Towards the integration of authoritative and OSM geospatial datasets in support of the European strategy for data

Alessandro Sarretta & Marco Minghini
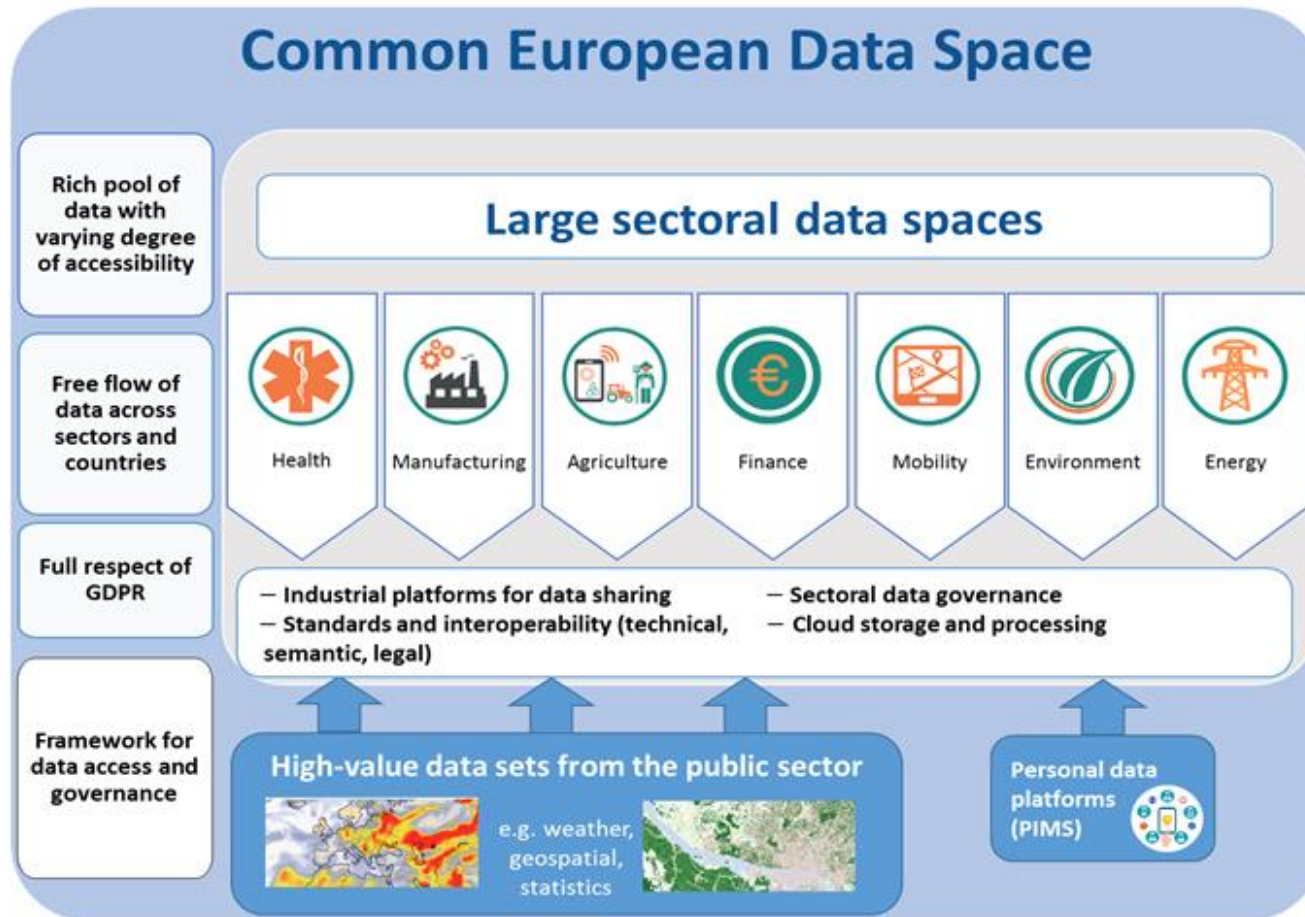
FOSS4G 2021 – 30th September 2021

European strategy for data

- Launched in February 2020 within the Europe Fit for the Digital Age priority.

- Aims at the creation of a European single market for data.

- Highlights the problems to address:

  - data availability (licensing), sharing, interoperability, quality, governance, infrastructure & technologies, skills & data literacy, cybersecurity

EUROPEAN COMMISSION

Brussels, 19.2.2020
COM(2020) 66 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

A European strategy for data

https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

European strategy for data

Common European Data Space

Rich pool of data with varying degree of accessibility	**Large sectoral data spaces**
Free flow of data across sectors and countries	Health · Manufacturing · Agriculture · Finance · Mobility · Environment · Energy
Full respect of GDPR	– Industrial platforms for data sharing – Standards and interoperability (technical, semantic, legal) – Sectoral data governance – Cloud storage and processing
Framework for data access and governance	High-value data sets from the public sector e.g. weather, geospatial, statistics · Personal data platforms (PIMS)

- Envisages the establishment of:
  - a common EU data space
  - sectoral data spaces
- benefiting society and economy
- through the combination of data across actors and sectors
  - including citizen-generated data!

https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

European Commission

Aim of the study

- We zoom on the domain of geospatial data and consider OpenStreetMap (OSM) as the most popular source of citizen-generated data.

- Goal: investigating enablers & barriers of integrating OSM with authoritative data from EU National Mapping Agencies, with a focus on:

  - the production of integrated datasets (not only comparison/OSM quality assessment)

  - the national level (in contrast to literature)

- Output: deriving recommendations on technical, semantic, organisational and legal interoperability to support the establishment of data spaces.

- First experiment: integration between the Finnish national address datasets from OSM and the National Land Survey of Finland.

European Commission

Data sources

- OpenStreetMap (OSM)
  - worldwide collaborative project
  - open data (ODbL licence)
  - simple flat data model using tags
  - bulk download of the full planet

- National Land Survey of Finland (NLS)
  - official national dataset
  - open data (CC-BY 4.0 licence)
  - INSPIRE compliant with Addresses data theme
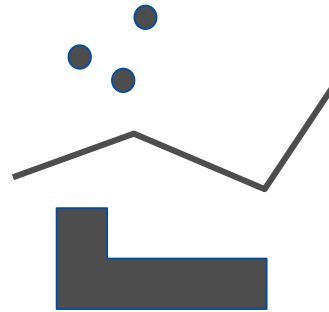 - OGC API Features

Data model: OSM

3 geometry types:
- nodes
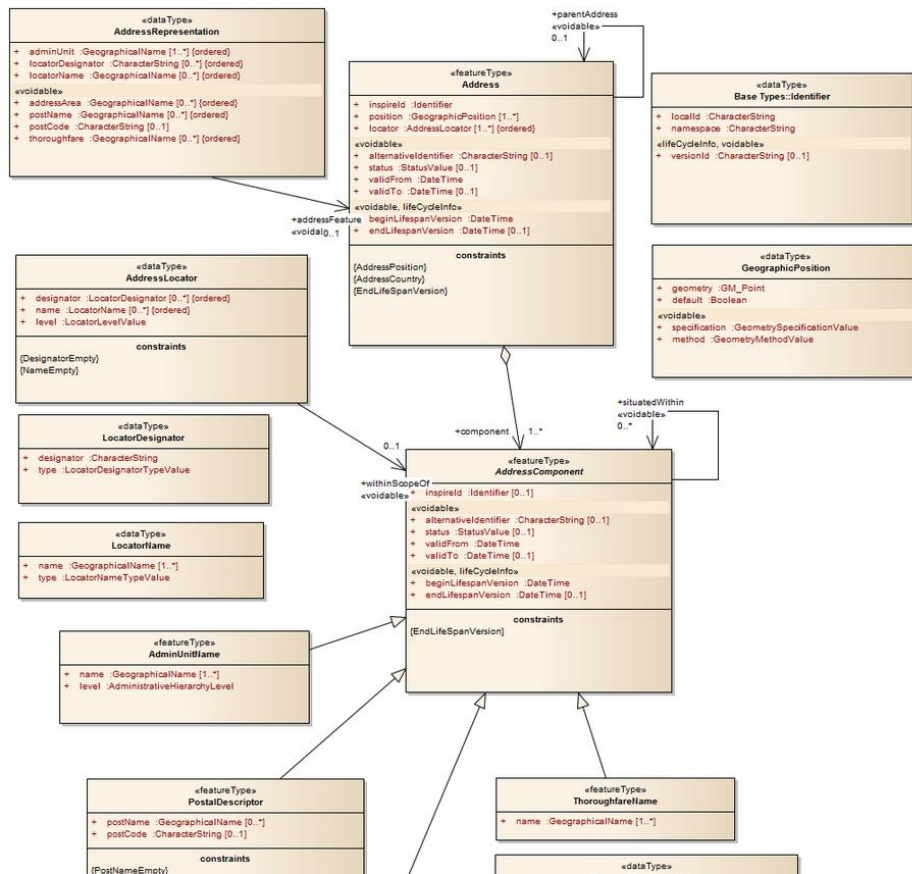- ways
- relations

1 attribute type:
- tags
  - <key> = <value>
  - <key2> = <value>
  - ...

Addresses

| OSM tag	Description
addr:country	country code of the address
addr:city	name of the city of the address
addr:street	name of the street of the address
addr:housenumber	building number of the address

European Commission

Data model: NLS

Based on the INSPIRE UML data model for Addresses data theme

NLS attribute	Description
component_ThoroughfareName_name fin	name of the street of the address in Finnish
component_ThoroughfareName_name swe	name of the street of the address in Swedish
component_ThoroughfareName_name sme	name of the street of the address in Sami
locator_designator_addressNumber	building number of the address
component_AdminUnitName_4	code of the city of the address
component_AdminUnitName_1	country name of the address

Data model integration

NLS

NLS attribute	Description
component_ThoroughfareName_name fin	name of the street of the address in Finnish
component_ThoroughfareName_name swe	name of the street of the address in Swedish
component_ThoroughfareName_name sme	name of the street of the address in Sami
locator_designator_addressNumber	building number of the address
component_AdminUnitName_4	code of the city of the address
component_AdminUnitName_1	country name of the address

OSM

OSM tag	Description
addr:country	country code of the address
addr:city	name of the city of the address
addr:street	name of the street of the address
addr:housenumber	building number of the address

INSPIRE/NLS attributes	OSM attributes
component_ThoroughfareName_name fin	addr:street
locator_designator_addressNumber	addr:housenumber
component_AdminUnitName_4	addr:city
component_AdminUnitName_1	addr:country

Integration process

- Step by step reproducible workflow
- Implementation through FOSS4G: QGIS Graphical Modeler
- GitHub repository
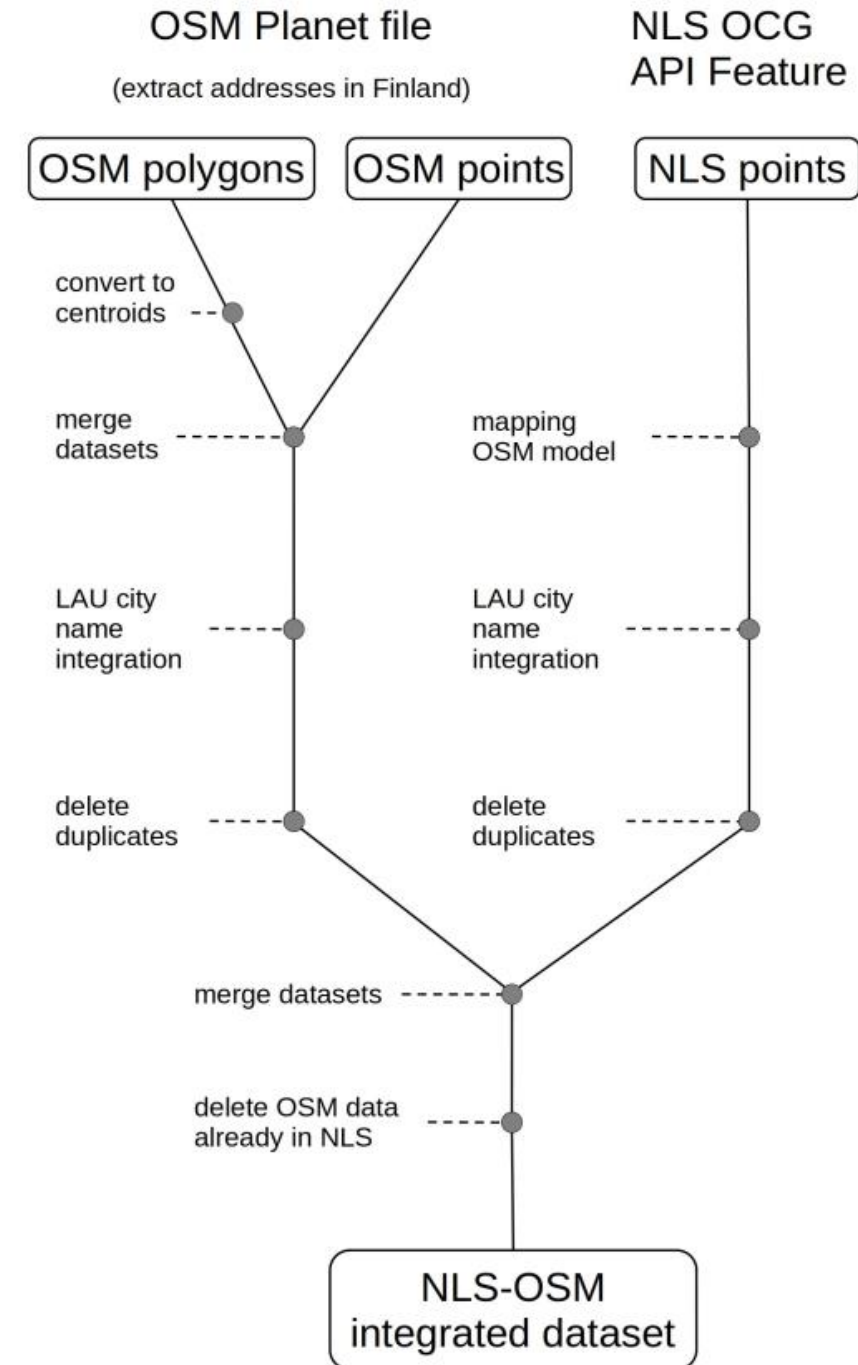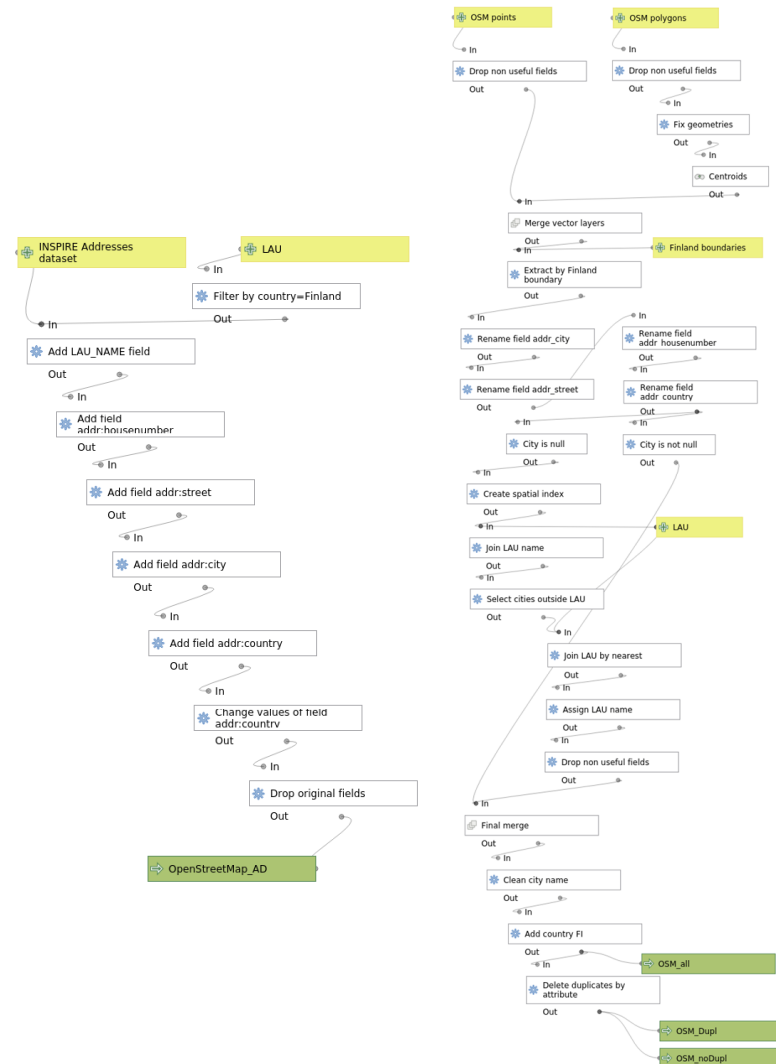
Results

- In original data, higher number of address in NLS than in OSM: 3.3M vs 0.5M
- Uneven geographical distribution of OSM data due to a few imports and mapping efforts
- High density of OSM data mainly in urban areas: high population density → more mappers → more updates
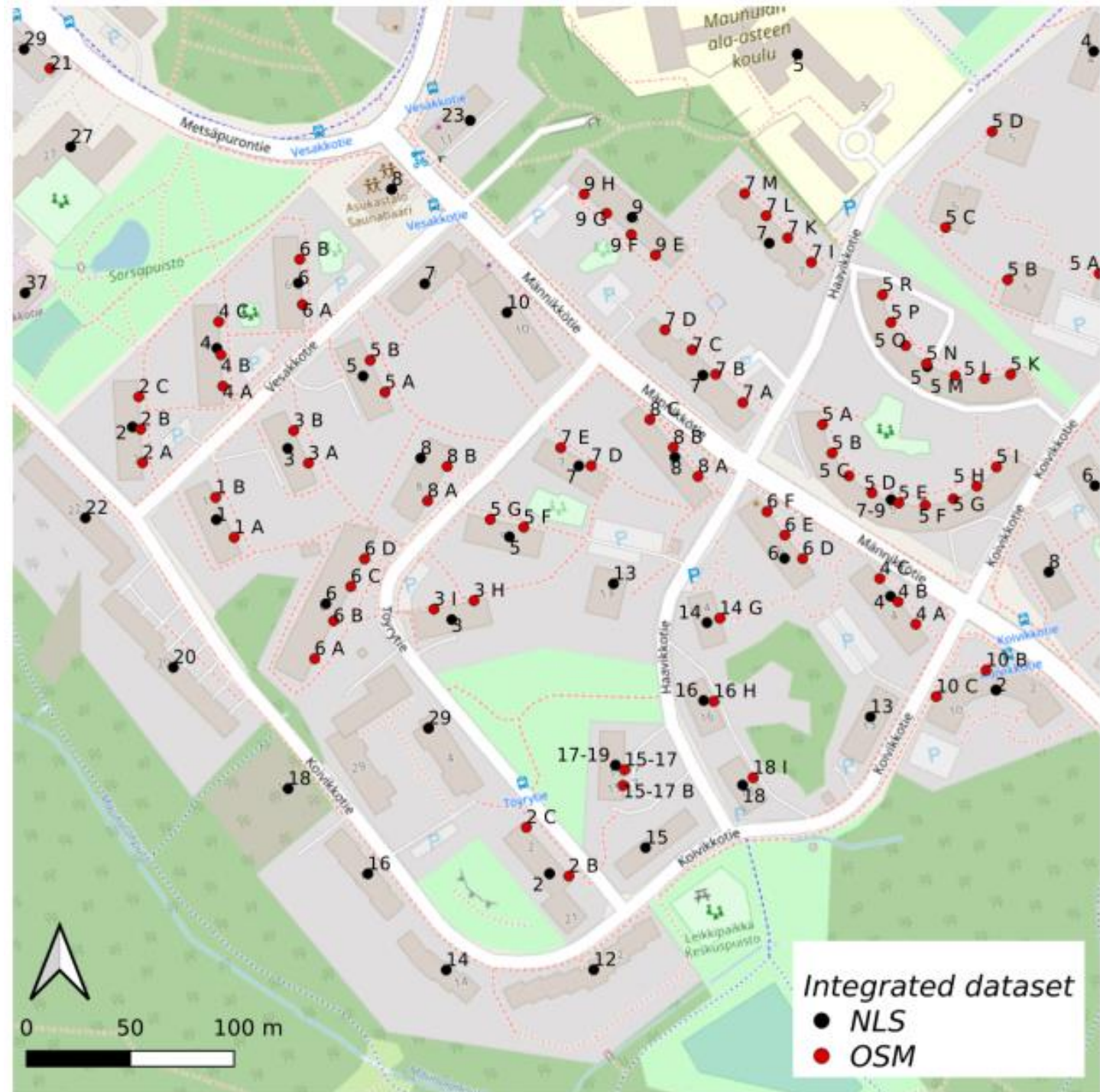
EEA reference grid 10x10km
- empty cells: no addresses both in NLS and OSM
- white cells: no OSM data
- dark cells: OSM data > NLS data

%OSM data compared to NLS
- 0
- < 5
- 5 - 10
- 10 - 50
- 50 - 100
- > 100

0   75   150 km
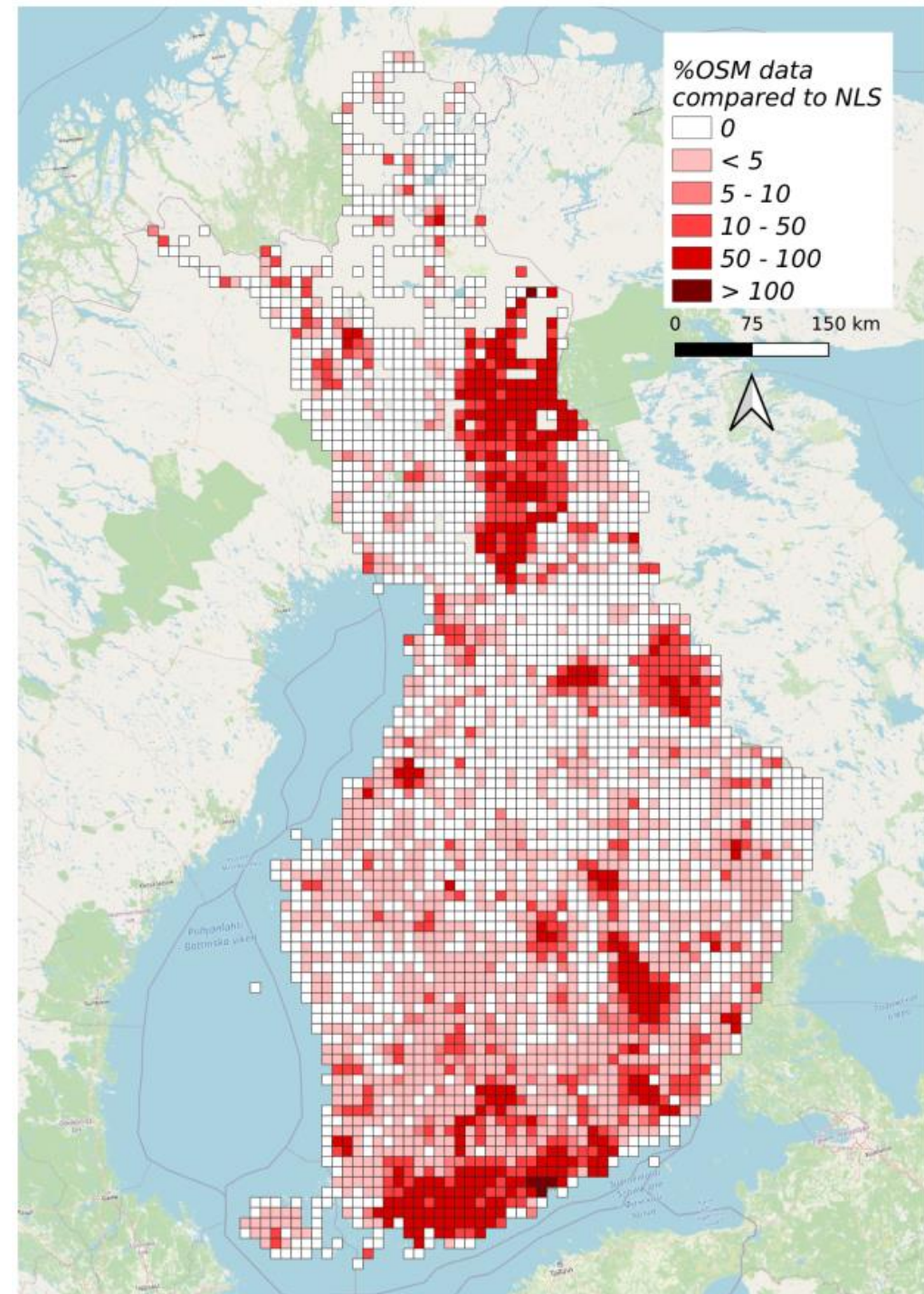
Results

- Integrated data: 96% NLS (only ~80.000 OSM)
- Some OSM data are indeed misspelled street or cities names
- cases where OSM actually includes more detailed or up-to-date information

Licences incompatibilities

Most of governmental data (and geospatial too) comes under the Creative Commons Attribution 4.0 International licence (CC-BY 4.0).

CC-BY 4.0 and ODbL are not fully compatible: one of the main issues discussed during interviews with national and regional mapping authorities.

CC-BY 4.0 → ODbL
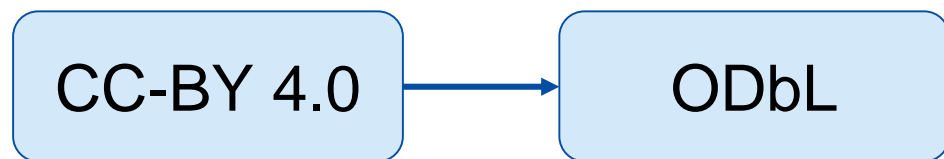
Importing CC-BY 4.0 data in OSM requires an additional explicit permission from licensors (OSMF clarification)

ODbL → CC-BY 4.0
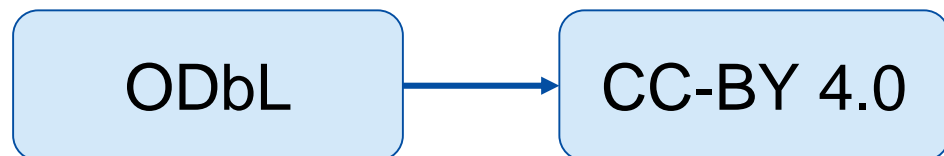
Including OSM data in some dataset, requires to release the integrated data under ODbL

European Commission

Conclusions & next steps

- any data integration process should be carefully prepared
- from a technical perspective, FOSS4G tools enable the integration between the OSM and NLS address datasets and could improve both datasets
- national specificities (in addition to OSM dishomogeneity) can complicate integration: INSPIRE interoperability can ease the application in other countries
- Licences incompatibility is a major problem for two-ways exchange of information between NMAs and OpenStreetMap
- OSM is becoming a more complex ecosystem (citizen, volunteers, governments, private/business companies): this has to be tackled and managed
- First step for wider discussion on permanent collaboration between institutional mapping agencies and OpenStreetMap

European Commission

Paper

TOWARDS THE INTEGRATION OF AUTHORITATIVE AND OPENSTREETMAP GEOSPATIAL DATASETS IN SUPPORT OF THE EUROPEAN STRATEGY FOR DATA

A. Sarretta[a, *], M. Minghini[b]

[a] National Research Council, Research Institute for Geo-hydrological Protection, Padua, Italy - alessandro.sarretta@irpi.cnr.it
[b] European Commission, Joint Research Centre (JRC), Ispra, Italy - marco.minghini@ec.europa.eu

Commission IV, WG IV/4

ABSTRACT:

Digital transformation is at core of Europe's future and the importance of data is well highlighted by the recently published European strategy for data, which envisions the establishment of so-called European data spaces enabling seamless data flows across actors and sectors to ultimately boost the economy and generate innovation. Integrating datasets produced by multiple actors, including citizen-generated data, is a key objective of the strategy. This study focuses on OpenStreetMap (OSM), the most popular crowdsourced geographic information project, and is the first step towards an exploration of pros and cons of integrating its open-licensed data with authoritative geospatial datasets from European National Mapping Agencies. In contrast to previous work, which has only tested data integration at the local or regional level, an experiment was presented to integrate the national address dataset published by the National Land Survey (NLS) of Finland with the corresponding dataset from OSM. The process included the analysis of the two datasets, a mapping between their data models and a set of processing steps—performed using the open source QGIS software—to transform and finally combine their content. The resulting dataset confirms that, while addresses from the NLS are in general more complete across Finland, in some areas OSM addresses provide a higher detail and more up-to-date information to usefully complement the authoritative one. Whilst the analysis confirms that an integration between OSM and authoritative geospatial datasets is technically and semantically feasible, future work is needed to evaluate enablers and barriers that also exist at the legal and organisational level.

https://doi.org/10.5194/isprs-archives-XLVI-4-W2-2021-159-2021

European Commission

Thank you!

✉ alessandro.sarretta@irpi.cnr.it

marco.minghini@ec.europa.eu

European Commission

Keep in touch

EU Science Hub: ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub – Joint Research Centre

EU Science, Research and Innovation

Eu Science Hub