2nd International Workshop on Spatial Data Quality
6 February 2018

Quality requirements for Core Data

François Chirié, Dominique Laurent, France

UN-GGIM: EUROPE
UNITED NATIONS INITIATIVE ON
GLOBAL GEOSPATIAL
INFORMATION MANAGEMENT

Plan

- Introduction – Context
- Defining core data content
- Quality requirements
  - Completeness
  - Accuracy
- Conclusions – main learnings

Introduction - Context

UN-GGIM (United Nations initiative on Global Geospatial Information Management)

- Objectives
  - coordination forum between Member States on geographic information
  - Targeting the Sustainable Development Goals

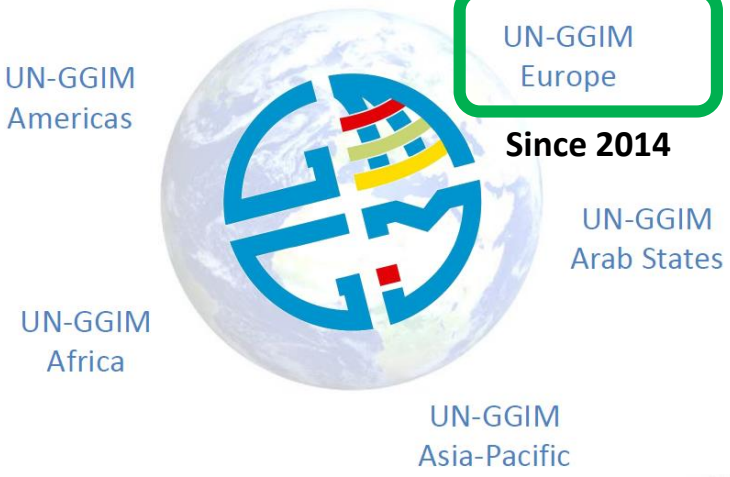
Unleashing the power of 'Where' ….

… to make the world a better place.

UN-GGIM

- Organisation
  - Activities at global level (since 2011)
    - Global Geodetic Reference Framework
    - Land registration
    - Institutional arrangements
    - Fundamental data
    - …..
  - Activities at regional level

Strong involvement of statistical community



Core data (chaired by France)

Data integration + SDG indicators (chaired by Germany)

What is core data?

- Core data is priority data
  - Geographic data
  - The most useful to analyse, achieve or monitor the SDGs (Sustainable Development Goals)
  - Directly or indirectly
    - Indirectly: used as background to georeference user data or to derive other data or to enable combination with other data

Sustainable Development Goals

Objectives of the Working Group on European core data

- **Describe core data** recommended to be produced and supplied by UN Member States of geographic Europe
  - Common requirements → common (minimum) content

- **Define priorities** for production of new data or for improvement of existing data
  - Recommendation meant for decision makers and data providers

Organisation

- 15 European countries
  - Austria
  - Belgium
  - Finland
  - France (chair)
  - Netherland
  - Switzerland
  - Turkey
  - Germany
  - Greece
  - Poland
  - Spain
  - Sweden
  - UK
- Observers
  - JRC, EEA, EuroSDR

Work plan

Phase 3: Economic model, political and financial frameworks for supporting core data availability
2018?

Phase 2: Recommendation for content of core data themes
2016 - 2018

Phase 1: Selection of core data themes
2015 - 2016

First step

Selection of core data themes from INSPIRE data themes

Annex I

Coordinate Reference Systems

Geographical Grid Systems

Geographical Names

Administrative Units

Addresses

Cadastral Parcels

Transport Networks

Hydrography

Protected Sites

Annex II

Elevation

Land Cover

OrthoImagery

Geology

Annex III

Statistical units

Buildings

Soil

Land use

Human health and safety

Utility and governmental services

Environmental monitoring facilities

Production and industrial facilities

Agricultural and aquaculture facilities

Population distribution - demography

Area management/restriction/regulation

Natural risk zones

Atmospheric conditions

Meteorological geographical features

Oceanographic geographical features

Sea regions

Bio-geographical regions

Habitats and biotopes

Species distribution

Energy resources

Mineral resources

Defining Core Data Theme Content

UN-GGIM: EUROPE

UNITED NATIONS INITIATIVE ON

GLOBAL GEOSPATIAL

INFORMATION MANAGEMENT

Objectives

- Work out '**Recommendations for Content**' for the selected core data themes
- Based on
  - Existing standards: mainly INSPIRE
  - User requirements with focus on SDG related use cases

User requirements: sources

- Bibliography
  - INSPIRE use cases
  - SDG and SDG targets
  - ….
- Interviews, workshops
- Questionnaires
- WG A expertise, discussions

Reasonable guess of **main** user requirements

Core data and INSPIRE

- INSPIRE is very good starting point
- But **limited to harmonisation of existing data**
- Very flexible data specifications
  - No indication about the expected level of detail
  - Most attributes are “voidable”
  - No quality requirements, few quality recommendations
  - Options

Expected content more or less clear according to themes

Core data profiling INSPIRE

- Selection of priority feature types and attributes
  - Extensions possible (but rare)
- Expected level(s) of detail
- Data capture rules
- **Quality recommendations**

Core data and INSPIRE

INSPIRE: STRUCTURE

CORE DATA: CONTENT

Core data

encourages

Production of common content

+

+

INSPIRE

mandates

Delivery of data according to common data model, format, services

Users get harmonised data with both common content and common structure

Comparison with INSPIRE

The INSPIRE big cheese
with lots of holes

Users begin to complain:
not so much to eat!

The core data cheese:
smaller but compact
and really filled

Principles

- 2 types of recommendations:
  - **Core recommendation**: first priority, highly required, achievable → ideally, short term action
  - Good practice: second priority, bring added value to core data → to be encouraged
- *Considerations for future: potential trends → long term potential actions*

Quality requirements

Completeness

Do users require completeness?

- Yes, of course
- Ideally, 100 % on all features and attributes
- But **core data is**
  - not about ideal data
  - but about **minimum data**
- Core recommendations must be achievable

How to get relevant completeness targets?

- Choose achievable targets
  - By selecting reasonable thresholds
    - 100% for cadastral parcels or administrative units
    - 95% for addresses
  - By selecting data that may be captured in all cases
    - Attributes for theme Buildings
      - On current use , on number of floors or height above ground
      - But not on date of construction
        - » Information may be lost

How to get relevant completeness targets?

- Only on “corissime” features
  - For theme GeographicalNames
    - populated places > natural features (landforms, landcover, …)
- Only if clear data capture rules
  - Completeness is against the nominal terrain
  - Nominal terrain has to be clearly defined.

The target includes all the populated places of interest for mapping or geocoding, such as cities, villages, neighbourhoods, hamlets, isolated buildings. In other words, names of single buildings in urban areas having also a classical address (house number + street name) are considered of minor interest and are not in the target of this rule.

Do users require more than completeness?

- Completeness is against the nominal terrain
- Nominal terrain is real world “seen” through data specification
- But real world is not always good enough!

Do users require more than completeness?

- Example 1: cadastral parcels
  - User requirements: on whole (land) territory
  - In some countries, no cadastral parcels on public domain

Good practice : Data for theme Cadastral Parcels should be available on whole land territory, including public domain.

  - May imply change in national law

Do users require more than completeness?

- Example 2: addresses
  - User requirements: “true” addresses everywhere
  - In some rural areas, poor address system
    - No street name, no house number
    - => Address only at village level

Good practice :All basic units of addressing should be provided with a unique address enabling their unambiguous location, i.e. an address with a locator and geographic position.

  - Municipalities to provide street names, address numbers
  - New addresses to be used by citizens in everyday life

Quality requirements

Accuracy

Absolute positional accuracy

- Gap between the position in data set and the position in nominal terrain
- Quality criteria widely used

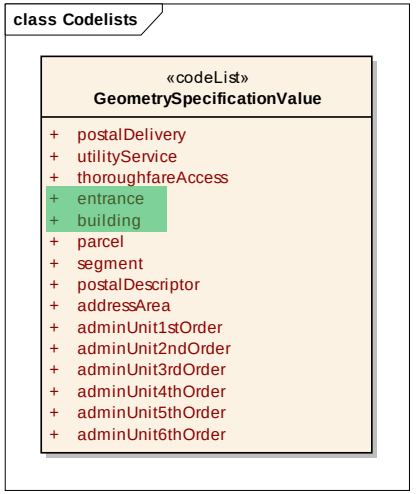
Core data and absolute positional accuracy

- Case 1: the real world is well-defined and the nominal terrain is the "real world"
  - Use of absolute accuracy
  - Examples: Cadastral Parcels, RoadLinks

Core recommendation: Cadastral parcels should have in general an absolute accuracy of 1m or better in urban areas and of 2,5m or better in rural areas. In case of new surveys, it is recommended to use methods enabling absolute accuracy better than 50 cm.

Core data and absolute positional accuracy

- Case 2: the real world is well-defined and the nominal terrain is a profile of the "real world"
  - Example 1: Multiple representation of Addresses
    - Preferred representation:

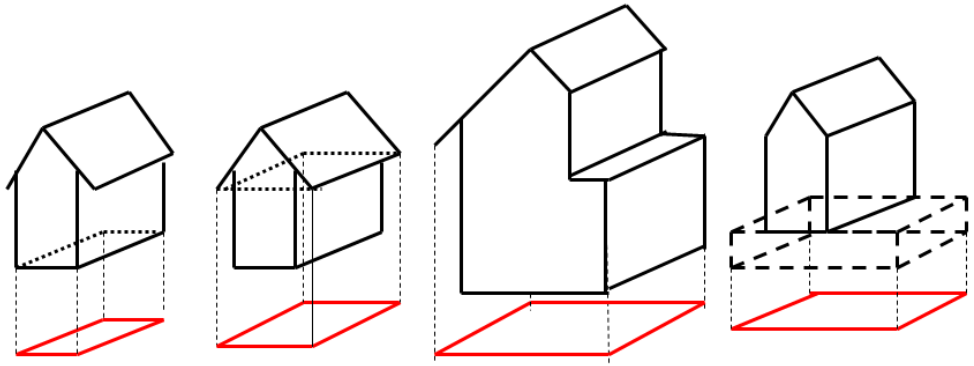


Good practice: wherever possible, building or entrance should be used, for reasons of precision.

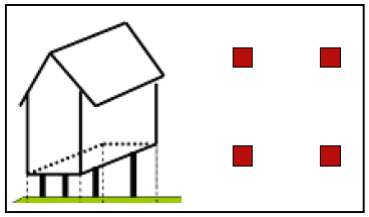
Co-ordinates of an address should be accurate to within 5 metres of the true position of the building centroid or entrance, where possible.

Core data and absolute positional accuracy

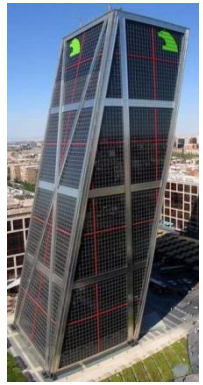
– Example 2: Multiple representation of Buildings

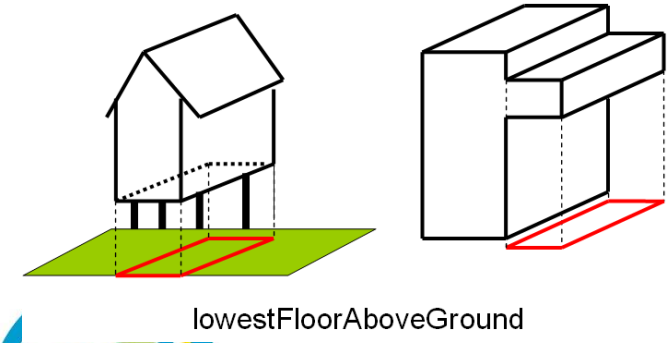


footprint roofedge aboveGroundEnvelope envelope



footprint





lowestFloorAboveGround

No preferred representation

Recommended accuracy (2m or better) is meaningful only if the horizontal geometry reference is documented.
(as in INSPIRE)

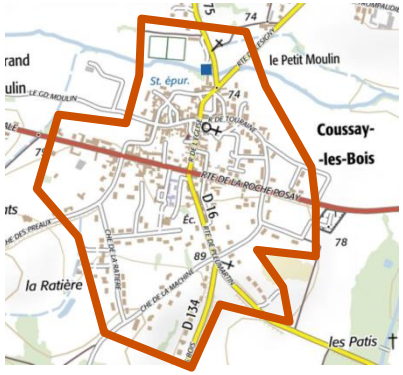
Core data and absolute positional accuracy

- Case 3: the real world is poorly defined
  - Limits of features are fuzzy
  - Examples: named places in theme GN
  - In existing data, generally represented just by a point
    - Location of the label on the map
  - But not close to “real world” (e.g. mountains, seas)

Core data and absolute positional accuracy

- Case 3: core data recommendation
  - Provide “true” geometry” – generally a surface
    - Gives an idea of “importance” of the named place
  - Document the reliability of this geometry

Good



Medium

Undefined

Core data and absolute positional accuracy

- Case 4: the “real world” hardly exists
  - Case of artificial boundaries
    - Administrative units (AU)
    - Regulated or managed areas (AM)
  - The “real world” is defined in legal texts not always explicit enough
    - Old text referring to objects that have changed or that no longer exist
    - Fuzzy text(s) with various possible interpretations

Core data and absolute positional accuracy

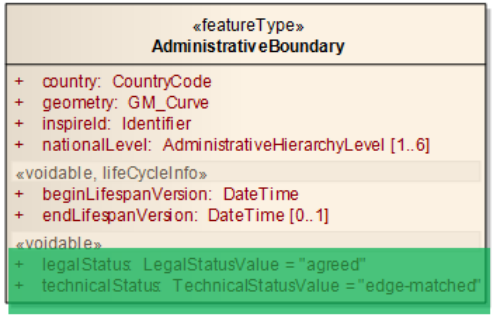
- Case 4: the “real world” hardly exists
  - User requirements:
    - get reference data => common representation used by all stakeholders

Some hundreds meters are not important in high sea. What we need is a common line agreed by all.
(about Maritime Units)

    - if geographic data doesn’t provide “truth”, it may be source of issues

Core data and absolute positional accuracy

- Case 4: the “real world” hardly exists
  - Producer point of view
    - Geographic data as **representation** of AU or AM: easy
    - Geographic data as **definition** of AU or AM: tricky
    - Example: international boundaries
      - Technically Edge-matched at Regional level
        - » Agreement between NMCAs
        - » GI seen as representation of AU
      - At master level, technically edge-matched only if legal agreement
        - » technical edge-matching to be submitted to official boundary commissions
        - » GI seen as definition of AU



Core data and absolute positional accuracy

- Case 4: the “real world” hardly exists
  - Core data recommendation
    - Step 1: document if geographic information provides legal “truth”
      - Using INSPIRE attributes for AU
      - Extending the INSPIRE model for AM
    - Step 2: encourage reference data (common convention)

It is recommended to have reference data on maritime and land administrative units, agreed and used by all stakeholders.

Relative positional accuracy

- Within a theme
  - Administrative units of same level forming partition of territory
    - Topology that can be checked automatically
  - Road data conform to the navigation logic (junctions, bridges)
- Between themes

Addresses should be located on the «right » side of the street and in the "right" administrative or statistic unit.

Relative positional accuracy

- Recommendations (relatively) easy to formulate

Data on administrative units should be consistent with data on cadastral parcels and with data on topographic features (roads, buildings, …)

- But not so easy to be applied
  - May be currently impossible to align both on cadastral parcels and on topographic features
  - Also in "considerations for future": encourage investigation, coordination of data producers

Conclusions

Main learnings

How to define quality requirements?

- Usual quality criteria (e.g. absolute positional accuracy)
  - are well adapted to topographic features
  - but not so much for artificial – administrative features
- For artificial features, users want
  - Reference data, common convention
  - Consistency with background data, if any

How to achieve quality requirements?

- Efforts of data producers to maintain and enhance their data
- Struggle to get reference data, to ensure cross-theme consistency
  - Regulations
  - Coordination of data producers
- The administrative system (AD, CP, AU, AM) may also have to be upgraded
- => Geographic Information Management (UN-G**GIM**) is required