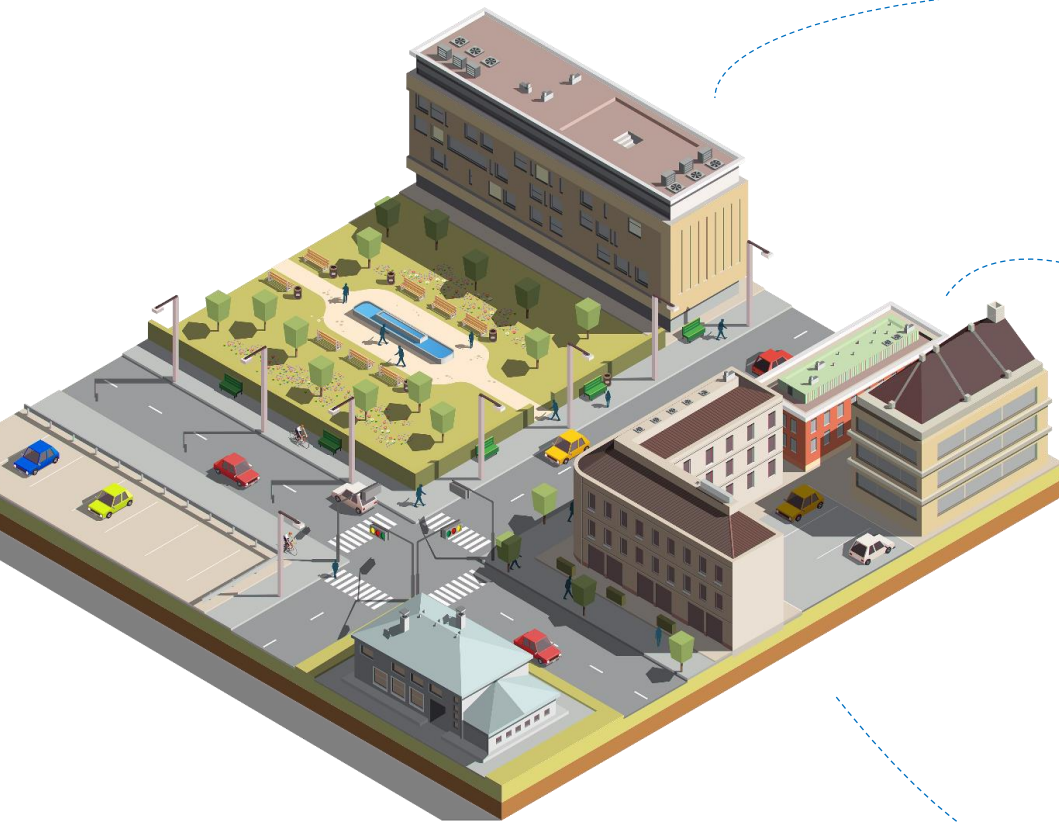# Evaluating Quality of Spatial Data Coming from Multiple Suppliers

Case Geospatial Platform of Finland

The Geospatial Platform in a nutshell

Spatial data is produced and maintained in the processes utilised by the municipalities, regions and the government, - **we will have our processes standardised by the year 2023.**

Decisions taken in society and the business world must be founded on reliable information – **machine readability is a key goal, building an accurate and up-to-date situational picture allowing people to make better decisions.**

More effective use of spatial data will save society's resources and create new innovations - **we will help companies and communities to succeed.**

Finnish National Topographic Database (NTDB)

- The database which Geospatial Platform in built on

- Based on:

  - conceptual models

  - International and national standards (for example ELF & INSPIRE)

- Consists of different feature types:

  - Buildings (2D + 3D)

  - Road links

  - Other manmade structures (such as building installations, dams, chimneys, pillars, platforms…)

  - Addresses

- Include quality validation and lifecycle management process for every feature (lifecycle processes are not presented in this presentation)

Different data providers

- Finnish National Land Survey has a comprehensive topographic database

- The most accurate data comes from municipalities

  - Integrating municipalities' data into NTDB

  - Different municipalities offer different kinds of data

  - Variety in format and quality

- No surprise: Large diversity of data is the biggest challenge in using multiple data sources building a centralized database

Overcoming versatility of data
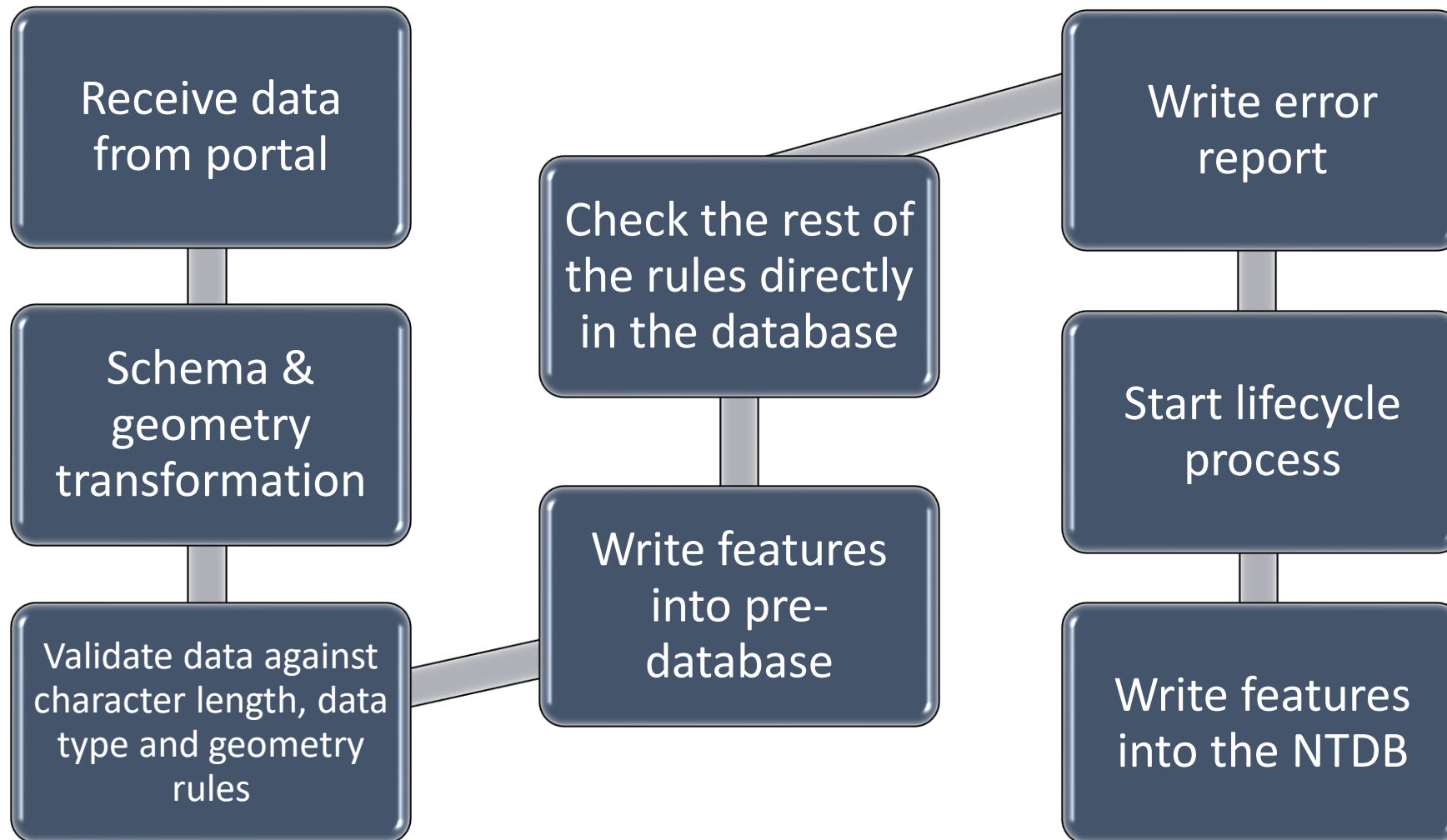
- Every data import goes through

1. **Schema transformation:**
   - Schema and format configuration is implemented on every dataset before they can be imported to the NTDB
   - Configurations are made manually, and is very prone to mistakes

2. **Quality check:**
   - To make sure no invalid features go into the NTDB
   - Based on quality rules: There are common rules that any feature should pass, and rules that are only implemented on certain feature type(s)
   - Produces an error report where data supplier can spot invalid features
   - Depends fully on correct schema transformation

Quality Guard and Data Upload Service

Receive data from portal

Schema & geometry transformation

Validate data against character length, data type and geometry rules

Check the rest of the rules directly in the database

Write features into pre-database

Write error report

Start lifecycle process

Write features into the NTDB

Quality Rules

- Divided into four different ISO 19157 quality elements:
    - Format consistency
    - Domain consistency
    - Topological consistency
    - Conceptual consistency

- 395 individual rules
    - A single rule can be implemented on 1+ feature types

- Make sure that contents of NTDB are in line with conceptual models
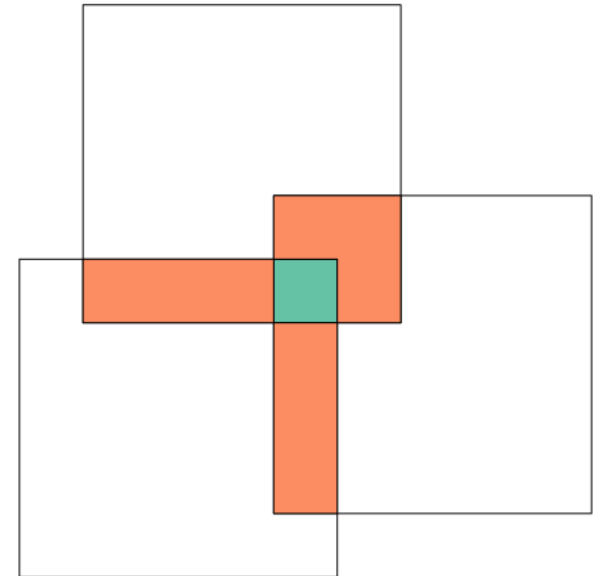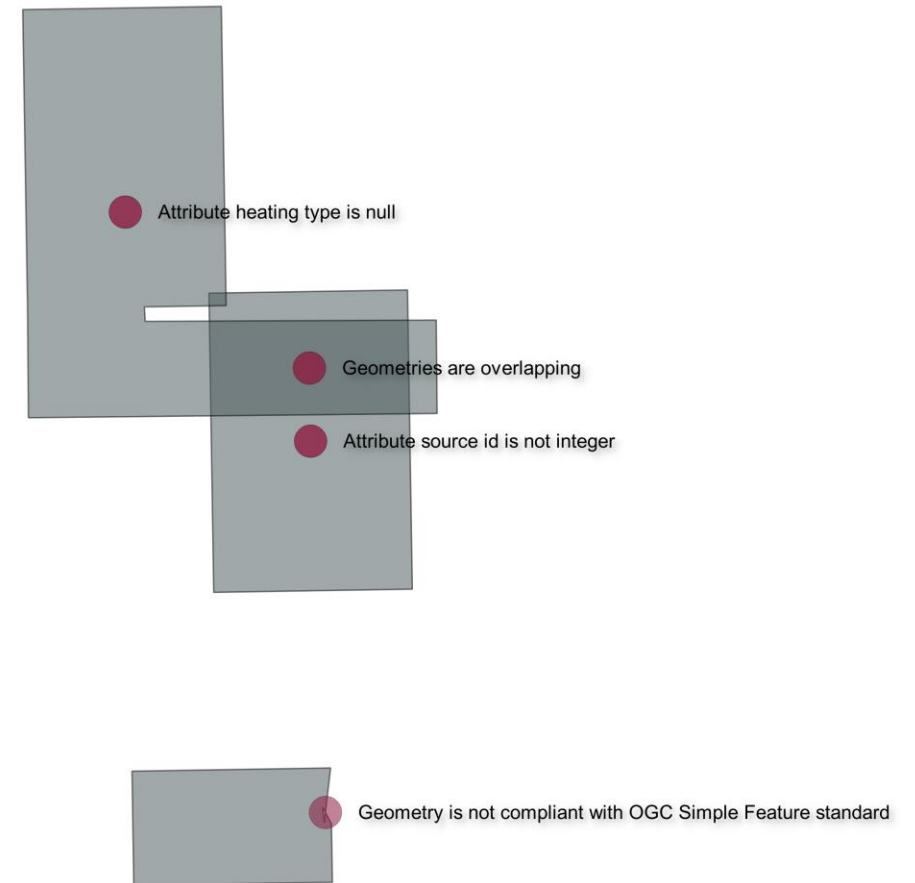
14 rule types

RULE TYPE	WHAT IS TESTED
Not null	Attribute has a value
Character length	Value consists of certain amount of characters
Geometry type	Geometry is the right type (area, point or line)
Value range	Value belongs in a predefined range of values
Belongs in a set	Value belongs in a predefined set of values
Data type	Data type is correct (integer, double, numeric, boolean, string, timestamp or date)
Distance	Distance between features (features that are linked to each other's can only have certain distance between them)
Compare	Value must be bigger, smaller or equal compared to another value
Overlaps	Features must not overlap more than defined ratio
Geometry validity	Geometry meets OGC SFSQL standards
Empty geometry	Feature has a geometry
RegEx	Value is consistent with a given regular expression
Within	Feature is within a given area
Name list	Value is found on a list (used to find out misspellings in address names)

Warning or error?

- Quality rule cause either an error or a warning

- Rules that cause error, prevent feature from going to the NTDB

  - For example: invalid data type in an attribute, wrong geometry type, invalid geometry

- Rules that cause warning are considered to be less severe and do not prevent features from going to the NTDB

  - Only on non-obligatory attributes

  - Invalid values are cleaned up

Quality report

- Errors and warnings are presented as an error report for data provider

- Downloadable and seen on the same user interface where data is imported from

- Includes:

  - Location of invalid feature (if topological error, error location)

  - Error description

  - Value causing the error

  - Severity (error / warning)

  - Attribute name that was disqualified (according to conceptual model)

  - Failed quality element

  - Invalid features identifier in the original system

Attribute heating type is null

Geometries are overlapping

Attribute source id is not integer

Geometry is not compliant with OGC Simple Feature standard

Errors and warnings presented in the user interface

Ylläpitoalue, virheet ja huomautukset kartalla

● Virhe ● Huomautus ☐ Ylläpitoalue ☐ Ladattu aineisto

LATAA TULOKSET

To conclude

- National Topographical Database can be built on multiple data providers

- Must have:

    1. Solid schema transformation

    2. Comprehensive quality check

    3. Dynamic process that takes care no invalid features can go to NTDB

Thank You!

mari.isomaki@maanmittauslaitos.fi