Towards a comparison of pan-European open building data

Sara Thabit Gonzalez, Marco Minghini, Lorenzo Gabrielli, Patrizia Sulis

European Commission, Joint Research Centre (JRC), Ispra, Italy

JRC Human Planet Forum – 21 November 2025

Context

- Building footprints are **key geospatial datasets** for several use cases
  - city planning, demographic analyses, modelling energy production/consumption, disaster preparedness/response, digital twins
- Buildings have been **traditionally produced, curated & updated by governmental organisations**
  - National Mapping/Cadastral Agencies responsible for Spatial Data Infrastructures (SDIs)
- Recent technological trends have seen **other players become producers** of building datasets
  - private sector, research/academia, citizen-generated data initiatives

Study Objectives

- **Identify** existing building datasets:
  - from governmental and **non-governmental organisations**
  - available (at least) at the **continental scale**, with a focus on the EU
  - accessible under an **open license**

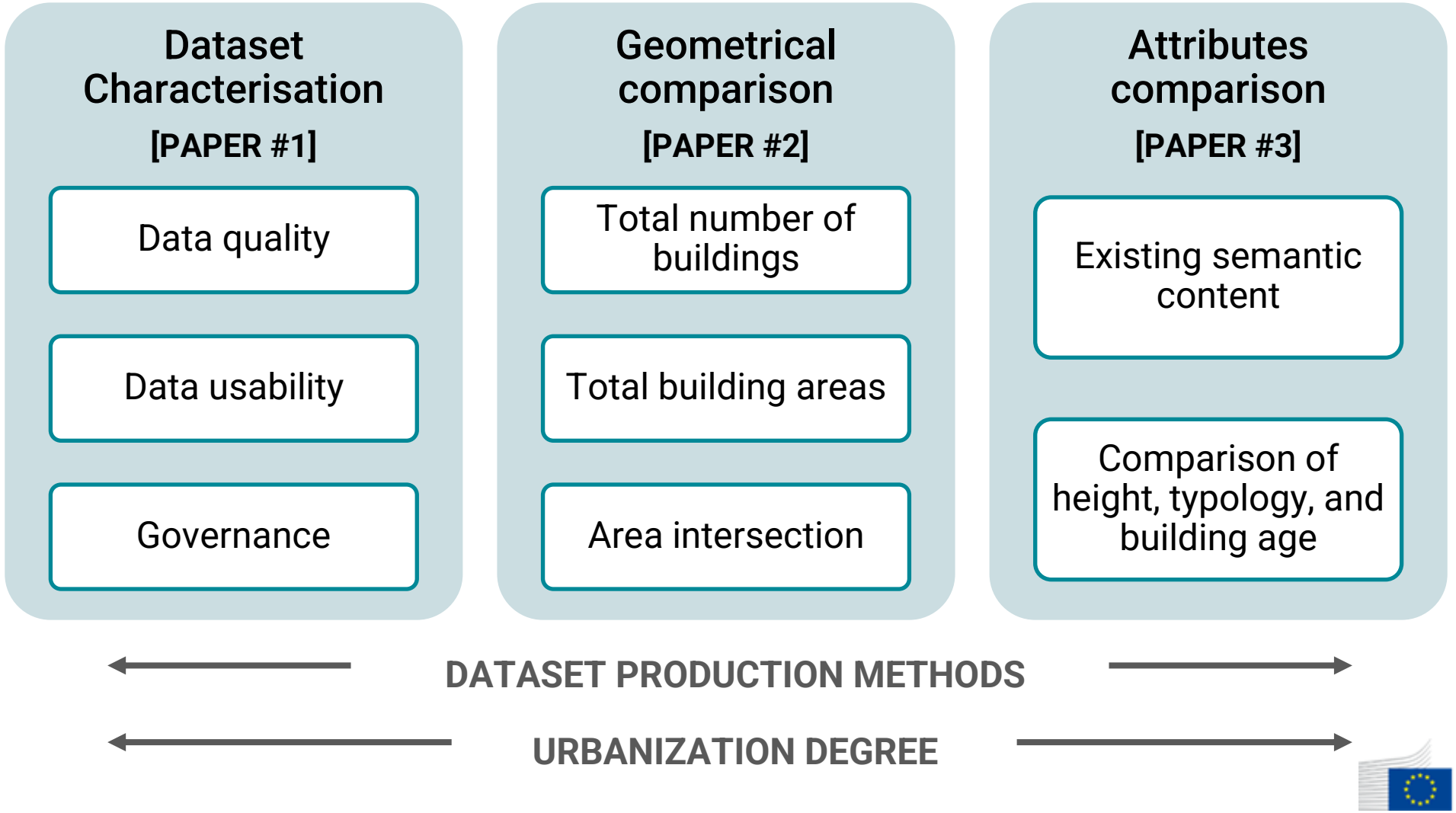
- **Analyse** and **compare** the building datasets
  - characterisation based on a self-developed assessment framework
  - comparison through quantitative analyses

- Derive insights on **implications** for policy-relevant use cases
  - which dataset(s) can better address which policy need

Open building footprints

	PUBLIC SECTOR-LED	COMMUNITY-LED	INDUSTRY-LED			RESEARCH-LED		
Dataset	INSPIRE & Open Data Directive	OpenStreetMap (OSM)	Microsoft Global ML Building Footprints	Google Open Buildings	Overture Maps	EUBUCCO	JRC Digital Building Stock Model (DBSM)	Global Open Building Attribute Table (GHS-OBAT)
Data production process	Free and open governmental datasets	Volunteered geographic information (VGI)	ML algorithm on satellite imagery	ML algorithm on satellite imagery	Integration of OSM, Microsoft and Google datasets	Integration of open governmental datasets and OSM	Integration of OSM, Microsoft, and ESM R2020	ML algorithm on satellite imagery
Geographic coverage	EU	Global	Global (almost)	Latin America, Africa, South-east Asia	Global	EU & Switzerland	EU	Global

Study approach



1. Buildings data characterisation

- **Assessment framework**
  - inspired by & adapted from existing research
  - composed of **3 dimensions** and **13 attributes**

1. DATA QUALITY	2. DATA USABILITY	3. GOVERNANCE
Attributes	***Attributes***	***Attributes***
Geographic coverage and completeness	Data findability	Organisational structure
Granularity, shape and positional accuracy	Data accessibility	Business model and sustainability
Timeliness	Data interoperability and manageability	Openness, transparency and reproducibility
Semantic content	Licence and reusability	Community engagement
Quality assurance mechanisms and indicators		

1. Buildings data characterisation

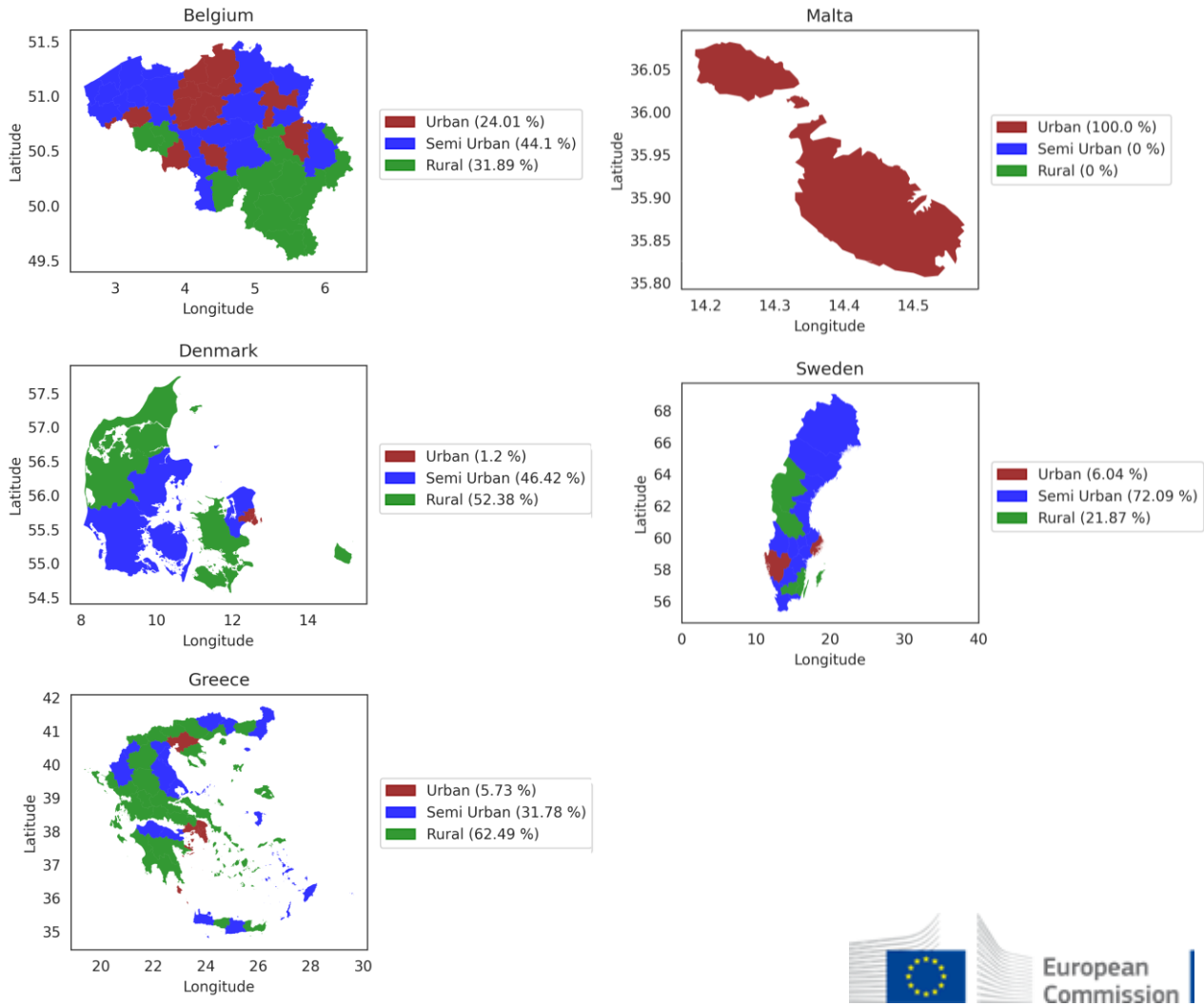
		PUBLIC SECTOR-LED	COMMUNITY-LED	INDUSTRY-LED			RESEARCH-LED	
		INSPIRE & Open Data Directive	**OpenStreetMap (OSM)**	**Microsoft GlobalML Building Footprints**	**Google Open Buildings**	**Overture Maps**	**JRC Digital Building Stock Model (DBSM)**	**EUBUCCO**
Data quality	*Geographic coverage and completeness*	Typically high	Influenced by demographic, socioeconomic, and contributor factors	Influenced by image resolution or distortions.	Influenced by image resolution or distortions.	Relatively higher than non-conflated datasets	Relatively high completeness compared with cadastral data	High variations per country, influenced by the dataset used
	Granularity, shape and positional accuracy	Tipically high	Influenced by demographic, socioeconomic, and contributor factors	Influenced by the specific LM algorythm and buildings characteristics	Influenced by the specific LM algorythm and buildings characteristics	*Insufficient existing comparative analysis*	Relatively high completeness compared with cadastral data	High variations per country, influenced by the dataset used
	Timeliness	Periodic releases by country	Constant/Real-time updates	Monthly/Bimonthly releases	Monthly/Bimonthly releases	Monthly/Bimonthly releases	First release Oct/2023, second June/2025	Single release (Nov/2022)
	Semantic content	Different list of attributes per country, comprehensive	High number of attributes, partial coverage	Only building height, partial coverage	No semantic content	Multiple, imported from secondary sources	Height, use and building age	Height, use, and building age. Different coverage depending on the country/dataset
	Accuracy indicators and quality assurance mechanisms	Standardised data collection protocols and audits	Community of peers & academic analysis	AI-developed "Confidence score" per building	AI-developed "Confidence score" per building	No	No	No
Data usability	*Data Findability*	Different access tools depending on national portals	Highest variety of pre-visualization and search tools and platforms	No pre-visualization, only API search	Pre-visualization and search tools	Pre-visualization, but search tools only with account	No pre-visualization or search tools	No pre-visualization, only search tools
	Data Accessibility		Highly accessible by different user profiles	Limited to skilled users	Accessible by average user	Accessible by average user	Limited to high computational power	Accessible by average user
	Data Interoperability and Manageability	No cross-border harmonization; standard GIS formats	Harmonized data; highest variety of data formats	Harmonized data. Export only as .csv.gz	Harmonized data. Exports as .csv and .geojson	Harmonized data. Exports only as GeoParquet	Harmonized data. Exports as RDF/JSON, RDF/XML, Turtle formats.	Harmonized data. Exports in .gpkg and .csv
	Licence and reusability	CC0, CC BY-4.0 or equivalent or less restrictive license	ODbL	ODbL	CC BY-4.0 and ODbL	CDLA except from data derived from OSM (ODbL)	ODbL	ODbL except of 2 databases
Governance	*Organisational structure*	Hybrid (Top-down & bottom-up)	Bottom-up (open)	Top-down (closed)	Top-down (closed)	Top-down (open via annual fee)	Top-down (closed)	Top-down (closed)
	Business model and sustainability	Public sector funded	Foundation funding & voluntary mapping	Corporate funding	Corporate funding	Foundation funding	Project funding (limited timing)	Project funding (limited timing)
	Openness, transparency and reproducibility	Transparent and reproducible production process	Transparent and reproducible production process	Insufficient information for reproduction or process analysis	Insufficient information for reproduction or process analysis	Insufficient information for reproduction or process analysis	Insufficient information for reproduction or process analysis	Transparent and reproducible production process
	Community engagement	EU-led community of public sector collaborators	Very large open community	Not specified	Not specified	Payment-based membership tiers	Limited community of researchers	Limited community of researchers

2. Geometrical comparison

Focused on:

- **total area**
- **total number of buildings**
- **geometric intersections**

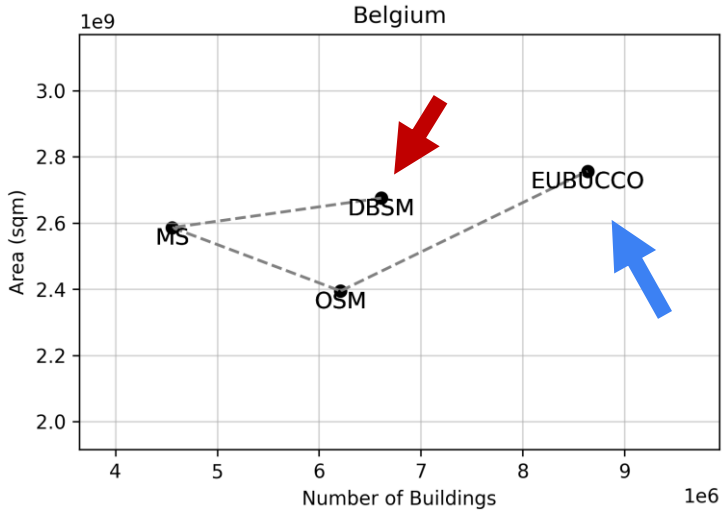
- considering the **degree of urbanisation**
- limited to **4 building datasets**: OSM, EUBUCCO, MS, DBSM
- limited to **5 EU countries**: Belgium, Denmark, Greece, Malta, Sweden

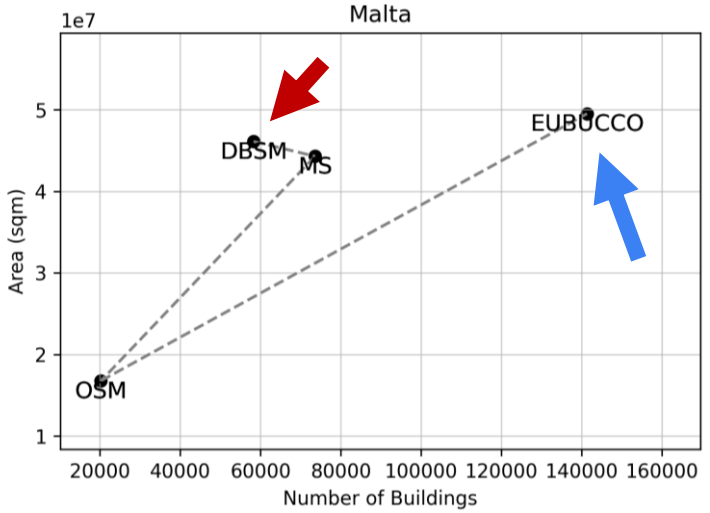


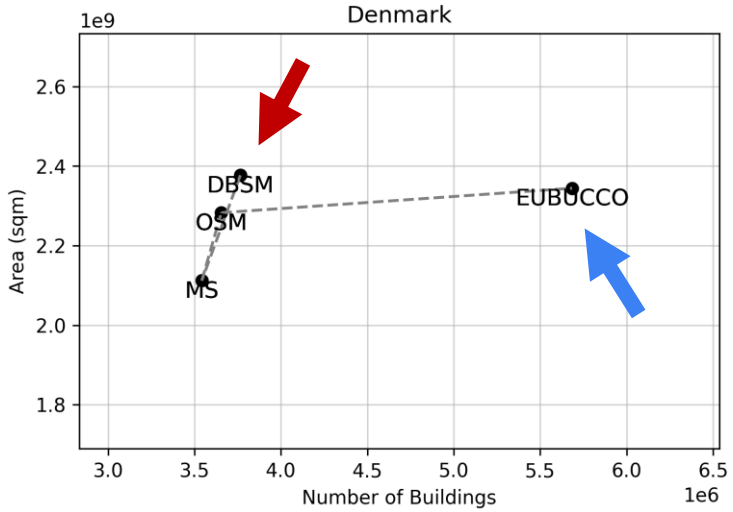
European Commission

2. Geometrical comparison

Dataset	Country	Number of buildings	Area of buildings [10^8 m^2]
EUBUCCO	Belgium	8,636,114	27.56
	Denmark	5,684,734	23.44
	Greece	856,140	3.04
	Malta	141,329	0.49
	Sweden	2,504,961	21.38
OSM	Belgium	6,211,451	23.94
	Denmark	3,654,875	22.82
	Greece	1,217,547	4.71
	Malta	20,225	0.16
	Sweden	3,050,667	24.84
DBSM	Belgium	6,610,034	26.75
	Denmark	3,765,255	23.76
	Greece	4,540,228	15.03
	Malta	58,247	0.46
	Sweden	4,936,573	36.16
MS	Belgium	4,557,403	25.86
	Denmark	3,541,845	21.11
	Greece	5,722,750	14.74
	Malta	73,579	0.44
	Sweden	6,422,594	31.07







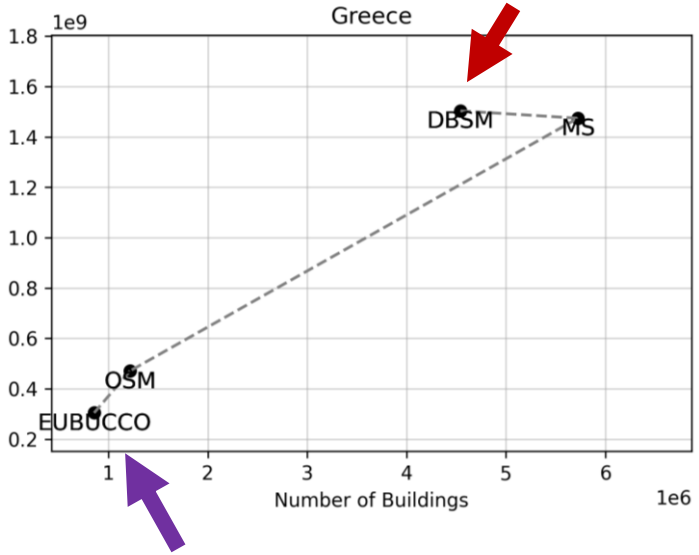
→ Authoritative data (public sector)

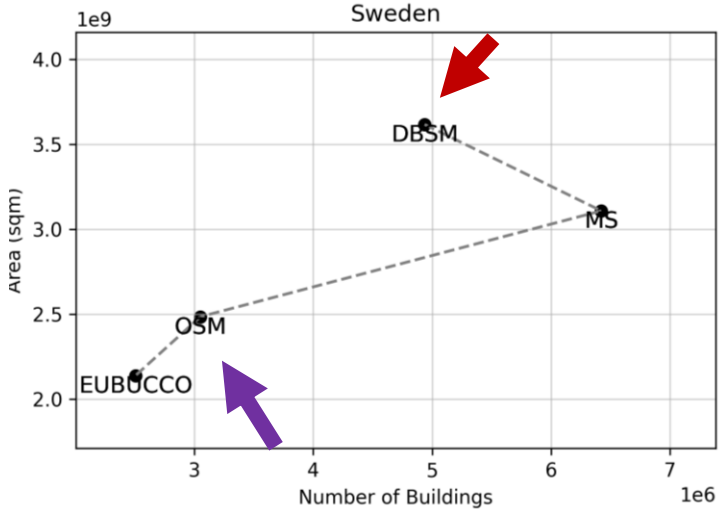
→ Conflation strategy (OSM and MS)

European Commission

2. Geometrical comparison

Dataset	Country	Number of buildings	Area of buildings [10^8 m^2]
EUBUCCO	Belgium	8,636,114	27.56
	Denmark	5,684,734	23.44
	Greece	856,140	3.04
	Malta	141,329	0.49
	Sweden	2,504,961	21.38
OSM	Belgium	6,211,451	23.94
	Denmark	3,654,875	22.82
	Greece	1,217,547	4.71
	Malta	20,225	0.16
	Sweden	3,050,667	24.84
DBSM	Belgium	6,610,034	26.75
	Denmark	3,765,255	23.76
	Greece	4,540,228	15.03
	Malta	58,247	0.46
	Sweden	4,936,573	36.16
MS	Belgium	4,557,403	25.86
	Denmark	3,541,845	21.11
	Greece	5,722,750	14.74
	Malta	73,579	0.44
	Sweden	6,422,594	31.07





→ No Authoritative data

→ Conflation strategy (OSM and MS)

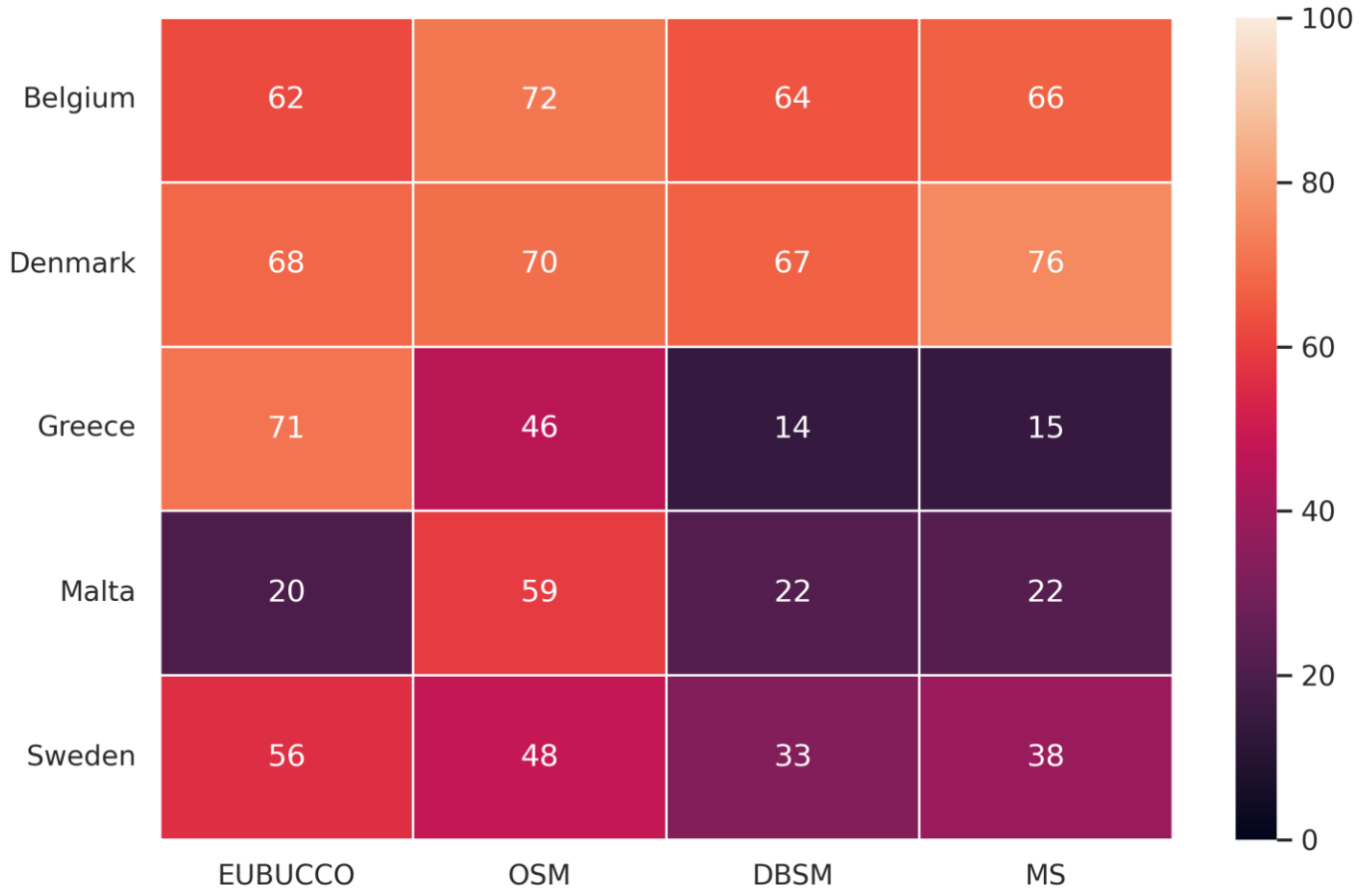
→ OSM data

European Commission

2. Geometrical comparison

Intersection between all the 4 datasets

- % of the area of each dataset represented by the area of intersection between the 4 datasets
- % **lower in rural areas** (minimum 7%) and **higher in urban areas** (maximum 79%)



3. Attributes comparison

Dataset	Spatial coverage	Semantic information available (relevant for our study)
OpenStreetMap	Global	• Height • Typology • Building age • Number of floors • Building material
EUBUCCO	Continental (EU & Switzerland)	• Height • Typology • Building age
Microsoft Global ML Building Footprints	Global (almost)	• Height
Overture Maps Foundation	Global	• Height • Typology • Building age • Number of floors • Building material
Digital Building Stock Model (DBSM) 2025	Continental (EU countries)	• Height • Typology • Building age
Global Human Settlement Building Attributes (GHS-OBAT)	Global	• Height • Typology • Building age

All EU countries

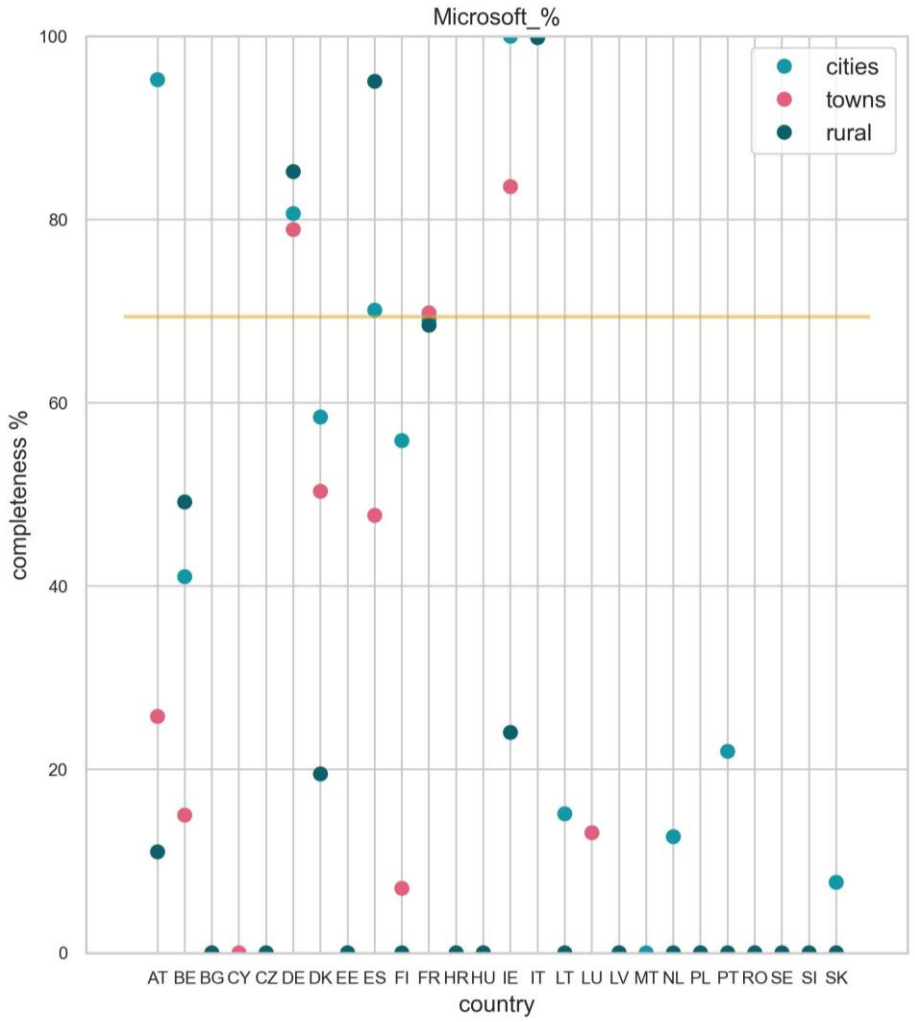
Considering the **degree of urbanization**

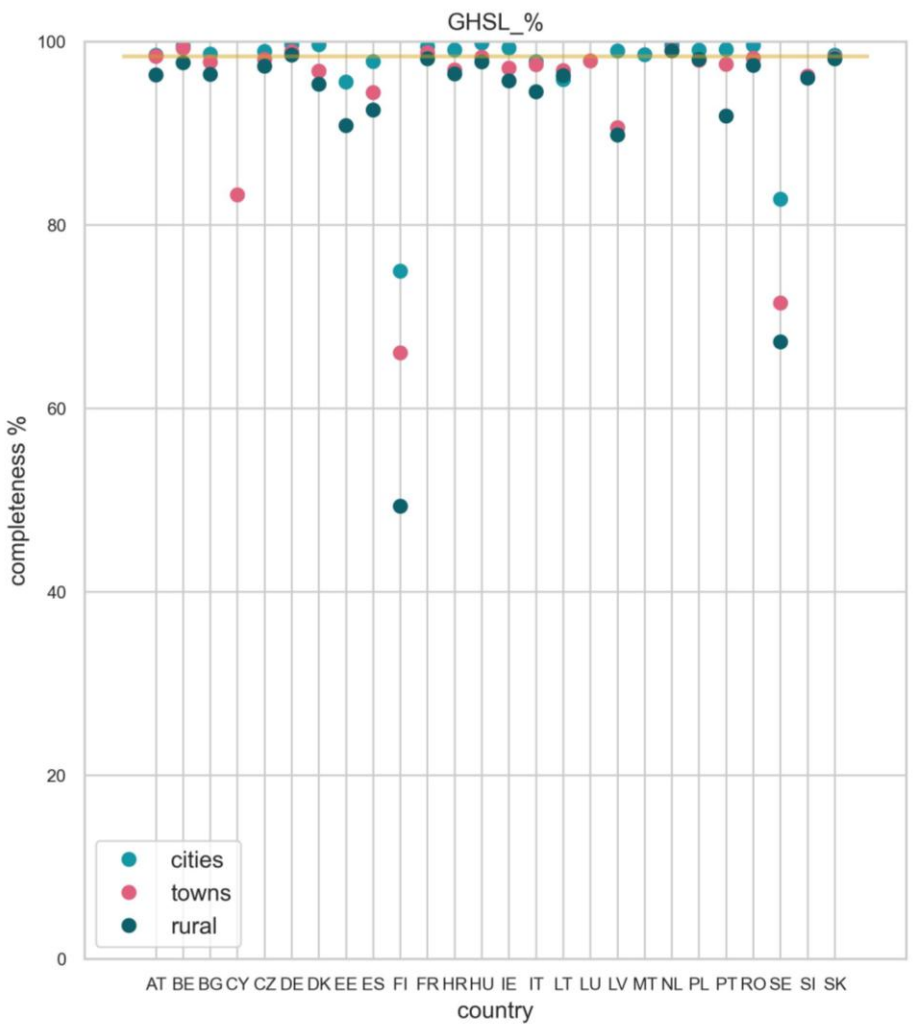
Methodology at a glance

- map/**harmonise attributes** across datasets
- for each dataset/country/attribute, **calculate the fraction of buildings with values,** including by degree of urbanization
- for each dataset/country/attribute, **derive statistical distributions of values**
- assess the **similarity of distributions** by country and dataset

European Commission

3. Comparison of attribute completeness

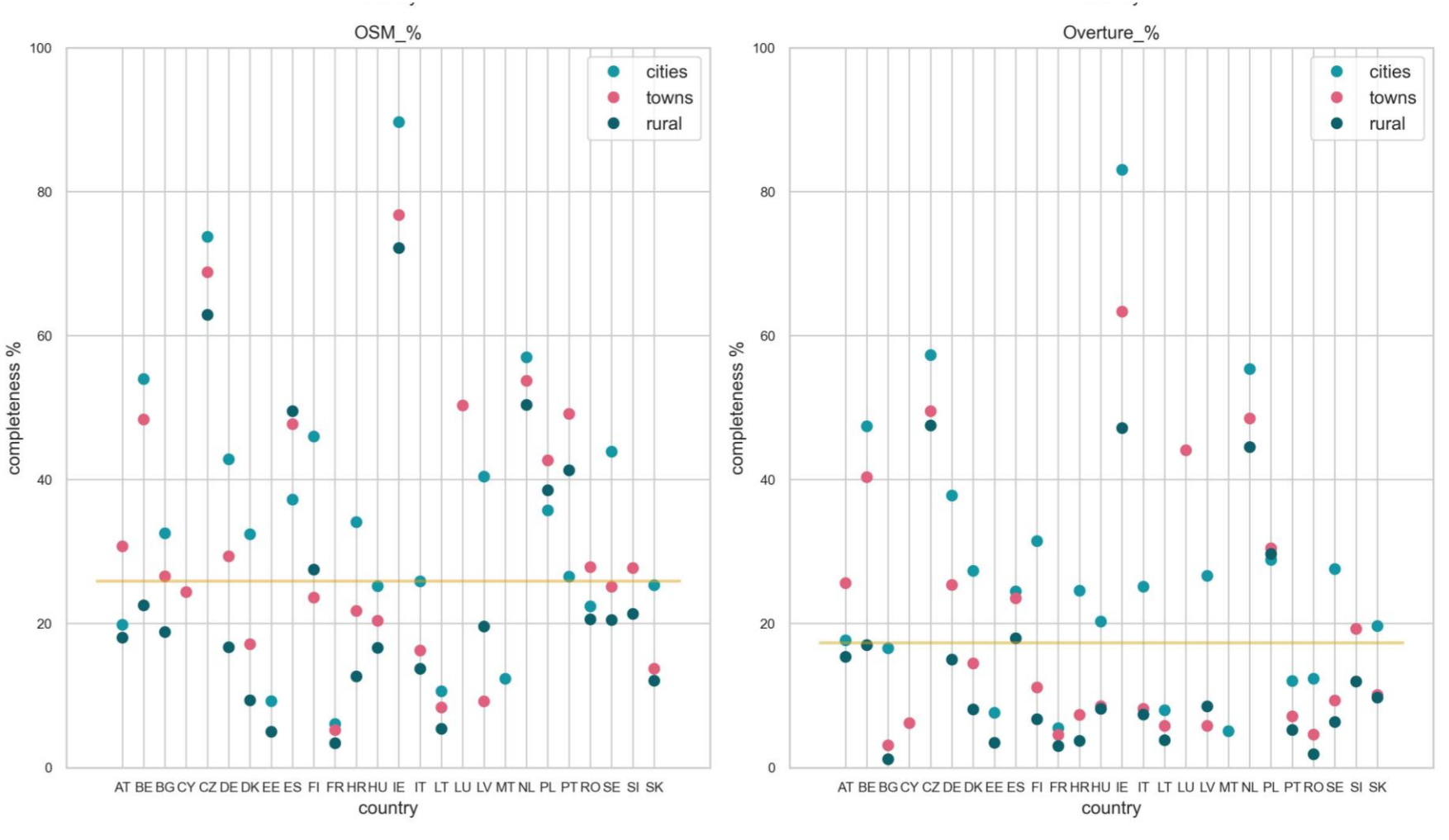




'Height' attribute completeness across 27 countries & degrees of urbanization

European Commission

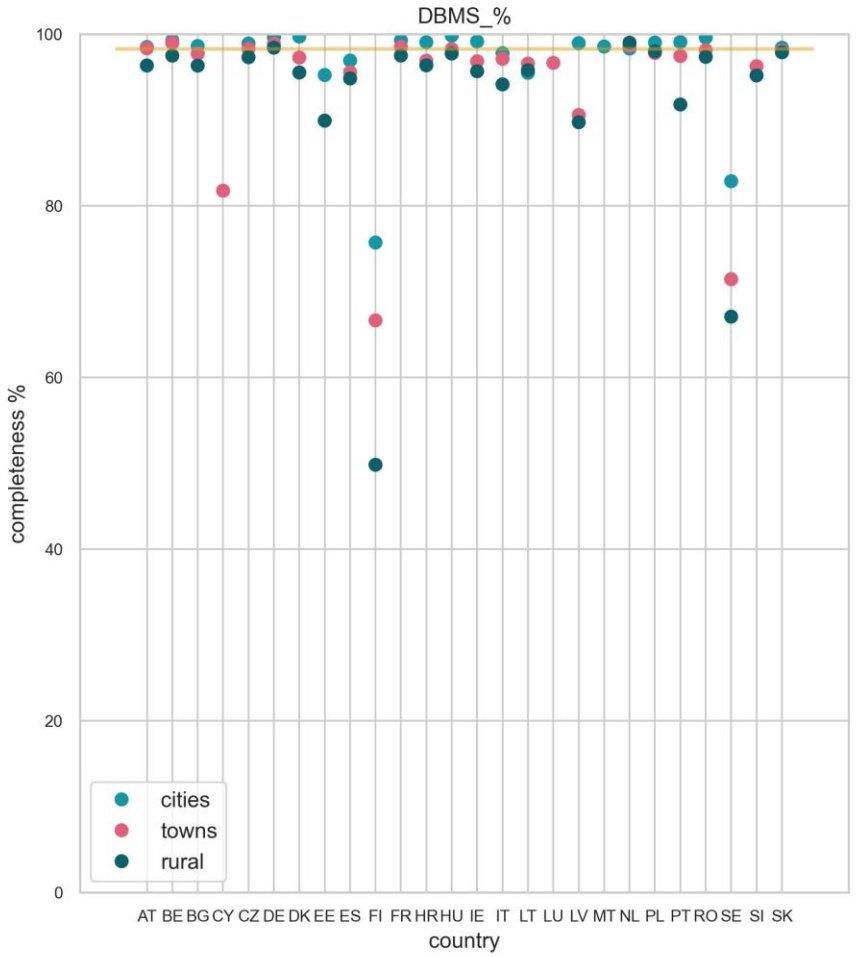
3. Comparison of attribute completeness

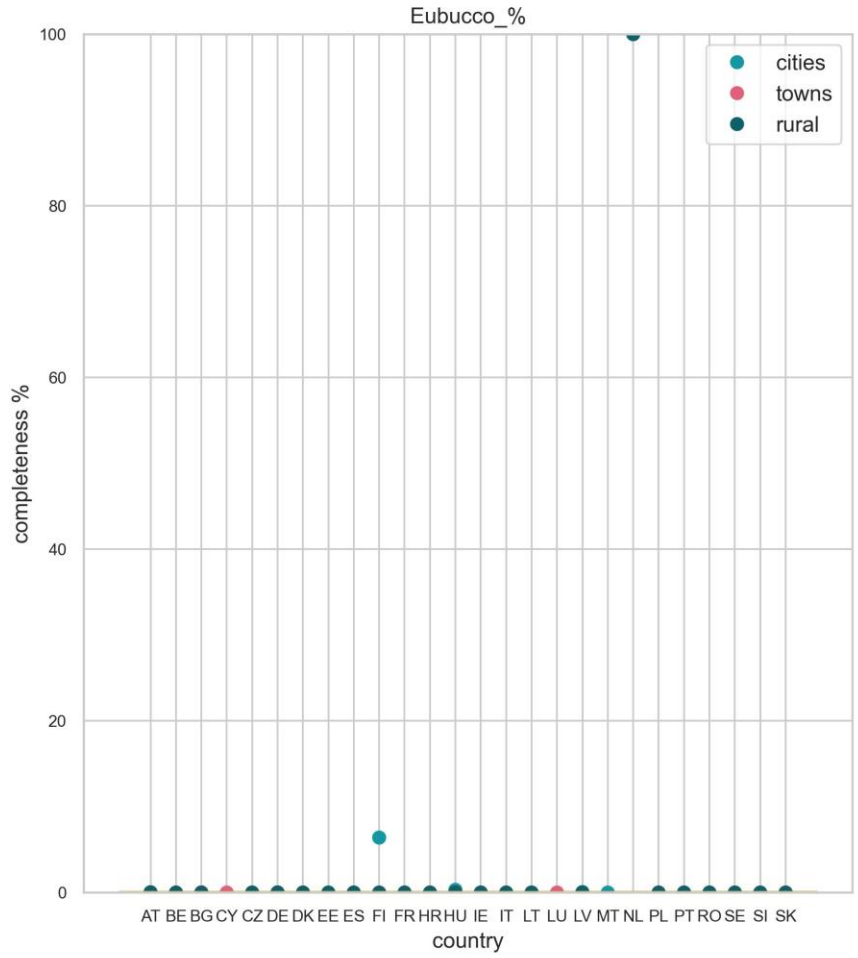


‘Typology’ attribute completeness across 27 countries & degrees of urbanization

European Commission

3. Comparison of attribute completeness





‘Building age’ attribute completeness across 27 countries & degrees of urbanization

European Commission

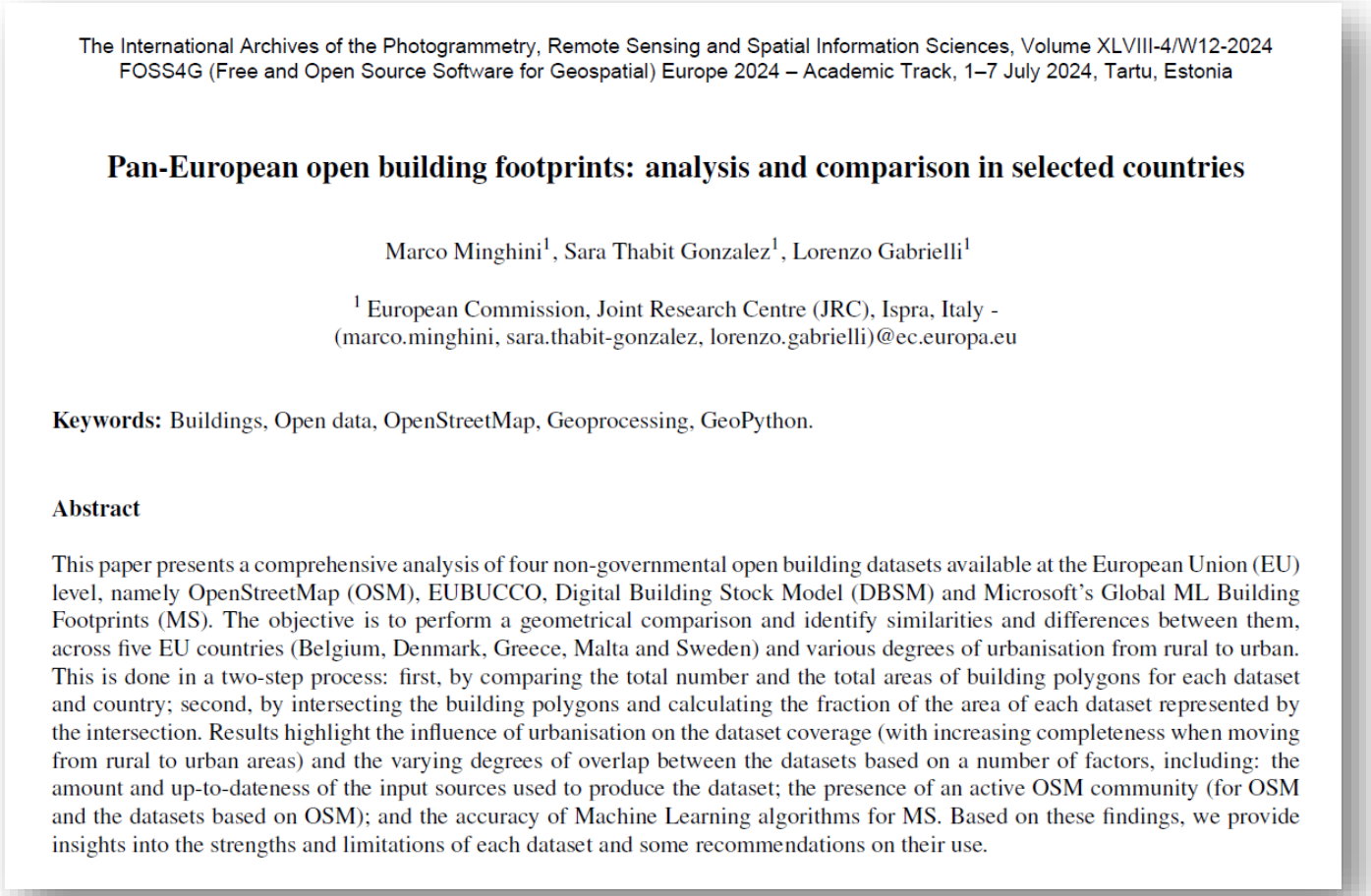
Conclusions

Discussion

- First comparison of (some) non-governmental open building datasets
- **Relative comparison** (not quality assessment) of datasets is the way to go
  - **different sources**: governments, citizens, private companies, research institutions
  - **different production/update approaches**: digitalisation, machine learning, conflation
- There is **no best dataset** in general – the choice depends on the specific area & use case/application
- **Conflation** of multiple datasets is a promising approach
- Datasets quickly become **outdated** without regular releases
- **Governance** (transparency, inclusivity & sustainability) considerations
- How will **production of open data** look like in the future?
  - non-governmental initiatives hold potential to **enhance public sector data**

References

- Paper: https://isprs-archives.copernicus.org/articles/XLVIII-4-W12-2024/97/2024



The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W12-2024
FOSS4G (Free and Open Source Software for Geospatial) Europe 2024 – Academic Track, 1–7 July 2024, Tartu, Estonia

Pan-European open building footprints: analysis and comparison in selected countries

Marco Minghini[1], Sara Thabit Gonzalez[1], Lorenzo Gabrielli[1]

[1] European Commission, Joint Research Centre (JRC), Ispra, Italy -
(marco.minghini, sara.thabit-gonzalez, lorenzo.gabrielli)@ec.europa.eu

Keywords: Buildings, Open data, OpenStreetMap, Geoprocessing, GeoPython.

Abstract

This paper presents a comprehensive analysis of four non-governmental open building datasets available at the European Union (EU) level, namely OpenStreetMap (OSM), EUBUCCO, Digital Building Stock Model (DBSM) and Microsoft's Global ML Building Footprints (MS). The objective is to perform a geometrical comparison and identify similarities and differences between them, across five EU countries (Belgium, Denmark, Greece, Malta and Sweden) and various degrees of urbanisation from rural to urban. This is done in a two-step process: first, by comparing the total number and the total areas of building polygons for each dataset and country; second, by intersecting the building polygons and calculating the fraction of the area of each dataset represented by the intersection. Results highlight the influence of urbanisation on the dataset coverage (with increasing completeness when moving from rural to urban areas) and the varying degrees of overlap between the datasets based on a number of factors, including: the amount and up-to-dateness of the input sources used to produce the dataset; the presence of an active OSM community (for OSM and the datasets based on OSM); and the accuracy of Machine Learning algorithms for MS. Based on these findings, we provide insights into the strengths and limitations of each dataset and some recommendations on their use.

Thank you!

sara.thabit-gonzale@ec.europa.eu

EU Science Hub

joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub – Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

@eu_science