



Evaluation of Novel approaches for governing (location) data and technology.

Combined use of public sector and citizen-generated data

Alessandro Sarretta
Marco Minghini

MIG-T meeting – 8th October 2021

Context of the study

European Strategy for data

- Aims at the creation of a **European single market for data**.
 - a common EU data space and sectoral **data spaces**
- Highlights the problems to address:
 - **data availability** (licensing), **sharing**, **interoperability**, **quality**, **governance**, **infrastructure & technologies**, **skills & data literacy**, **cybersecurity**
- Benefits for 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>

Aim of the study

- 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.

Data sources



- 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



- OpenStreetMap (OSM)

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

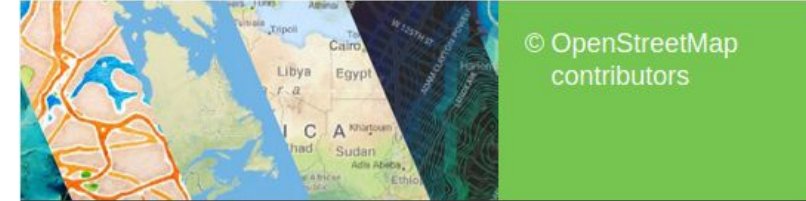
- Dutch cadastre of addresses and buildings - BAG (Basisregistratie Adressen en Gebouwen)

- official national dataset
- open data (CC-BY 4.0 licence)
- complex XML format, non-INSPIRE compliant
- bulk download of processed dataset



OpenStreetMap

- the widest worldwide **collaborative VGI project**
- database of geospatial **open data** ([ODbL](#) licensed)
- based on local knowledge
- community driven
 - over 7 million registered users & 1.6 million active contributors
 - around 50000 active contributors per month & around 120 million changes per month
- uneven geospatial community, contribution and coverage



OpenStreetMap powers map data on thousands of web sites, mobile apps, and hardware devices

OpenStreetMap is built by a community of mappers that contribute and maintain data about roads, trails, cafés, railway stations, and much more, all over the world.

Local Knowledge

OpenStreetMap emphasizes local knowledge. Contributors use aerial imagery, GPS devices, and low-tech field maps to verify that OSM is accurate and up to date.

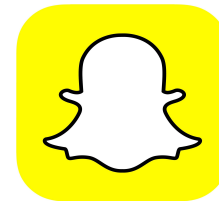
Community Driven

OpenStreetMap's community is diverse, passionate, and growing every day. Our contributors include enthusiast mappers, GIS professionals, engineers running the OSM servers, humanitarians mapping disaster-affected areas, and many more. To learn more about the community, see the [OpenStreetMap Blog](#), [user diaries](#), [community blogs](#), and the [OSM Foundation](#) website.

Open Data

OpenStreetMap is *open data*: you are free to use it for any purpose as long as you credit OpenStreetMap and its contributors. If you alter or build upon the data in certain ways, you may distribute the result only under the same licence. See the [Copyright and License page](#) for details.

Who else is using/contributing to OSM?



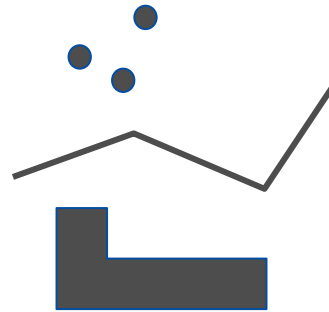
... and many local, national and international authorities/institutions.

Data model: OSM



3 geometry types:

- nodes
- ways
- relations



1 attribute type:

- tags
 - `<key> = <value>`
 - `<key2> = <value>`
 - ...

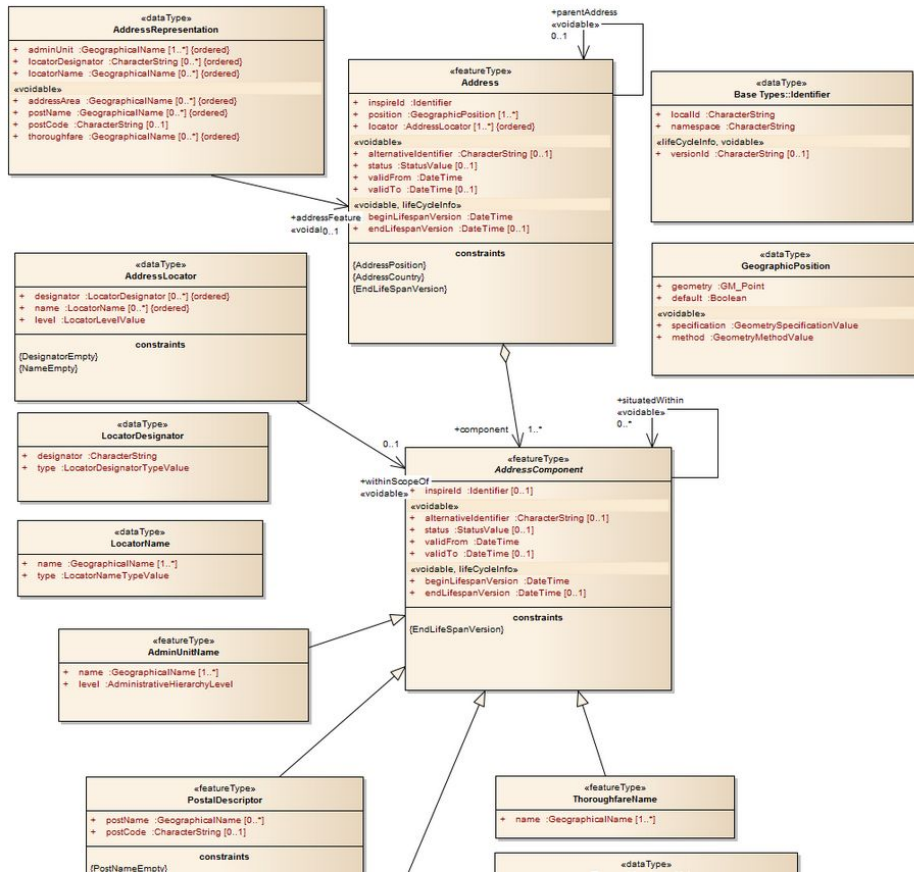
Addresses

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

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

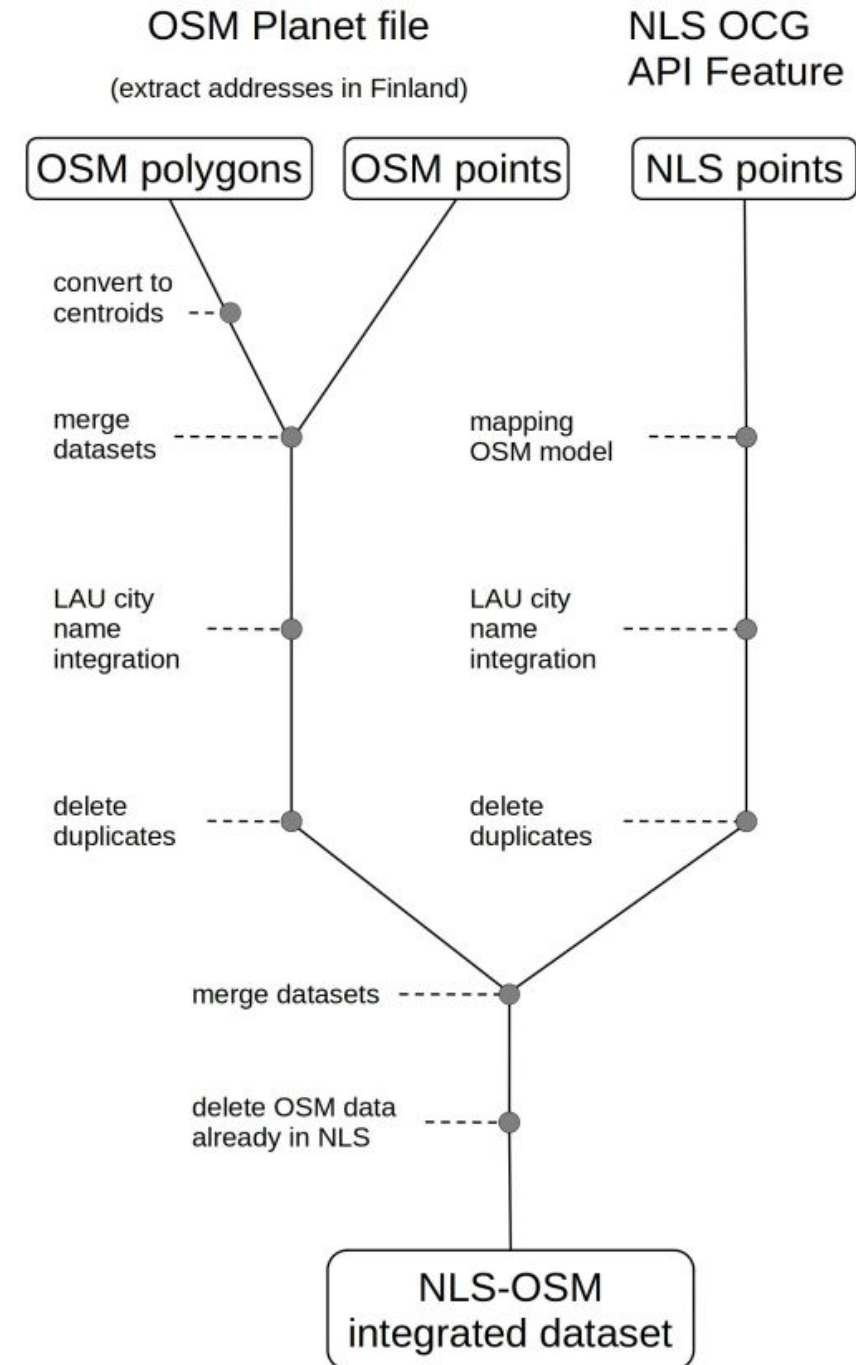
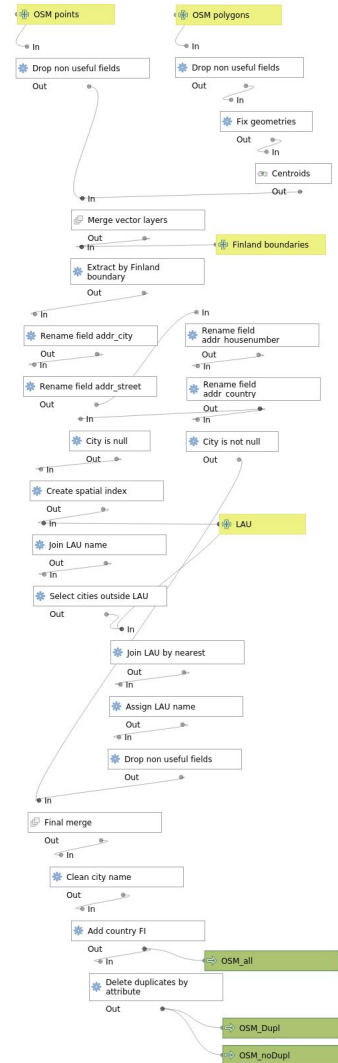
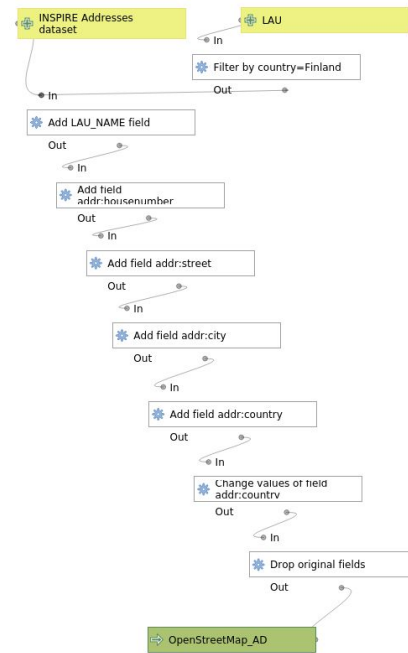
OSM attributes

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](#)

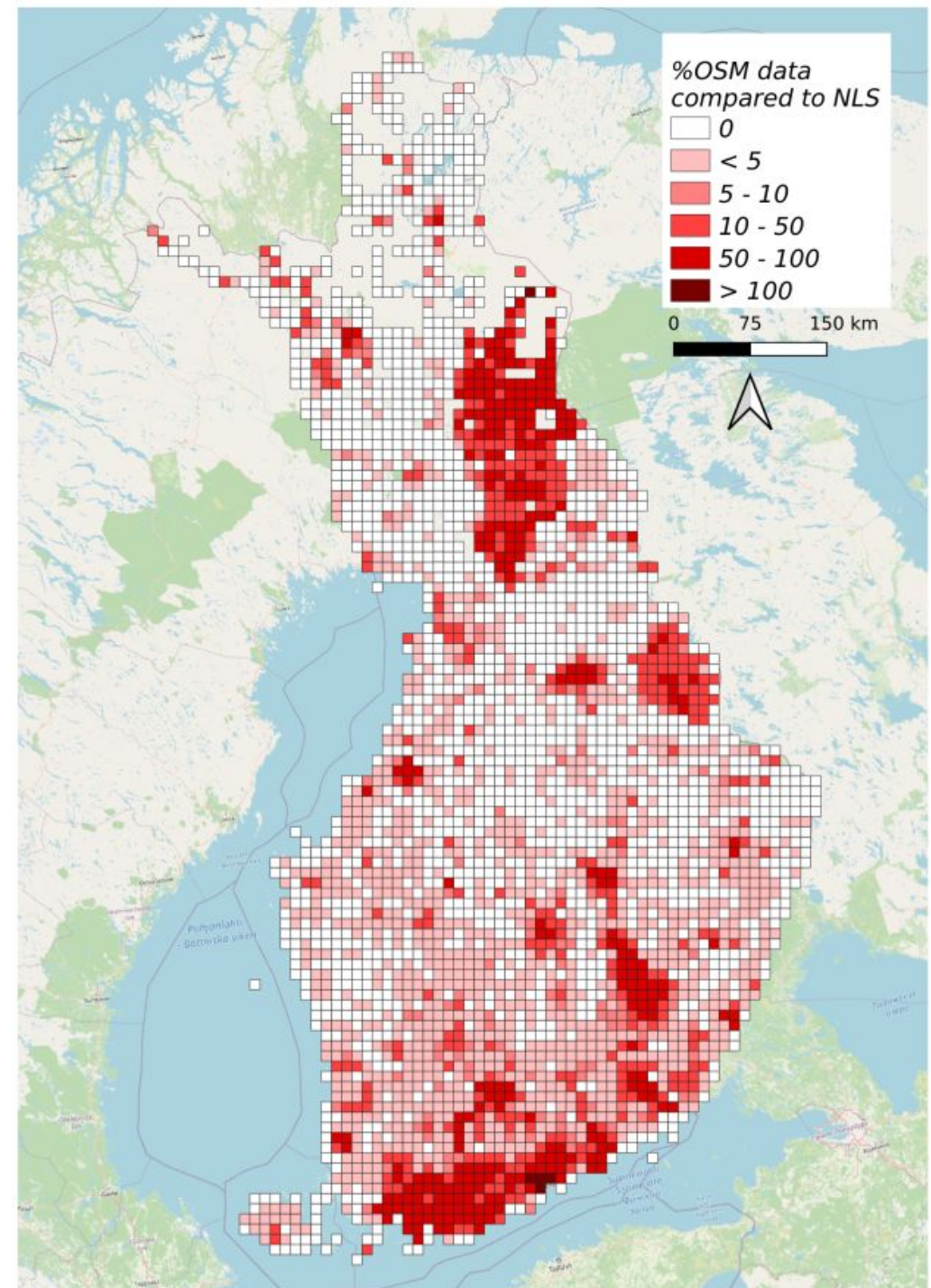


Results

- 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

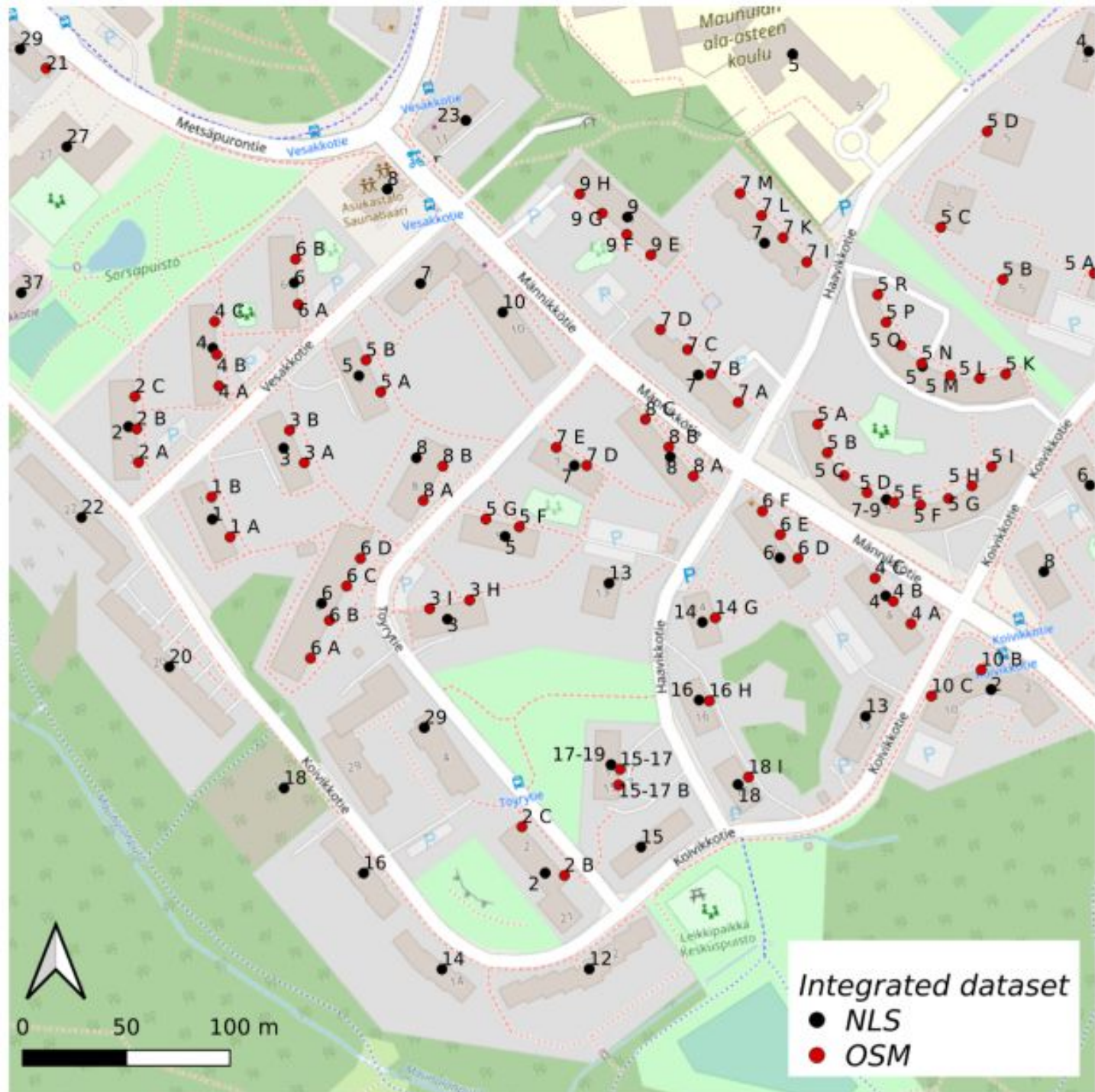
[EEA reference grid](#) 10x10km

- empty cells: no addresses both in NLS and OSM
- white cells: no OSM data
- dark cells: OSM data > NLS data



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



Interviews with experts and stakeholders

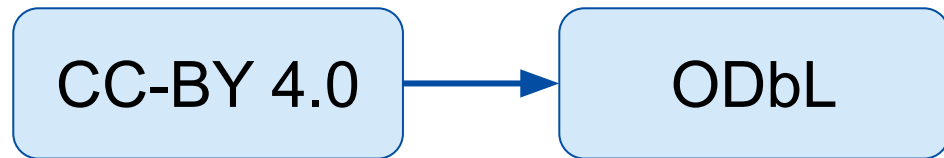
3 interviews with national and regional authorities of Spain, France and Italy on recent experiences in using OSM together with authoritative data:

- **licence incompatibility** is always mentioned as an important barrier in fully reusing and collaborating with OSM data

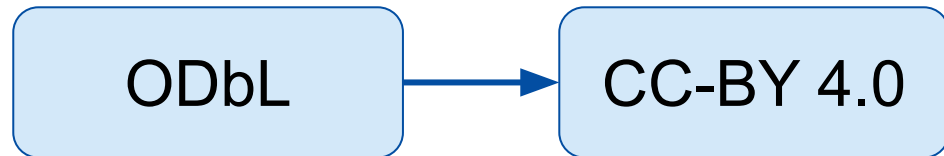
Licences incompatibilities

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



Importing CC-BY 4.0 data in OSM requires an additional explicit permission from licensors ([OSMF clarification](#))



Including OSM data in some dataset, requires to release the integrated data under ODbL

Interviews with experts and stakeholders

3 interviews with national and regional authorities of Spain, France and Italy on recent experiences in using OSM together with authoritative data:

- **licence incompatibility** is always mentioned as an important barrier in fully reusing and collaborating with OSM data
- some **specific topics** (e.g. hydrography, buildings, directions) are of particular interest to add additional information through OSM data
- interest not only in data but also in OSM **tools and technology** both for editing data and allow citizen to contribute
- procedures for automated **comparison of relative completeness** between authoritative/OSM can activate actions to improve both datasets

Interviews with experts and stakeholders

- one of the main goals for official cartographic authorities is that **citizens** should be **able to easily access and use authoritative data** (which costs public money to collect and maintain); a key means of disseminating and allowing people to use that data would be to include it in a global, open and widely used project and database such as OpenStreetMap.

Conclusions & next steps

- any data integration process should be carefully prepared
- from a technical perspective, the integration between the OSM and NLS address datasets is feasible (e.g through simple FOSS4G tools) 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

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

KEY WORDS: Citizen-generated data, Europe, Interoperability, National Mapping Agencies, OpenStreetMap, Spatial Data Infrastructures

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.

Thank you!



alessandro.sarretta@gmail.com



The authors acknowledge the support of the European Commission - Joint Research Centre (JRC) through contract number CT-EX2013D133588-101, entitled “Evaluation of Novel approaches for governing (location) data and technology. Combined use of public sector and citizen-generated data”.

© European Union 2020

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.



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