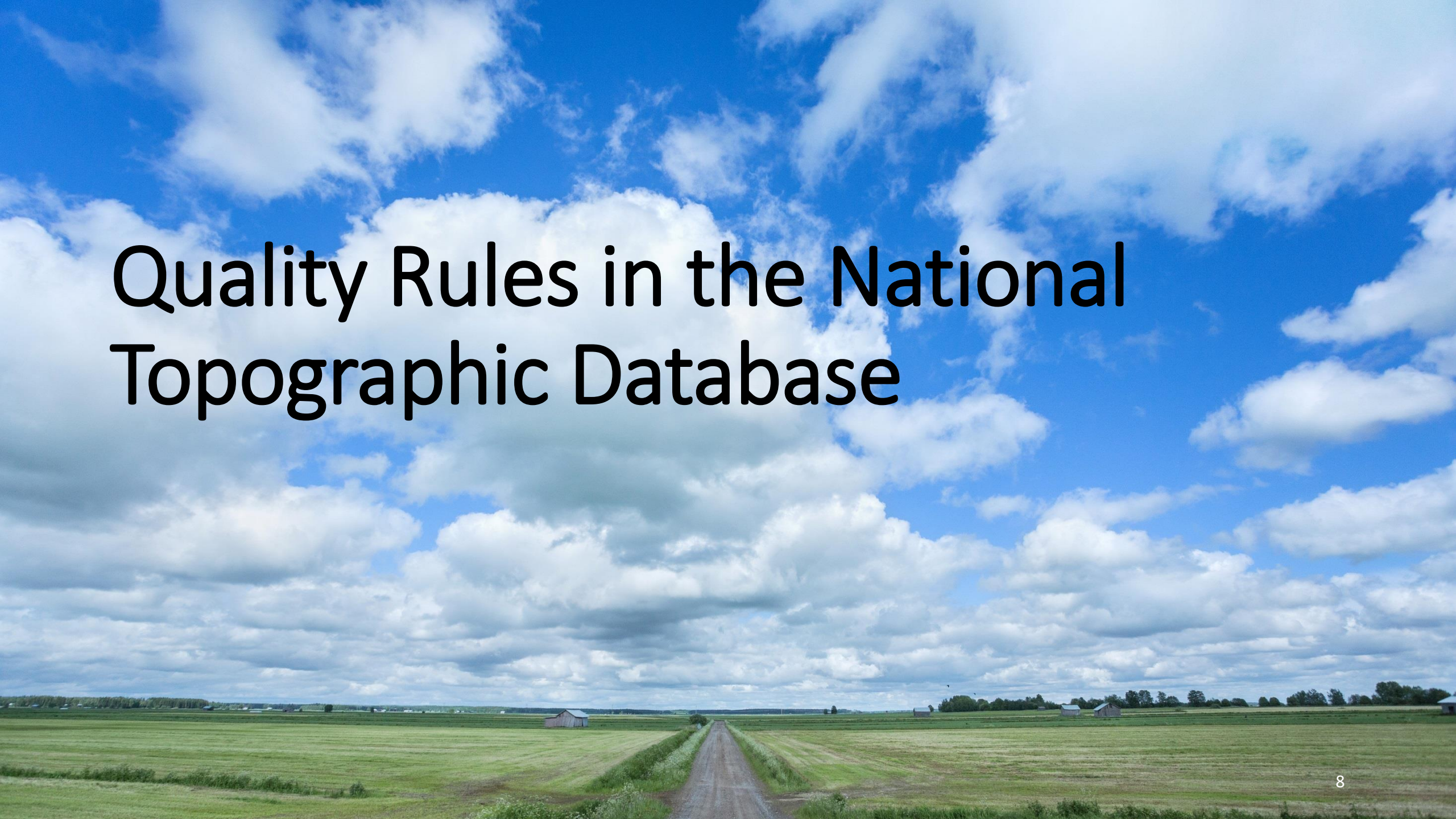
Who are involved in the project?

Governmental organizations

Municipalities

Other organizations

Metsätieto 2020



Quality Rules in the National Topographic Database

NTDB Quality Rules

- Built on what was learned in ELF and ESDIN projects
- Automatically testable rules regarding logical consistency
- Rules are basically definitions of what can be wrong with the data
- RuleSpeak provides structure and consistency for definitions
- Each NTDB theme gets their own set of quality rules
  - Buildings theme is nearly done
  - Addresses will be next

KMTK
Kansallinen
Maastotietokanta

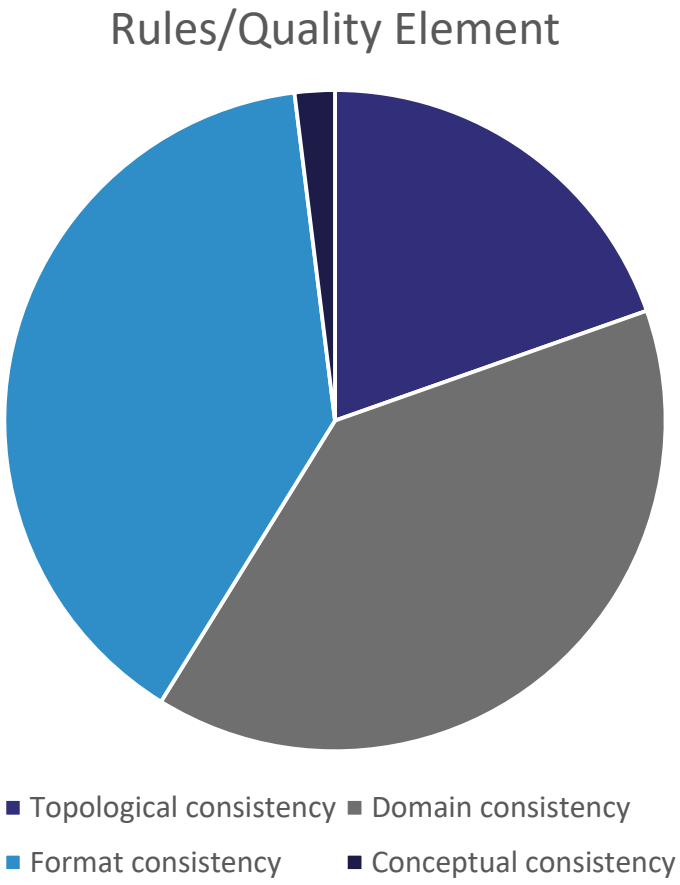
Conjuring Quality Rules for NTDB

- Process started with attempting to extract requirements from available sources
  - Feature descriptions of the existing TDB
  - INSPIRE Data Specifications
  - International standards, e.g. OGC Simple Features, ISO 19157 Data Quality
- These requirements were baked into initial quality rules
- NTDB specifications have been applied on the next iterations
  - Specs are still in a state of development…
  - Yes, it’s definitely a challenge

KMTK
Kansallinen
Maastotietokanta

What We Ended Up With

- 8 re-usable rule types
- ~50 rules for the Buildings theme
- Severity levels: WARNING and ERROR
  - In other words…



KMTK
Kansallinen
Maastotietokanta


YOU SHALL NOT PASS

What Shall…

NOT PASS

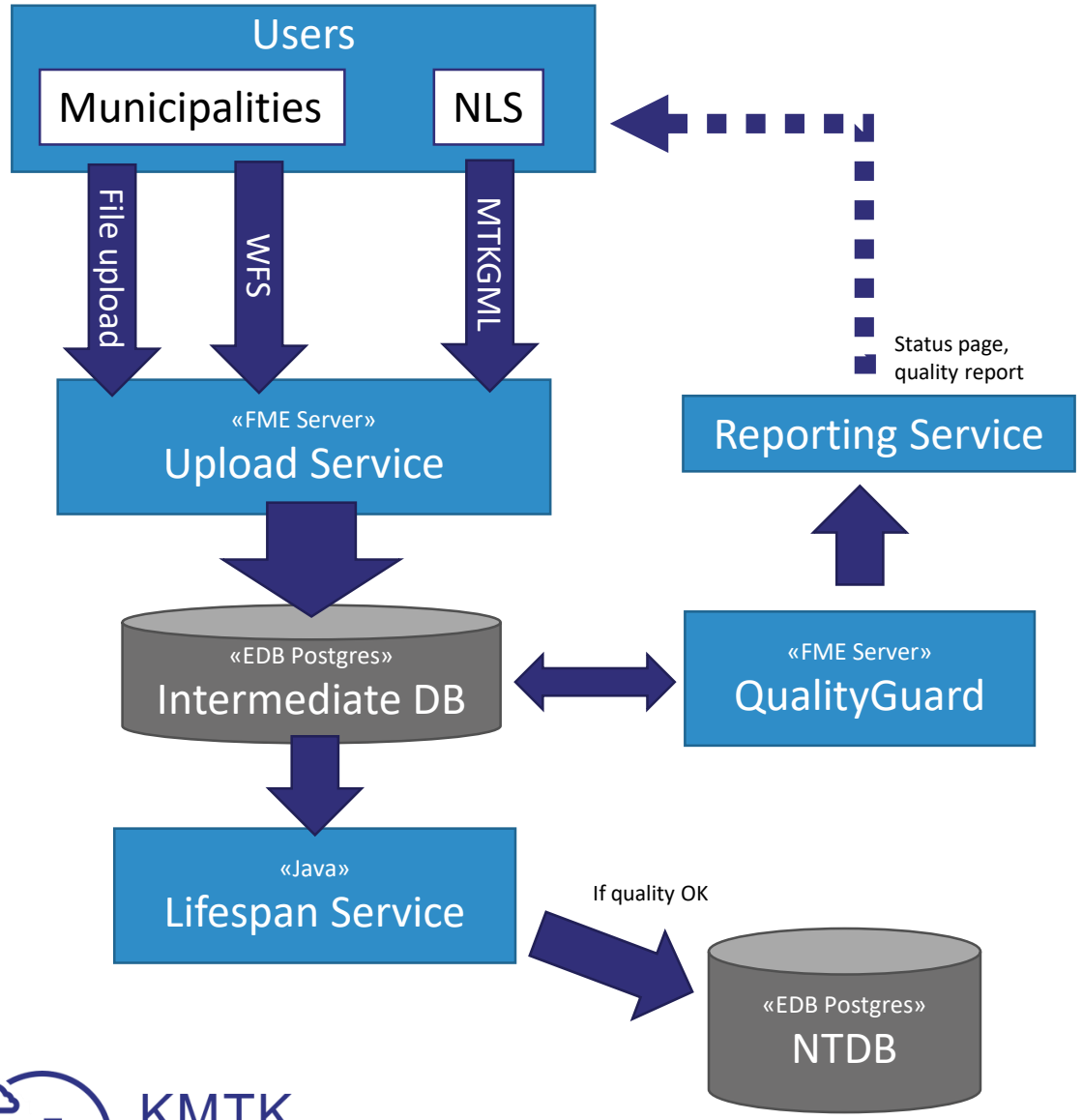
- Most issues regarding topological consistency
- Invalid and empty geometries
- Overlapping buildings and building parts

PASS

- Most issues regarding domain and format consistency
- ”Floor number must be between 1 and 100”
- ”Area in meters must be presented using the numeric data type”

KMTK
Kansallinen
Maastotietokanta

Technical Architecture and Implementation



KMTK
Kansallinen Maastotietokanta

Upload options:

- File upload via a UI
- Timed WFS download
- Timed processing of TDB exchange format

Process:

1. Data upload is iniated via a UI or automatically.
2. Upload service transforms data to the NTDB schema and inserts transformed features to an intermediate database. Format consistency is validated during this phase.
3. QualityGuard performs topological, conceptual and domain consistency validation.
4. Lifespan Service provides fresh identifiers to modified features if the change is significant enough
5. Data is inserted to the NTDB
6. Validation results and upload statistics are delivered to the user via the Reporting Service

QualityGuard

- FME® software implementation
  - Validation workflow and logic
  - Error dataset generation
- PostgreSQL database as a data and configuration store
- Quality Rule Bank
  - Quality rules
  - Rule types
  - Rule sets and user interface (soon™)

KMTK
Kansallinen
Maastotietokanta

Reporting Service

- Reporting service is basically a web page that shows the status and statistics of each session
- Modifications error locations can be observed on a map
- Users can download an error dataset for their sessions

KMTK
Kansallinen
Maastotietokanta

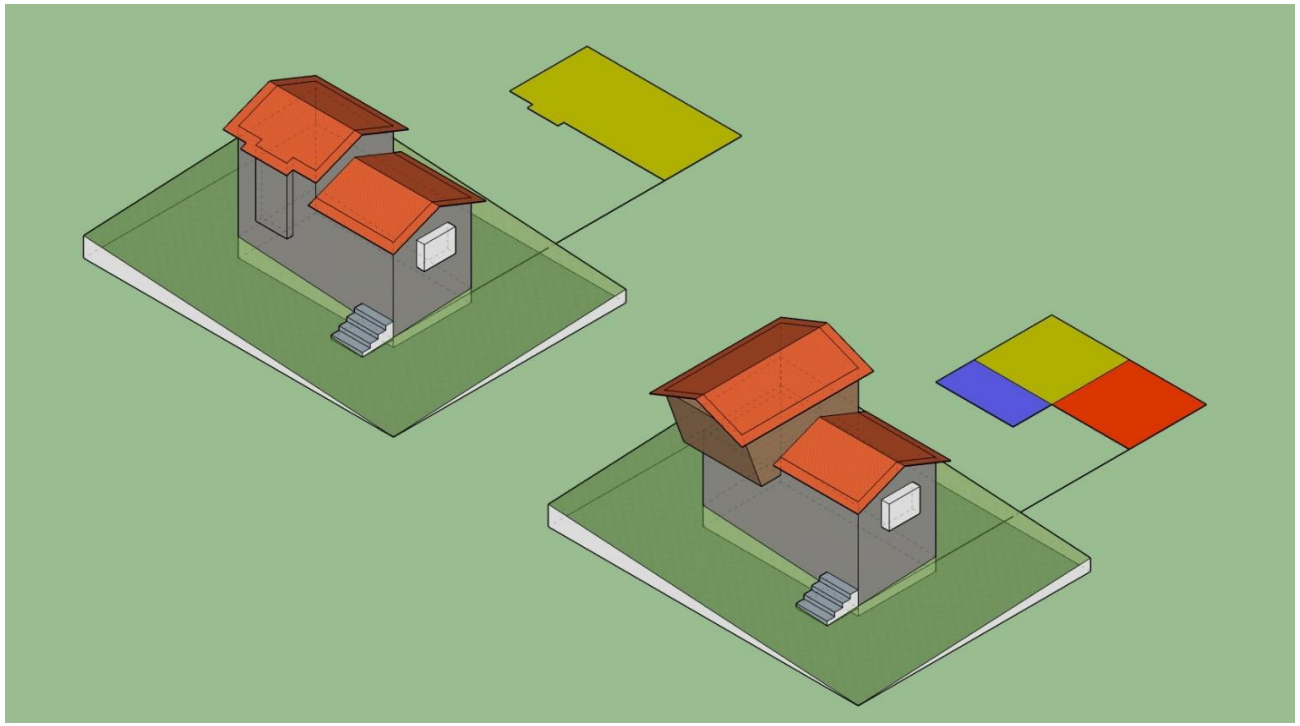
DQ Error Dataset

- Rule ID
- Quality element
- An original ID of the feature, if available
- Short description of the issue
- Point geometry indicating:
  - Location of geometry errors, e.g. self intersections and spikes
  - A generic location of a conflicting feature for attribute-related errors

KMTK
Kansallinen
Maastotietokanta

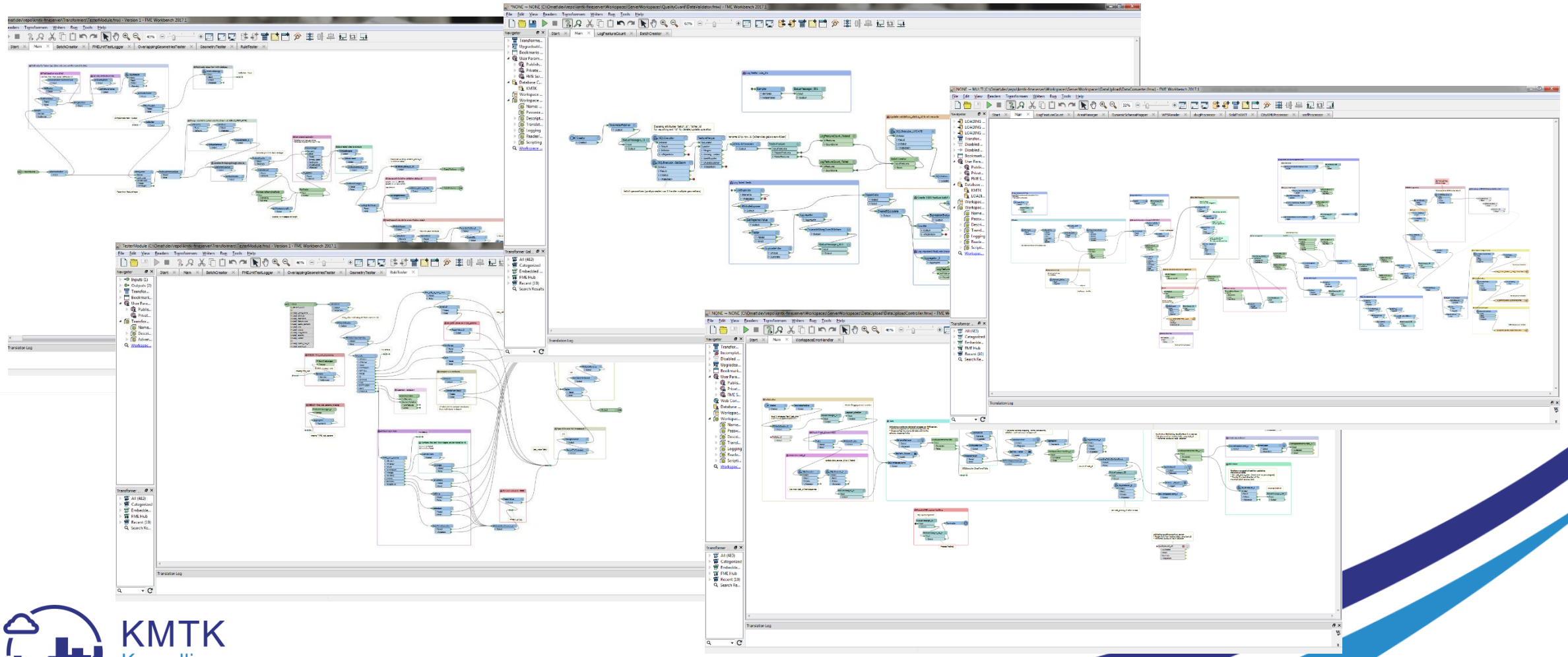
Implementation Plans, soon™

- Mechanism for chaining rule types (AND/OR)
- Rule sets
- Rules for validating 3D solids
- User interface for defining rules and rule sets for different themes and feature classes



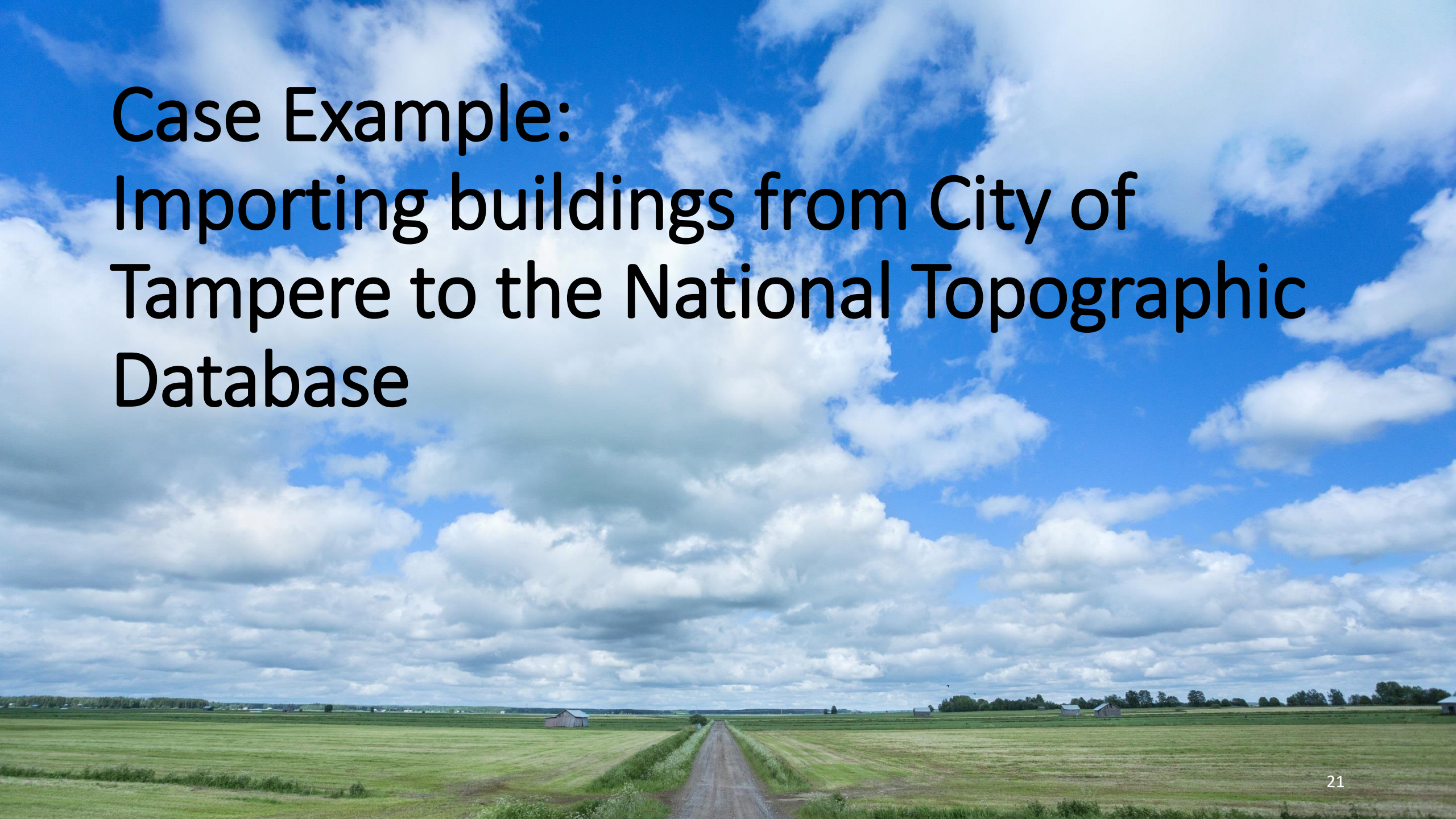
KMTK
Kansallinen
Maastotietokanta

Workspaces!



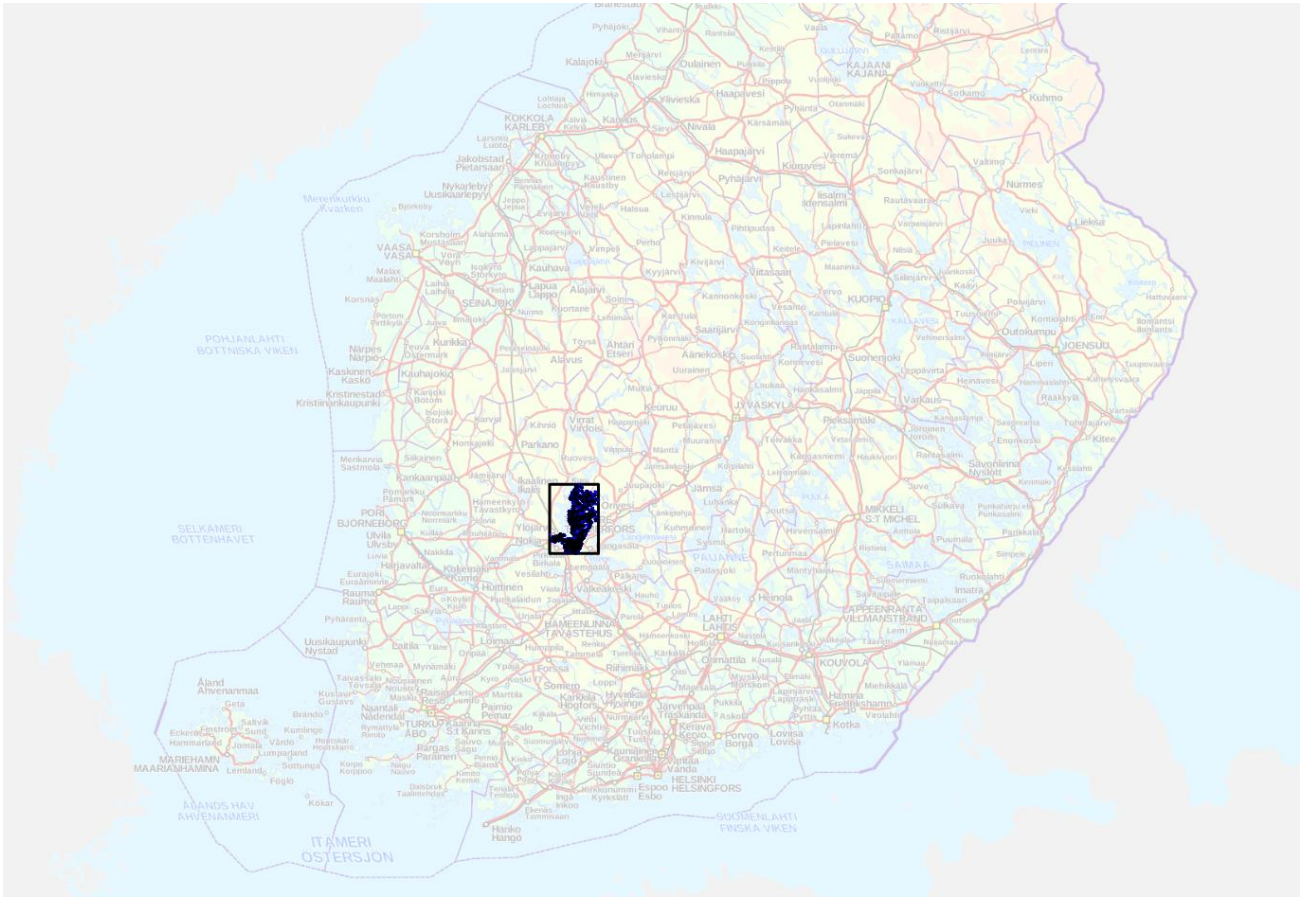
KMTK
Kansallinen
Maastotietokanta

Case Example: Importing buildings from City of Tampere to the National Topographic Database



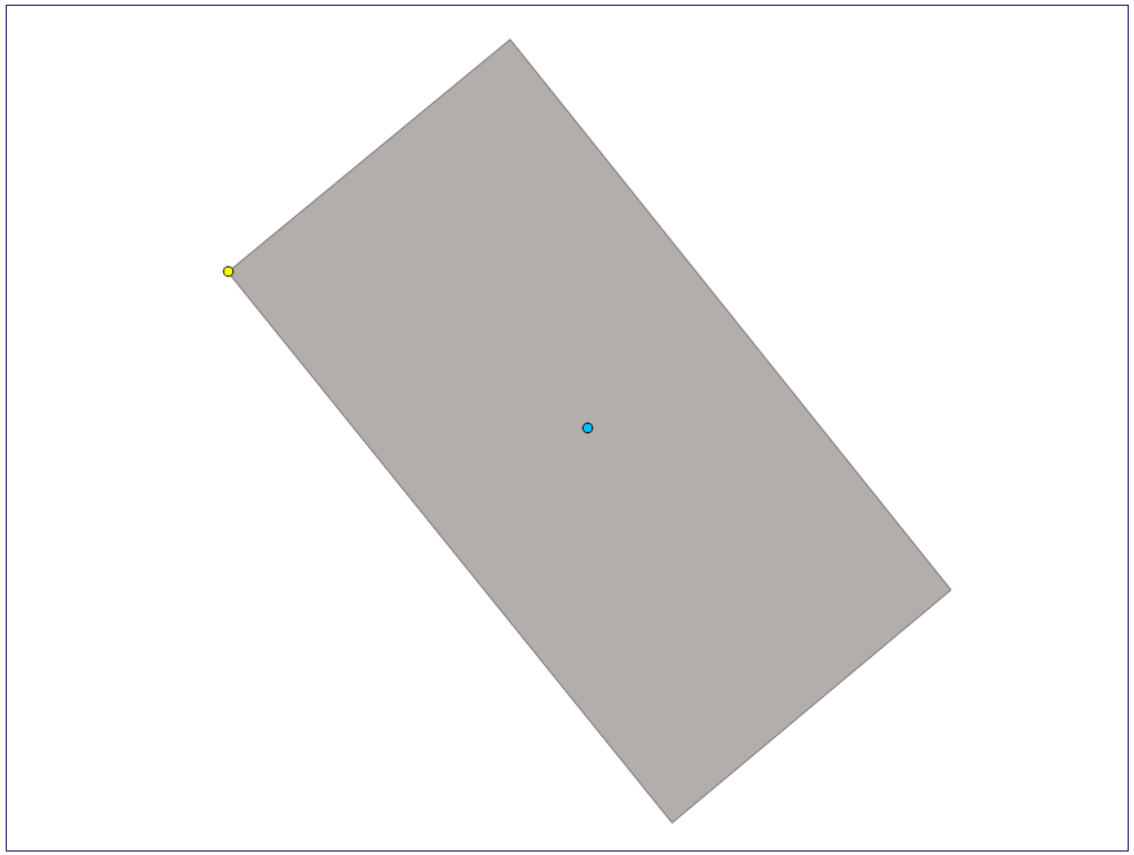
Tampere

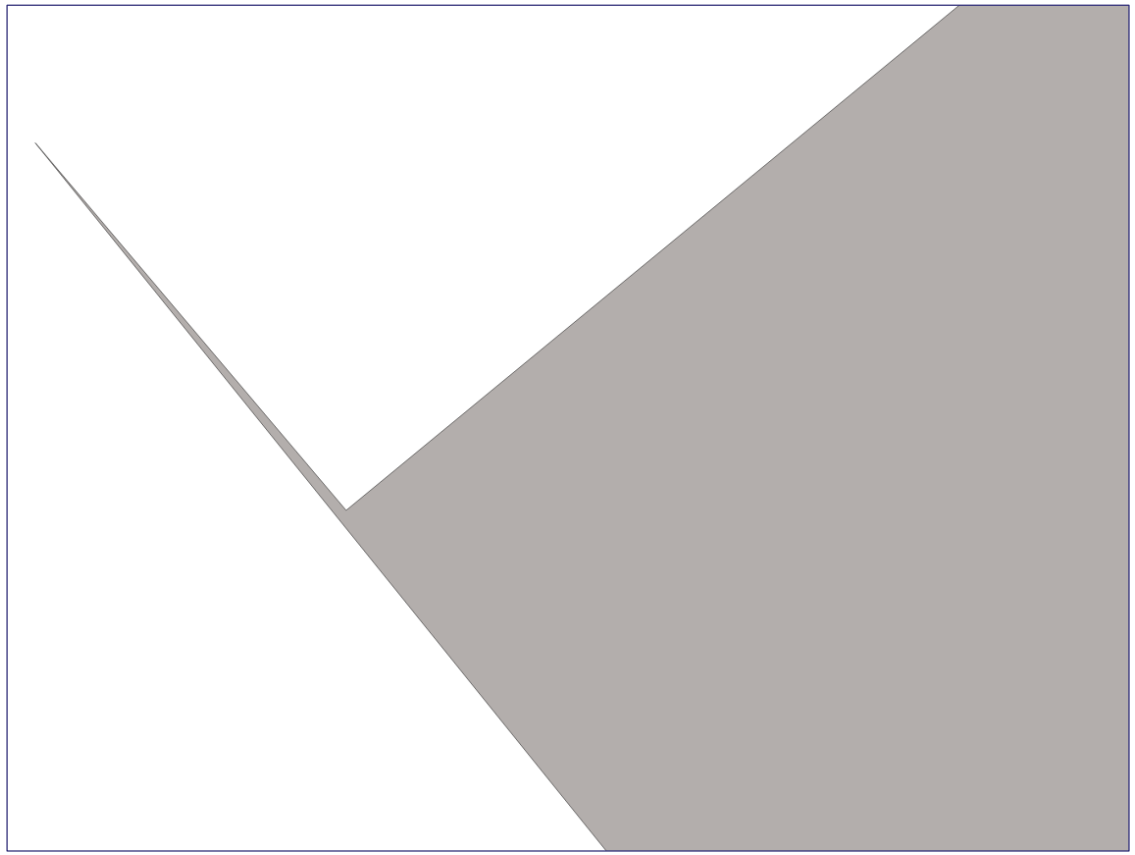
- Third most populous municipality and second largest urban area in Finland, with ~300k inhabitants in a ~5000 km^2 area
- WFS service offers ~50k buildings as polygons



KMTK
Kansallinen
Maastotietokanta

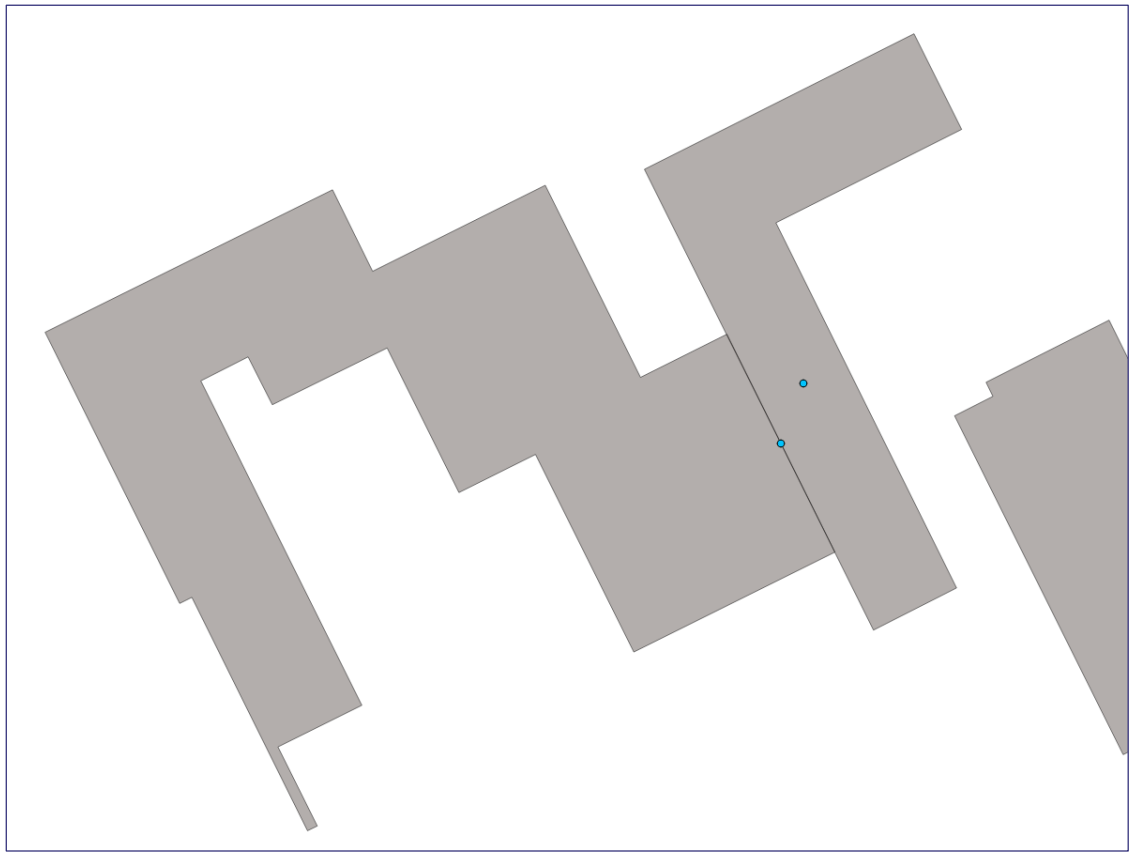
Invalid Polygons

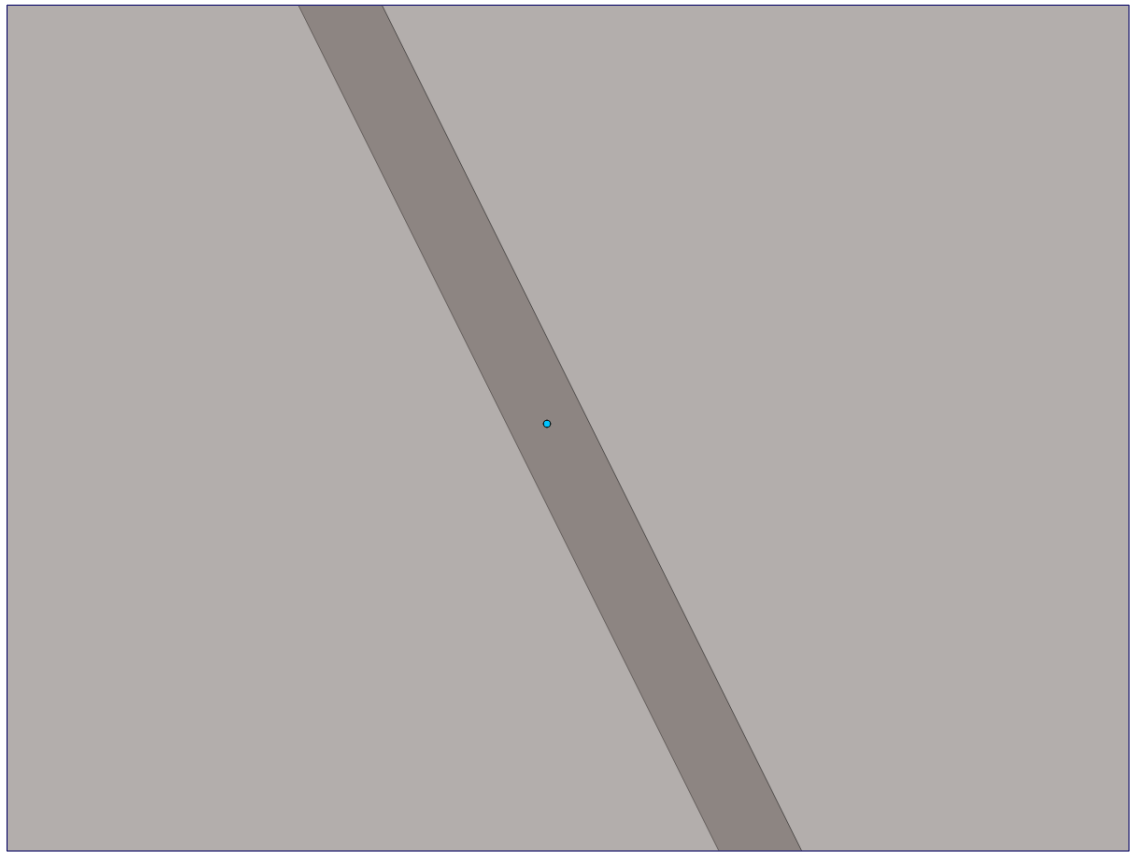




KMTK
Kansallinen
Maastotietokanta

Overlaps

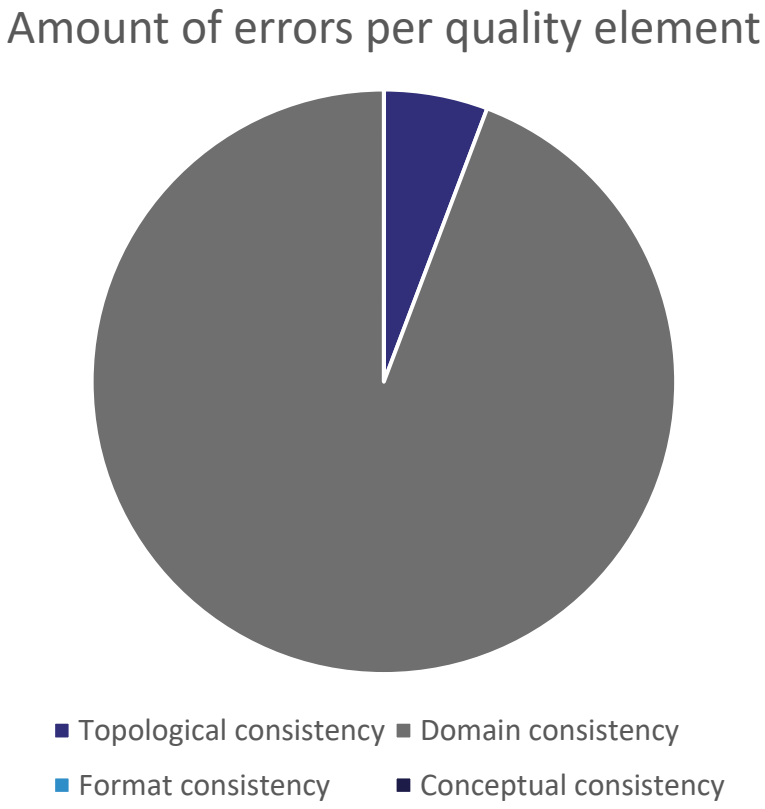




KMTK
Kansallinen
Maastotietokanta

Some Statistics

- 1060 errors, 17367 warnings
- Topological consistency: 1060
- Domain consistency: 17367



KMTK
Kansallinen
Maastotietokanta


Thank you!
Any questions?
Contact: nils.mesterton@nls.fi