Volunteered Geographic Information and OpenStreetMap: governance, applications and perspectives

Marco Minghini

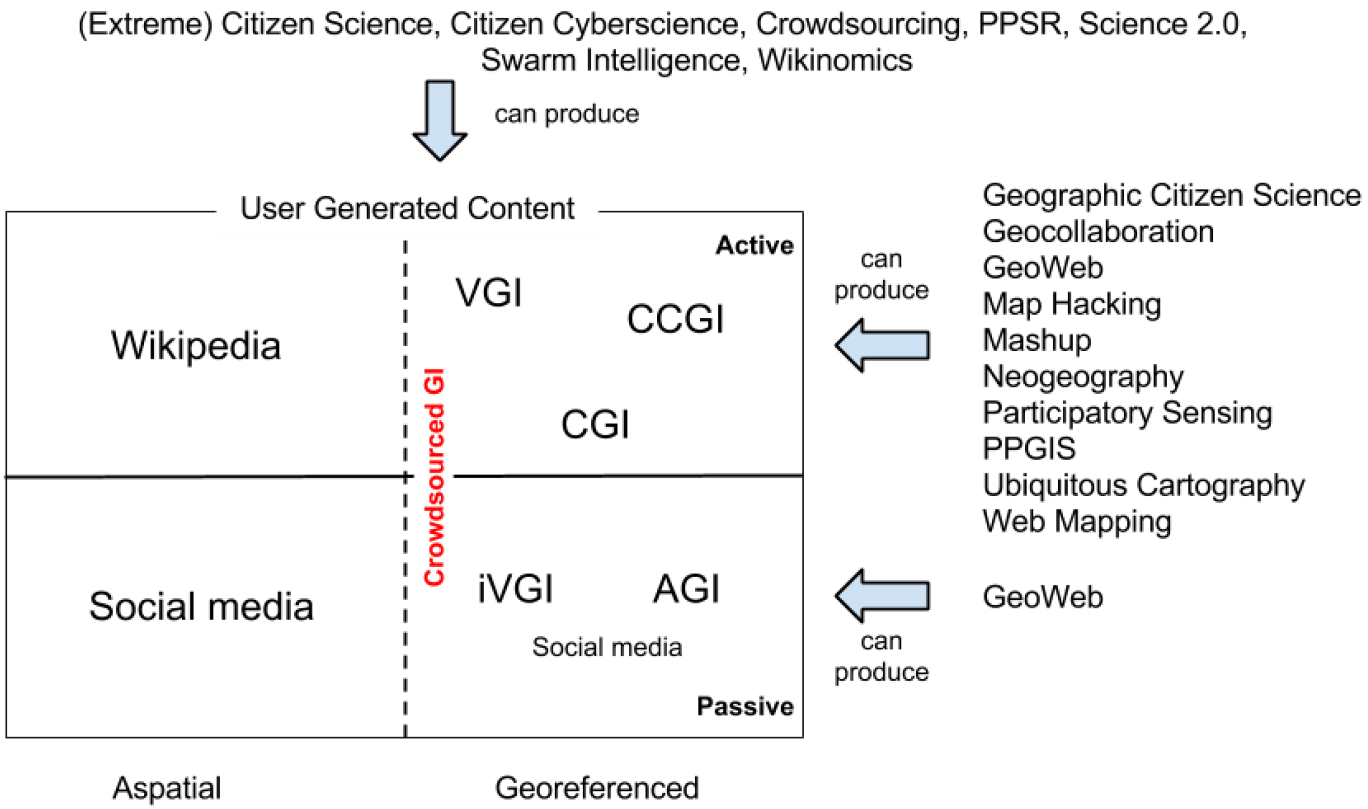
The European Commission's science and knowledge service

Joint Research Centre

Ispra, 18/01/2019

European Commission

User-generated geospatial content



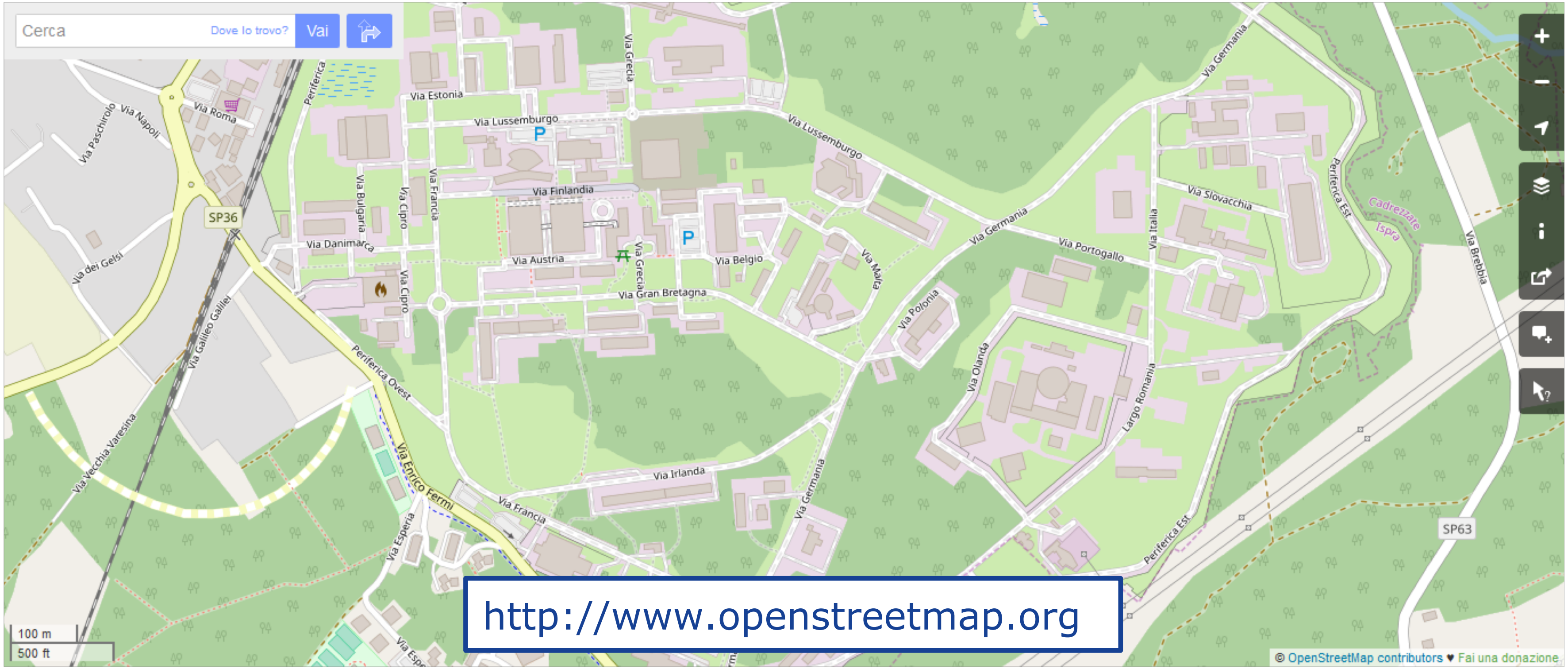
Source: See, L et al. (2016). Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information*, 5(5), 55.

Traditional (authoritative) mapping vs. VGI

Traditional mapping	VGI
official agencies	people
professional expertise	no specific skills
centralized	distributed
top-down approach	bottom-up approach
expensive	almost free
checked for quality	not verified
authoritative	asserted
global information	local information
slow, not updated	fast, updated

OpenStreetMap (OSM) – What is it?

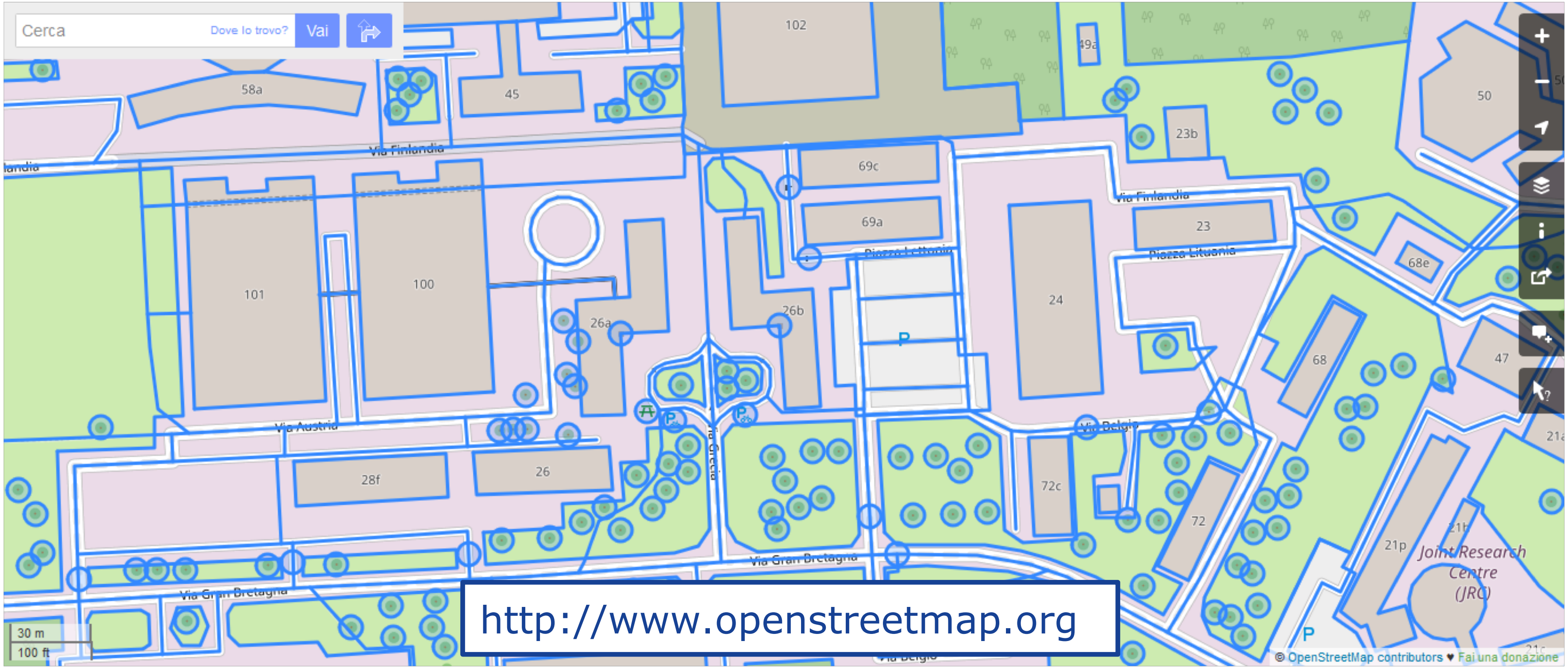
- The most popular VGI project, started in 2004
  - a free, editable vector map of the whole world built by volunteers



http://www.openstreetmap.org

OpenStreetMap (OSM) – What is it?

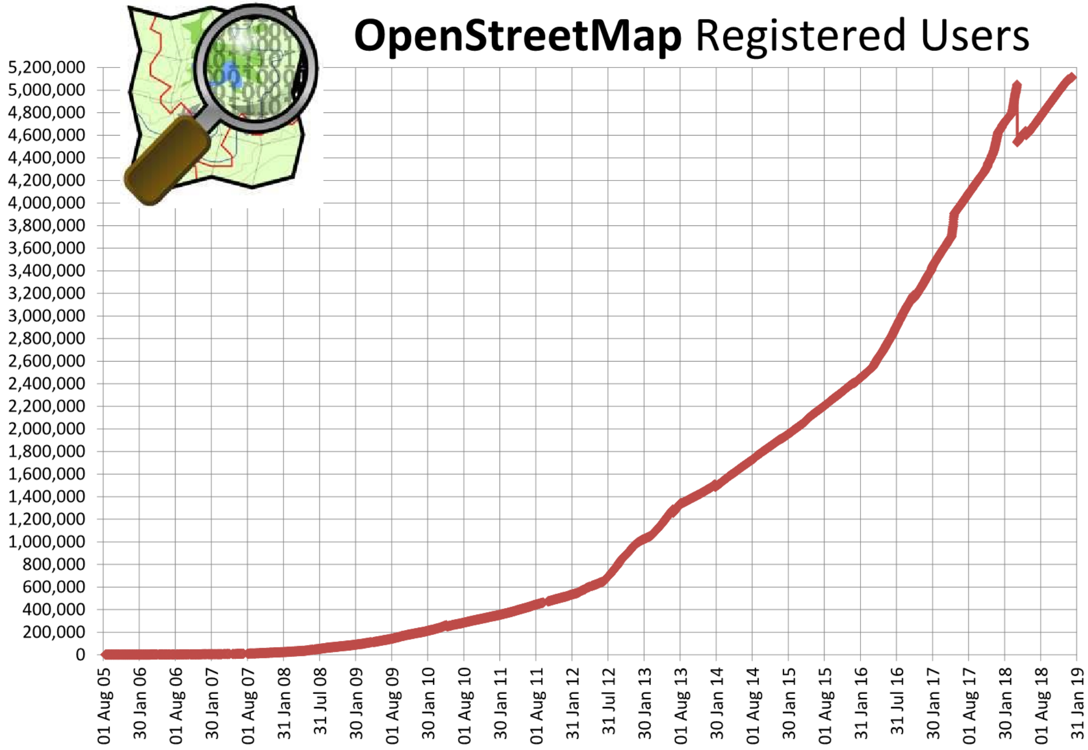
- The most popular VGI project, started in 2004
    - a free, editable vector map of the whole world built by volunteers



http://www.openstreetmap.org

OpenStreetMap (OSM) – How many users?

- Currently (January 2019) there are more than 5 million registered users.



 http://wiki.openstreetmap.org/wiki/Stats

OpenStreetMap License

- OpenStreetMap is open data, licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF).

You are free:

To Share: To copy, distribute and use the database.

To Create: To produce works from the database.

To Adapt: To modify, transform and build upon the database.

As long as you:

Attribute: You must attribute any public use of the database, or works produced from the database, in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, you must make clear to others the license of the database and keep intact any notices on the original database.

Share-Alike: If you publicly use any adapted version of this database, or works produced from an adapted database, you must also offer that adapted database under the ODbL.

Keep open: If you redistribute the database, or an adapted version of it, then you may use technological measures that restrict the work (such as DRM) as long as you also redistribute a version without such measures.

http://opendatacommons.org/licenses/odbl/summary

OpenStreetMap data model – Geometries

- Each object is the result of a combination between a geometry and one or more tags.

- There are three different types of geometry:
  - node - a single point: 
    - point objects: tree, bench, gate, trash bin, etc.
  - way - an ordered list of up to 2000 nodes: 
    - linear objects: road, river, wall, hedge, etc.
    - polygonal objects: building, lake, residential area, etc.
  - relation - an ordered list of nodes, ways and/or other relations: 
    - specific data structures: transportation lines, multipolygon, etc.

https://wiki.openstreetmap.org/wiki/Elements

European Commission

OpenStreetMap data model – Tags

- Each object is the result of a combination between a geometry and one or more tags.

- Tags are the geometry attributes. The object tagging is a very critical process:
  - each tag consists of two elements: a key and a value
  - example: building = hotel

    key value

OpenStreetMap data model – Tags

- Each object is the result of a combination between a geometry and one or more tags.

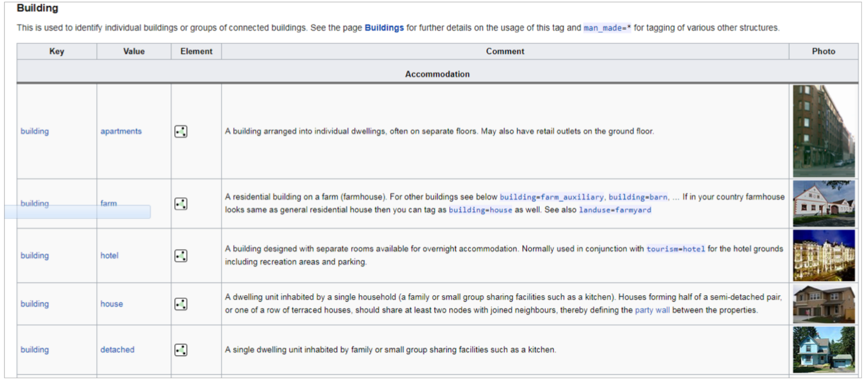
- Tags are the geometry attributes. The objects tagging is a very critical process:
  - the tagging is flexible:
    - each object must have at least one tag
    - there is no limit to the number of tags
    - the reference set of tags are maintained on a wiki page
  - each tag must be verifiable: "…a tag/value combination is verifiable if and only if independent users when observing the same feature would make the same observation every time…"

http://wiki.openstreetmap.org/wiki/Verifiability

OpenStreetMap Data Model – Tags

- Each object is the result of a combination between a geometry and one or more tags.

- Tags are the geometry attributes. The objects tagging is a very critical process:
  - manuals and services are available to guide the tagging process

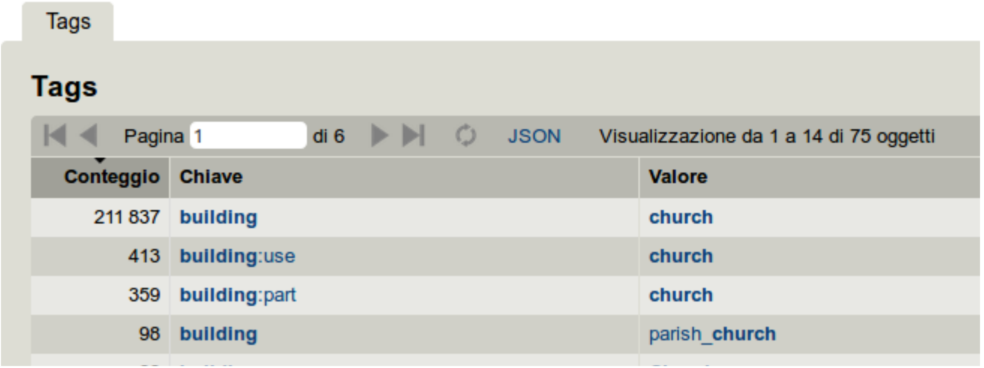


Building

This is used to identify individual buildings or groups of connected buildings. See the page **Buildings** for further details on the usage of this tag and man_made=* for tagging of various other structures.

Key	Value	Element	Comment	Photo
Accommodation				
building	apartments		A building arranged into individual dwellings, often on separate floors. May also have retail outlets on the ground floor.	
building	farm		A residential building on a farm (farmhouse). For other buildings see below building=farm_auxiliary, building=barn, ... If in your country farmhouse looks same as general residential house then you can tag as building=house as well. See also landuse=farmyard	
building	hotel		A building designed with separate rooms available for overnight accommodation. Normally used in conjunction with tourism=hotel for the hotel grounds including recreation areas and parking.	
building	house		A dwelling unit inhabited by a single household (a family or small group sharing facilities such as a kitchen). Houses forming half of a semi-detached pair, or one of a row of terraced houses, should share at least two nodes with joined neighbours, thereby defining the party wall between the properties.	
building	detached		A single dwelling unit inhabited by family or small group sharing facilities such as a kitchen.	

http://wiki.openstreetmap.org/wiki/Map_Features



Tags

Tags

Pagina 1 di 6 JSON Visualizzazione da 1 a 14 di 75 oggetti

Conteggio	Chiave	Valore
211 837	building	church
413	building:use	church
359	building:part	church
98	building	parish_church

https://taginfo.openstreetmap.org

OpenStreetMap Data Model

- Each object is the result of a combination between a geometry and one or more tags.

OpenStreetMap XML Data Format

- OSM objects are stored into XML documents with a .osm extension:
  - static representation (no track of changes)
  - node example

```
<node id="3654216212" visible="true" version="1" changeset="32673542"
      timestamp="2015-07-16T13:09:34Z" user="alexs68" uid="3068291" lat="45.8101242" lon="9.0830260">
 <tag k="addr:city" v="Como"/>
 <tag k="addr:country" v="IT"/>
 <tag k="addr:housenumber" v="34"/>
 <tag k="addr:postcode" v="22100"/>
 <tag k="addr:street" v="Via Bernardino Luini"/>
</node>
```

OpenStreetMap XML Data Format

- OSM objects are stored into XML documents with a .osm extension:
  - static representation (no track of changes)
  - way example

```xml
<way id="30565720" visible="true" version="14" changeset="47145821" timestamp="2017-03-25T10:48:18Z"
     user="mingo23" uid="1902578">
  <nd ref="337748134"/>
  <nd ref="337750203"/>
  <nd ref="337750206"/>
  <nd ref="1259348677"/>
  <nd ref="337750207"/>
  <nd ref="337750211"/>
  <nd ref="337750215"/>
  <nd ref="1259347237"/>
  <nd ref="337750218"/>
  <nd ref="337750221"/>
  <nd ref="337750224"/>
  <nd ref="1259347971"/>
  <nd ref="337750228"/>
  <nd ref="337750232"/>
  <nd ref="337748139"/>
  <nd ref="337748134"/>
  <tag k="amenity" v="place_of_worship"/>
  <tag k="building" v="church"/>
  <tag k="denomination" v="catholic"/>
  <tag k="name" v="Basilica di San Fedele"/>
  <tag k="religion" v="christian"/>
  <tag k="wikidata" v="Q2887102"/>
  <tag k="wikipedia" v="it:Basilica di San Fedele (Como)"/>
</way>
```

OpenStreetMap governance

- The OSM project is supported by the OpenStreetMap Foundation (OSMF):
  - communicates OSM's work to the world
  - acts as a legal entity for the OSM project
  - is the custodian for the OSM technical infrastructure
  - ensures the financial sustainability of the project
  - organizes the annual "State of the Map" conference
  - maintains the license of the data and ensure that users comply with it
  - allocates funds to software projects through microgrants
  - does NOT manage software projects
  - does NOT decide what to map and not to map

https://wiki.osmfoundation.org/wiki/Mission_Statement

OpenStreetMap governance

- The OpenStreetMap Foundation is composed of:
  - Normal/Associate Members: entitled to vote in the affairs of the OSMF
  - Corporate Members: for companies & organisations
  - Board of Directors: support and further the interests of the OSMF
  - Advisory Board: a group the Board may consult on important decisions
  - Local Chapters: national/regional organizations supporting OSMF mission
  - Working Groups: provide support to OSM in specific areas
    - Licensing Working Group
    - Data Working Group
    - Operations Working Group
    - Engineering Working Group
    - Communication Working Group
    - State of the Map Working Group
    - Membership Working Group

https://wiki.osmfoundation.org/wiki/Main_Page

Who is using OpenStreetMap?

- Applications using the OSM database include:
  - data editors
  - data download applications and services
  - disaster and humanitarian applications
  - government and industry usage
  - visualization services
  - routing services
  - quality assurance for OSM
  - games and leisure
  - education and research

Reference

CHAPTER 3

A Review of OpenStreetMap Data

Peter Mooney* and Marco Minghini†

*Department of Computer Science, Maynooth University, Maynooth, Co. Kildare, Ireland, Peter.mooney@nuim.ie

†Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Research works on VGI and OSM

- These topics/applications will be presented in the following:
  - 1. OSM extrinsic data quality assessment
    - a. road networks
    - b. buildings
  - 2. OSM intrinsic data quality assessment
  - 3. analysis of OSM contributions patterns
  - 4. production of Land Use/Land Cover (LULC) maps from OSM

European Commission

1a

OSM extrinsic data quality assessment – road networks

VGI & OSM quality

- Increasing availability of open data from National Mapping Agencies and Commercial Mapping Companies usable as a source of comparison for VGI (and OSM) data, i.e. for extrinsic quality assessment.
- Literature provides plenty of works assessing or comparing OSM quality against that of authoritative datasets:
    - strongly focused on road network
    - OSM compared to data from NMAs (UK Ordnance Survey, French NMA, USGS TNM/TIGER, etc.) and CSCs (Navteq, TeleAtlas, etc.)
    - semi- or fully-automated
- Comparison techniques are very strong and fit for purpose, but mostly application and dataset specific:
    - hard to replicate
    - difficult to extend to other dataset comparisons

European Commission

Our methodology

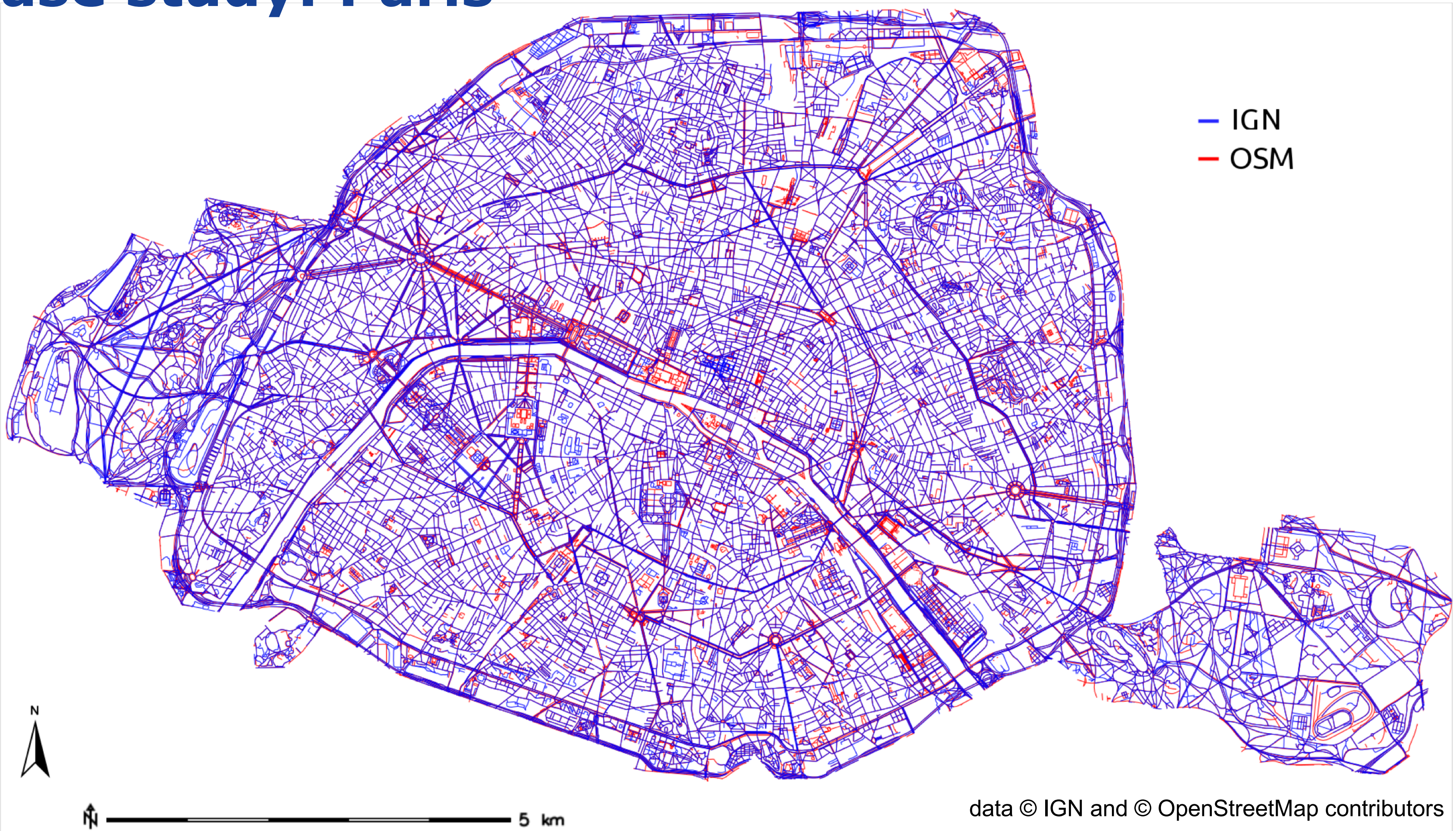
- Novel methodology to compare OSM and reference road datasets:
  - fully automated
  - focused on positional accuracy and completeness
  - flexible, i.e. not developed for a specific dataset
  - built with FOSS4G (Free and Open Source Software for Geospatial)
    - reusable and extensible in case of need

European Commission

Our methodology – Overview

- Currently developed as 3 GRASS GIS modules:
  - written in Python
  - available with a Graphical User Interface (GUI)
- Comparison between OSM and reference road network datasets composed of 3 consecutive steps:
  - 1. Preliminary comparison of the datasets and computation of global statistics
  - 2. Geometric preprocessing of the OSM dataset to extract a subset which is fully comparable with the reference dataset
  - 3. Evaluation of OSM quality using a grid-based approach

Case study: Paris



data © IGN and © OpenStreetMap contributors

Step 1: Preliminary comparison of the datasets

- Compute the total length of OSM and IGN datasets and their length difference, both in map units and percentage [*required*]
  - output values are returned in a text file

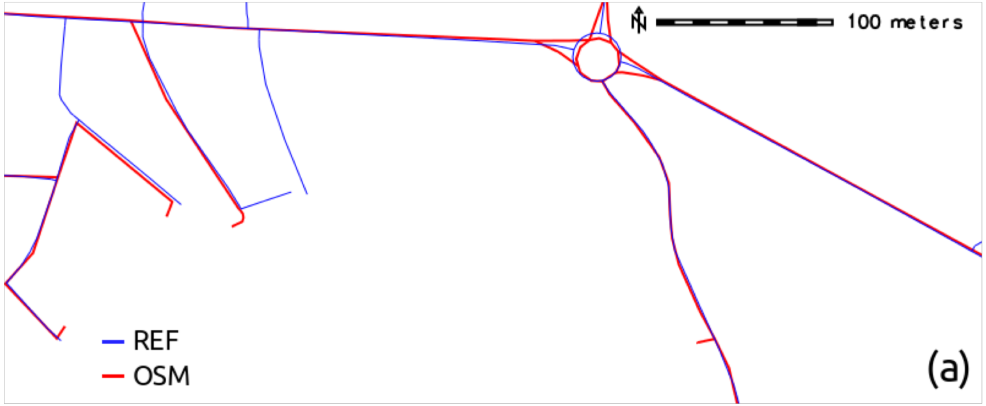
```
REF length: 2686373.1 m
OSM length: 3124627.0 m
REF-OSM difference: -438253.9 m (-16.3%)
```

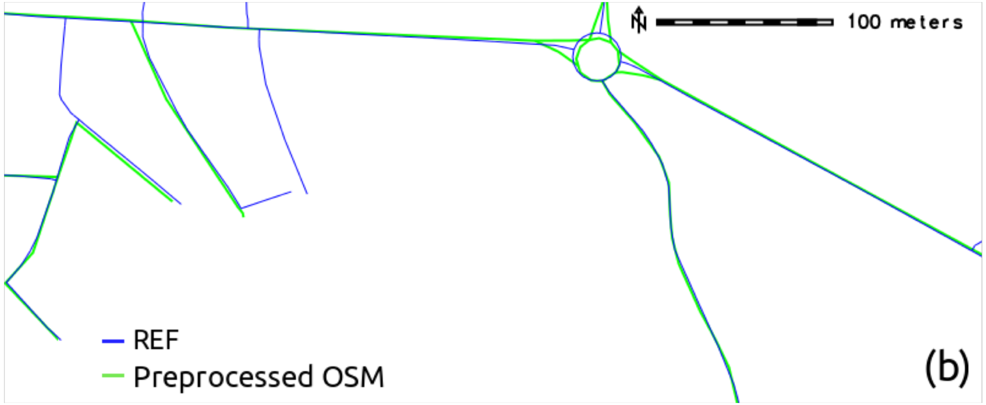
- ≅450 km more in OSM than IGN dataset!

```
BUFFER(m)|OSM_IN(m)|OSM_IN(%%)|OSM_OUT(m)|OSM_OUT(%%)|REF_IN(m)|REF_IN(%%)|REF_OUT(m)|REF_OUT(%%)
1.0|1374755.9|44.0|1749871.2|56.0|1366471.0|50.9|1319902.1|49.1
2.0|2014259.9|64.5|1110367.2|35.5|1982713.7|73.8|703659.4|26.2
3.0|2298072.4|73.5|826554.6|26.5|2223153.5|82.8|463219.6|17.2
4.0|2464185.0|78.9|660442.0|21.1|2329270.3|86.7|357102.8|13.3
5.0|2582784.2|82.7|541842.9|17.3|2387687.7|88.9|298685.4|11.1
6.0|2671758.8|85.5|452868.2|14.5|2424463.5|90.3|261909.6|9.7
7.0|2738327.0|87.6|386300.0|12.4|2451476.9|91.3|234896.2|8.7
8.0|2792053.8|89.4|332573.2|10.6|2471557.1|92.0|214816.0|8.0
9.0|2828903.0|90.5|295724.1|9.5|2488514.1|92.6|197859.0|7.4
10.0|2859512.1|91.5|265114.9|8.5|2501974.7|93.1|184398.4|6.9
11.0|2886190.1|92.4|238436.9|7.6|2513592.9|93.6|172780.2|6.4
12.0|2908071.9|93.1|216555.1|6.9|2523138.5|93.9|163234.6|6.1
13.0|2925602.0|93.6|199025.1|6.4|2532070.5|94.3|154302.6|5.7
14.0|2941922.8|94.2|182704.2|5.8|2540322.9|94.6|146050.2|5.4
15.0|2956112.7|94.6|168514.3|5.4|2548274.0|94.9|138099.1|5.1
16.0|2967813.5|95.0|156813.5|5.0|2555431.5|95.1|130941.6|4.9
17.0|2977318.7|95.3|147308.3|4.7|2562238.1|95.4|124135.0|4.6
18.0|2986371.8|95.6|138255.2|4.4|2568276.5|95.6|118096.6|4.4
19.0|2994833.4|95.8|129793.7|4.2|2574052.2|95.8|112320.9|4.2
20.0|3001796.0|96.1|122831.1|3.9|2579434.1|96.0|106939.0|4.0
```

Step 2: Preprocessing of the OSM dataset

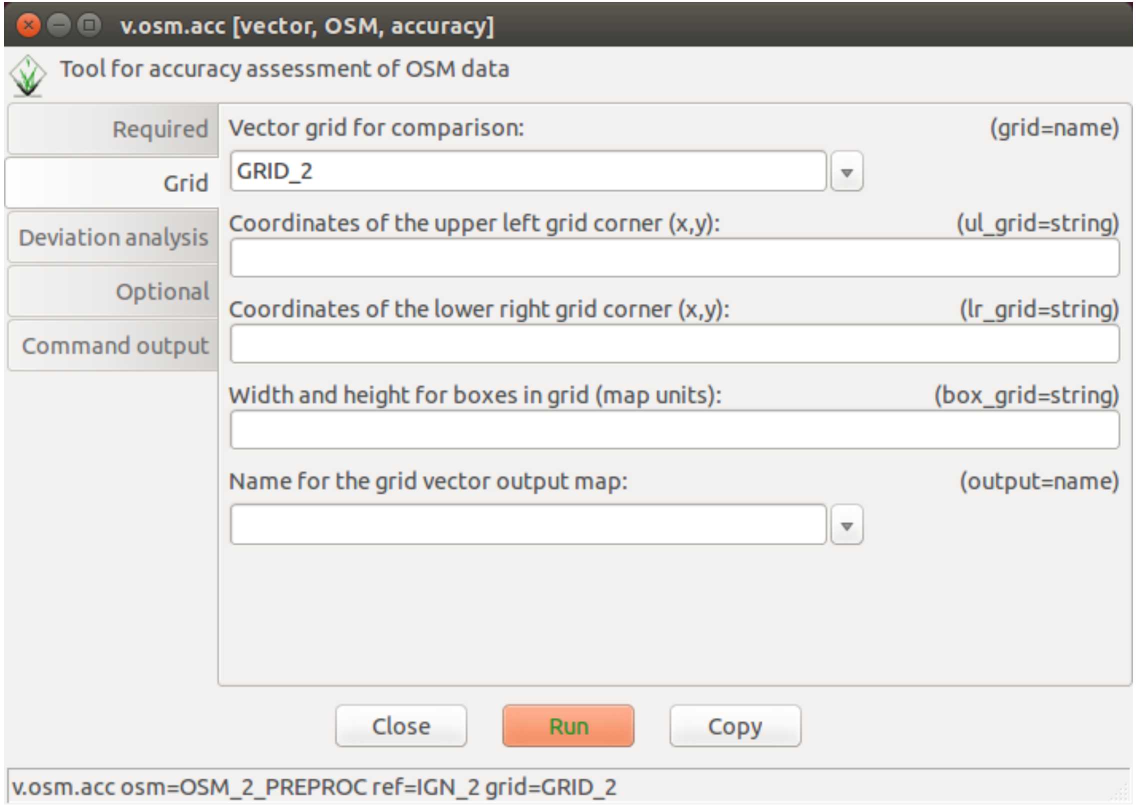
- Cleaning of OSM dataset to make it comparable with IGN dataset





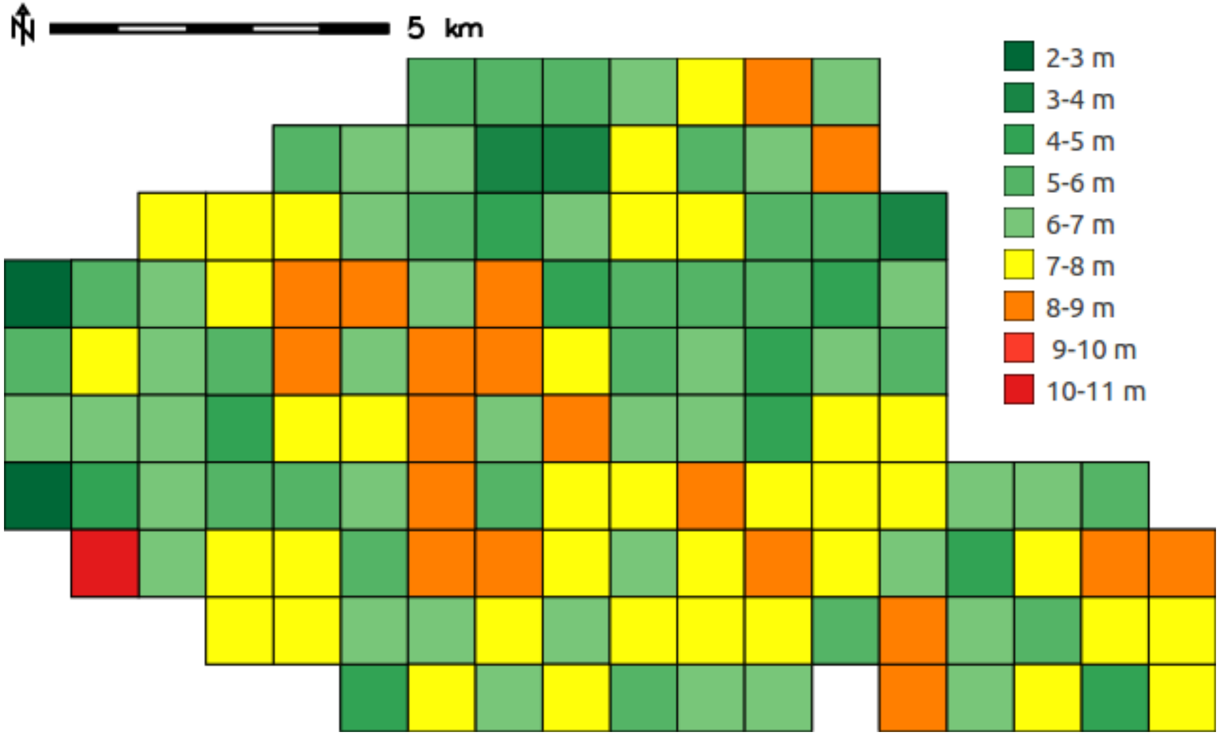
Step 3: Grid-based evaluation of OSM accuracy

- Use a grid to take into account OSM heterogeneous nature



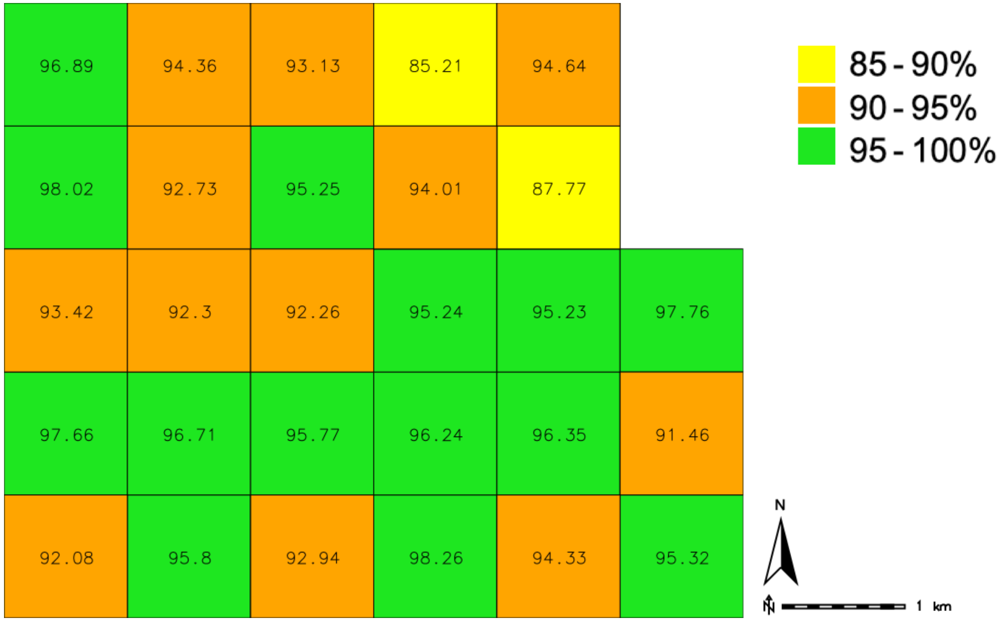
Step 3: Grid-based evaluation of OSM accuracy

- For each grid cell, find the OSM maximum deviation from IGN:
  - generalization threshold = 0.5 m, buffer = 11 m, OSM length % = 95



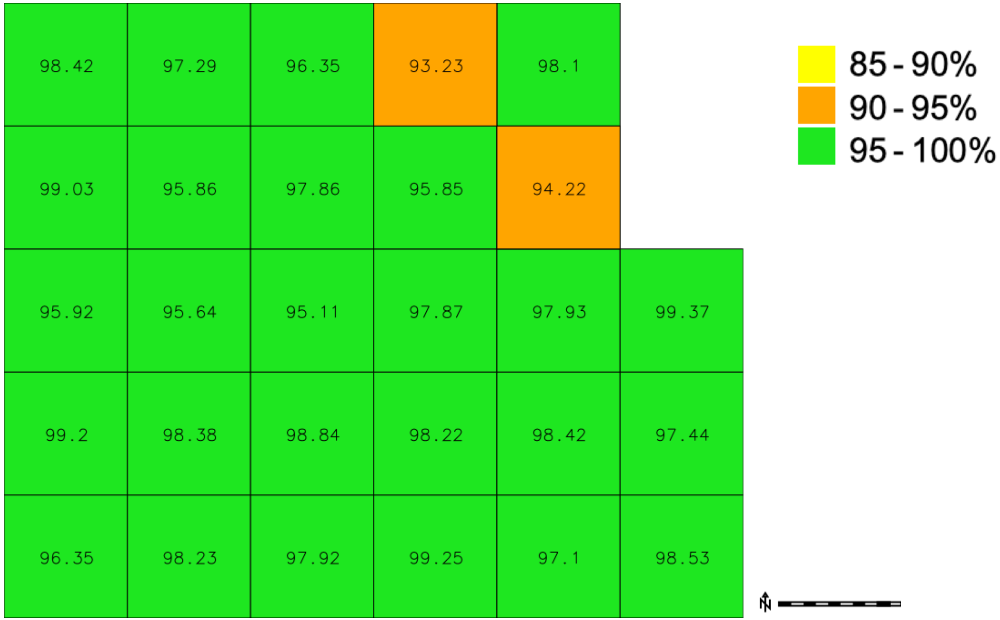
Step 3: Grid-based evaluation of OSM accuracy

- For each grid cell, evaluate OSM accuracy against one or more target values of OSM deviation from IGN:
    - length percentage of OSM roads included in the target buffer
    - Area 2: target buffer = 6 m



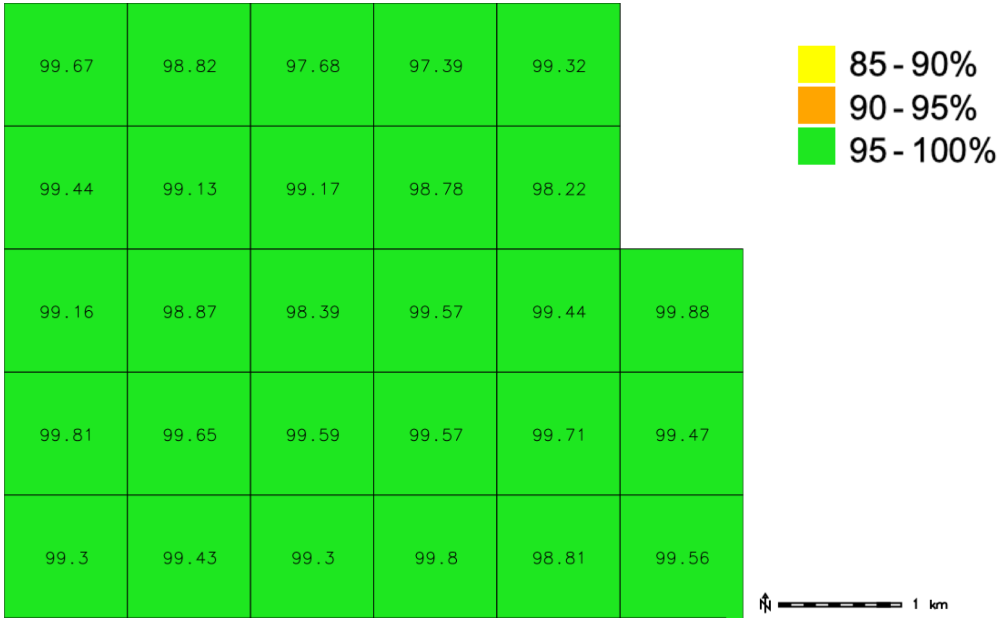
Step 3: Grid-based evaluation of OSM accuracy

- For each grid cell, evaluate OSM accuracy against one or more target values of OSM deviation from IGN:
  - length percentage of OSM roads included in the target buffer
  - Area 2: target buffer = 8 m



Step 3: Grid-based evaluation of OSM accuracy

- For each grid cell, evaluate OSM accuracy against one or more target values of OSM deviation from IGN:
  - length percentage of OSM roads included in the target buffer
  - Area 2: target buffer = 10 m



References

Transactions in GIS

Research Article

Transactions in GIS, 2016, 00(00): 00–00

Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets

Maria Antonia Brovelli,* Marco Minghini,* Monia Molinari* and Peter Mooney†

*Department of Civil and Environmental En

†Department of Computer Science, Maynoo

ORDEM DOS ENGENHEIROS

INFORMAÇÃO GEOESPACIAL PARA AS GERAÇÕES FUTURAS

OPORTUNIDADES E DESAFIOS

VIII CNCG

CONFERÊNCIA NACIONAL DE CARTOGRAFIA E GEODESIA

29 e 30 de outubro de 2015

Assessing OSM Road Positional Quality With Authoritative Data

Francisco ANTUNES[1], Cidália C. FONTE[1], Maria Antonia BROVELLI[2], Marco MINGHINI[2], Monia MOLINARI[2] and Peter MOONEY[3]

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B7, 2016
XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic

AN AUTOMATED GRASS-BASED PROCEDURE TO ASSESS THE GEOMETRICAL ACCURACY OF THE OPENSTREETMAP PARIS ROAD NETWORK

M. A. Brovelli, M. Minghini *, M. E. Molinari

Department of Civil and Environmental Engineering, Politecnico di Milano, Como Campus, Via Valleggio 11, 22100 Como, Italy – (maria.brovelli, marco.minghini, moniaelisa.molinari)@polimi.it

1b

OSM extrinsic data quality assessment – buildings

Extrinsic quality assessment of OSM buildings

- The purpose of this study is to contribute to the quality assessment of OSM buildings in Milan (Italy).

- The quality assessment has been performed by comparing the OSM data (downloaded in January 2016) against the building layer of the official vector cartography of Milan Municipality (produced in 2012).

- Two different quality parameters were evaluated:
  - completeness evaluation based on methods suggested by literature
  - positional accuracy evaluation based on a novel, quasi-automated matching algorithm

European Commission

Completeness assessment

- The completeness analysis was performed through the area ratio unit-based method proposed by Hecht et al. (2013):

$$C = A_{OSM} / A_{REF}$$

C = completeness

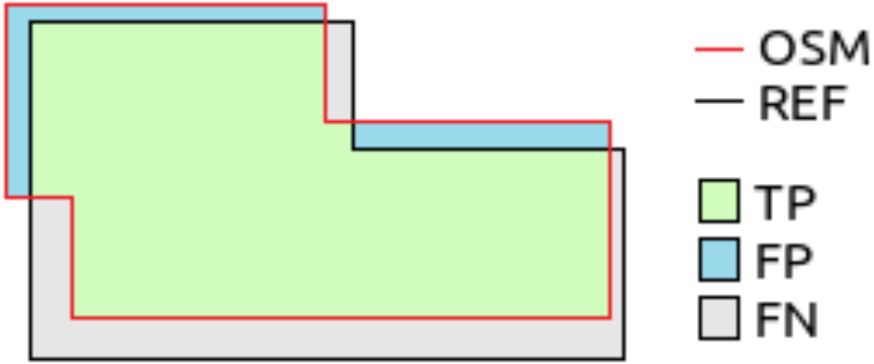
A_{REF} = total area of reference buildings

A_{OSM} = total area of OSM buildings

- The completeness parameter should be calculated within a predefined (e.g. administrative or geometrical) spatial unit to take into account the heterogeneity of OSM data.

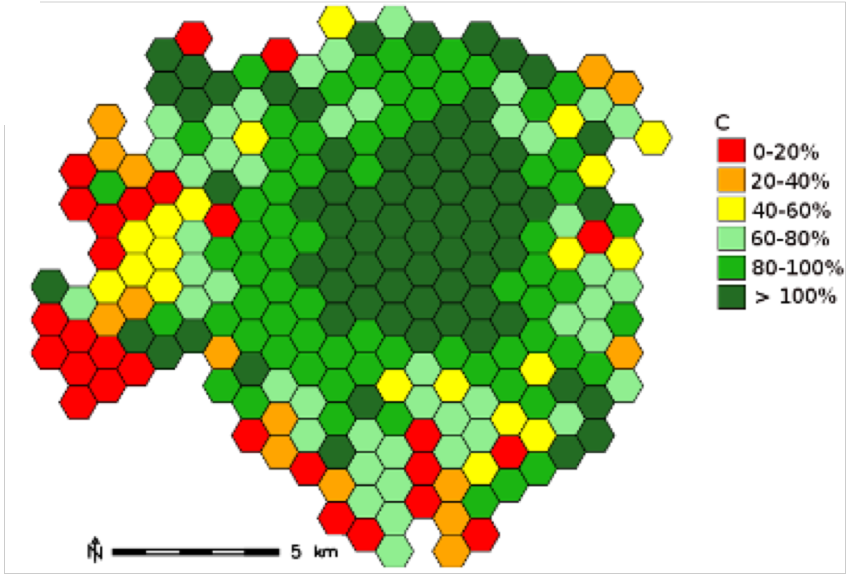
Completeness assessment

- The area ratio method can introduce an overestimation of C due to exceeding data available in OSM. For this reason, the computation of 3 additional rates is recommended:
    - True Positive (TP): the areas of agreement between the datasets
    - False Positive (FP): the OSM building areas which do not exist in the REF dataset
    - False Negative (FN): the REF building areas which do not exist in the OSM dataset



Completeness results

- Spatial distribution of completeness rate in Milan area:



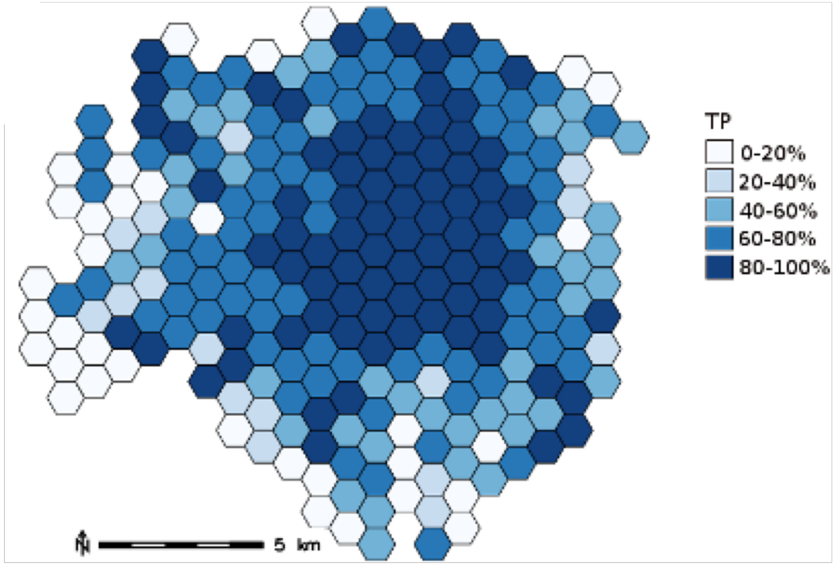
C VALUES	%
> 100%	28.9%
80% < C < 100%	27.7%
60% < C < 80%	18.5%
40% < C < 60%	8%
C < 40%	16.9%

- The completeness of the OSM dataset is very high in the city centre and gradually decreases when moving towards the periphery.

European Commission

TP results

- Spatial distribution of TP rate in Milan area:

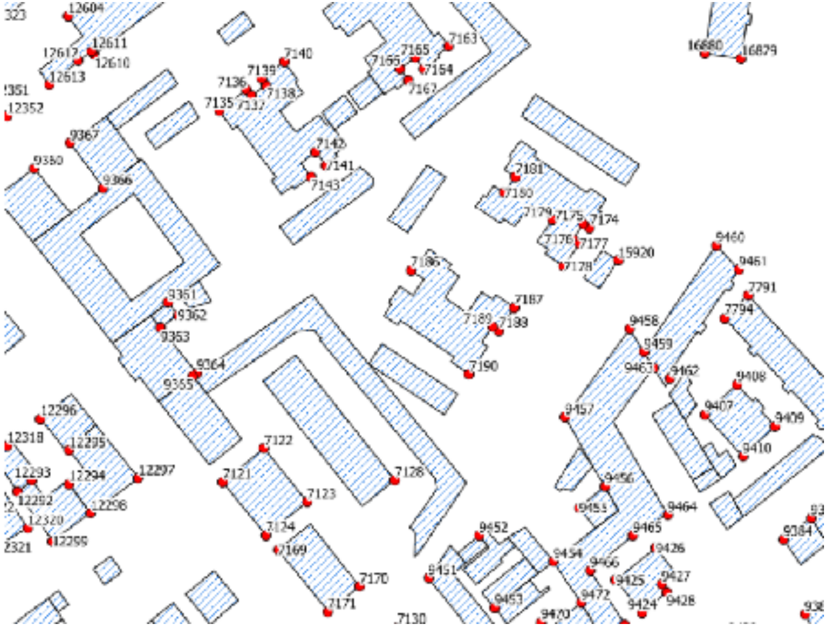


TP VALUES	%
100% > TP > 60%	63.9%
60% < TP < 40%	16.0%
TP < 40%	20.1%

- Results largely confirm the trend observed for C: OSM completeness is high in the city center and gradually lower in the peripheral areas.

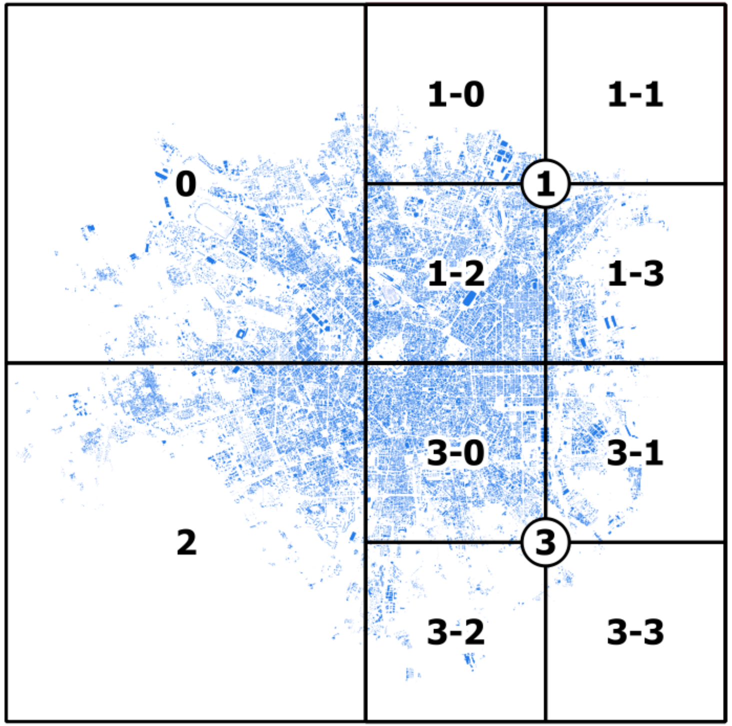
Positional accuracy assessment

- An advanced algorithm allowed the quasi-automated detection of homologous points between REF and OSM by means of geometric, topological and semantic analyses:



Positional accuracy results

- The number of homologous pairs detected is approximately 100’000.



Cell	Points	Trasf.	ΔY μ [m]	ΔX μ [m]	d μ [m]
0	19135	None	0.45	0.46	0.81
1-2	16480	None	0.35	0.46	0.77
2	18732	None	0.44	0.43	0.79
3-1	4318	None	0.28	0.41	0.71

- The positional accuracy:
  - is very high
  - is the same in both Milan city center and periphery.

Reference

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B2, 2016
XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic

POSITIONAL ACCURACY ASSESSMENT OF THE OPENSTREETMAP BUILDINGS LAYER THROUGH AUTOMATIC HOMOLOGOUS PAIRS DETECTION: THE METHOD AND A CASE STUDY

M. A. Brovelli *, M. Minghini, M. E. Molinari, G. Zamboni

Department of Civil and Environmental Engineering, Politecnico di Milano, Como Campus, Via Valleggio 11, 22100 Como, Italy - (maria.brovelli, marco.minghini, moniaelisa.molinari, giorgio.zamboni)@polimi.it

European Commission

2

OSM intrinsic data quality assessment

OSM intrinsic quality assessment

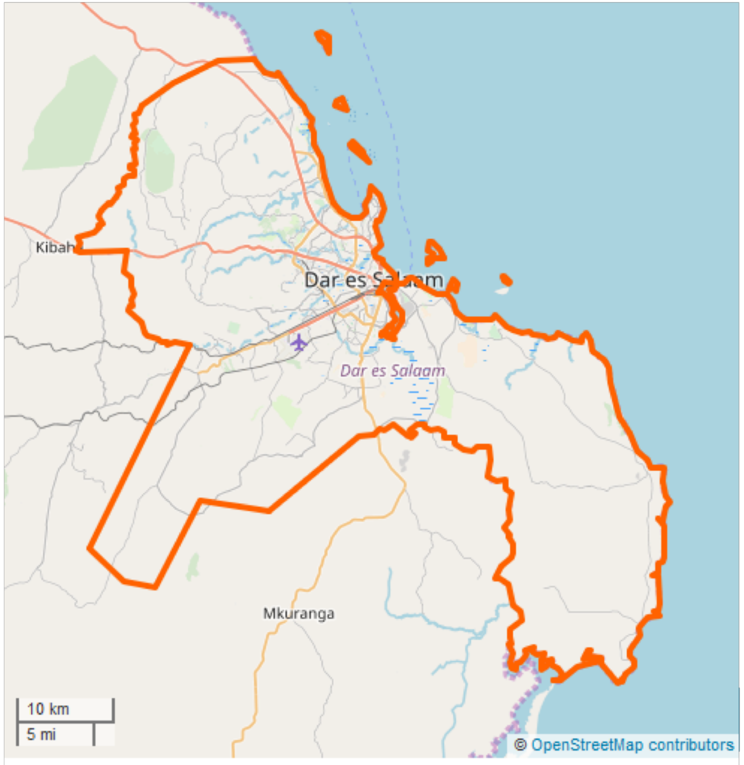
- It often makes little sense to limit OSM quality assessment to comparison against reference datasets:
  - OSM (VGI) and authoritative datasets are too different
  - reference datasets may be not available, or not accurate/up-to-date

- In an intrinsic quality assessment, OSM quality is evaluated by only looking at OSM itself, i.e. at the history of data:
  - OSM history is available through the OSM API and the Planet file
  - another quality parameter can be assessed: usability or fitness-for-use

European Commission

Methodology – Study area & hypotheses

- Study area: Dar es Salaam, Tanzania
  - densely mapped in OSM, mainly thanks to the Dar Ramani Huria project
- Analysis on nodes (POIs) and ways:
  - nodes/ways deleted not considered
  - edits considered are only those with changes in tags
  - changes made in a single changeset count as one single new version
- Data was downloaded on May 3, 2018:
  - 129572 nodes and 1156948 ways
  - edits by 1959 different contributors
  - 150716 and 1592221 versions for nodes and ways, respectively



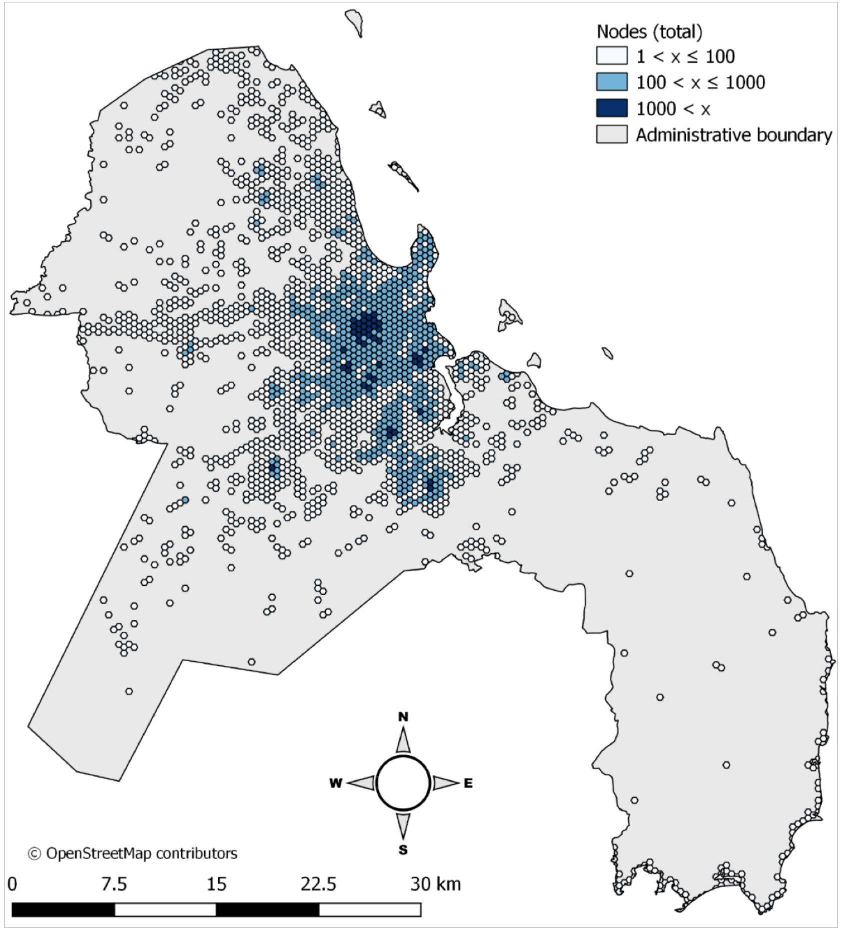
Methodology – Architecture

- More extensive analysis on a predefined area:
  - aggregate and store results in a database
  - suitable for further GIS processing

Frontend	QGIS
Interface	*SQLite file*
Processing	SQLite + Spatialite
Conversion	osh2sql / libosmium
Source	Planet

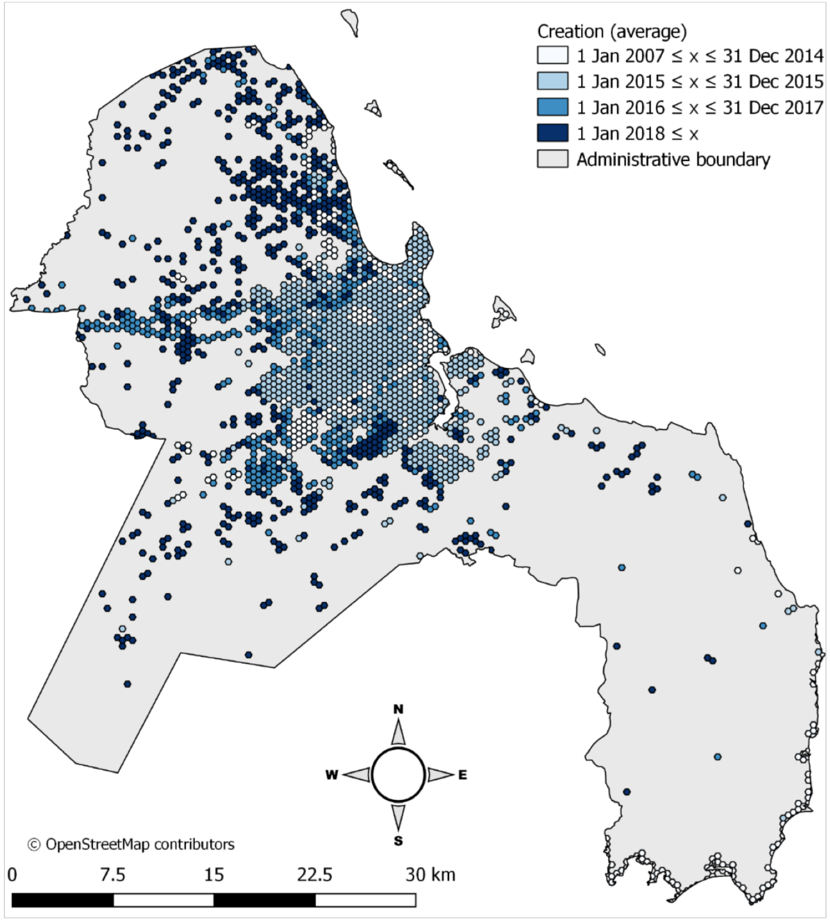
Results

- Total number of OSM nodes:
  - 79% of the total area does not contain any node
  - density of nodes progressively increasing from the rural to the most urbanized areas



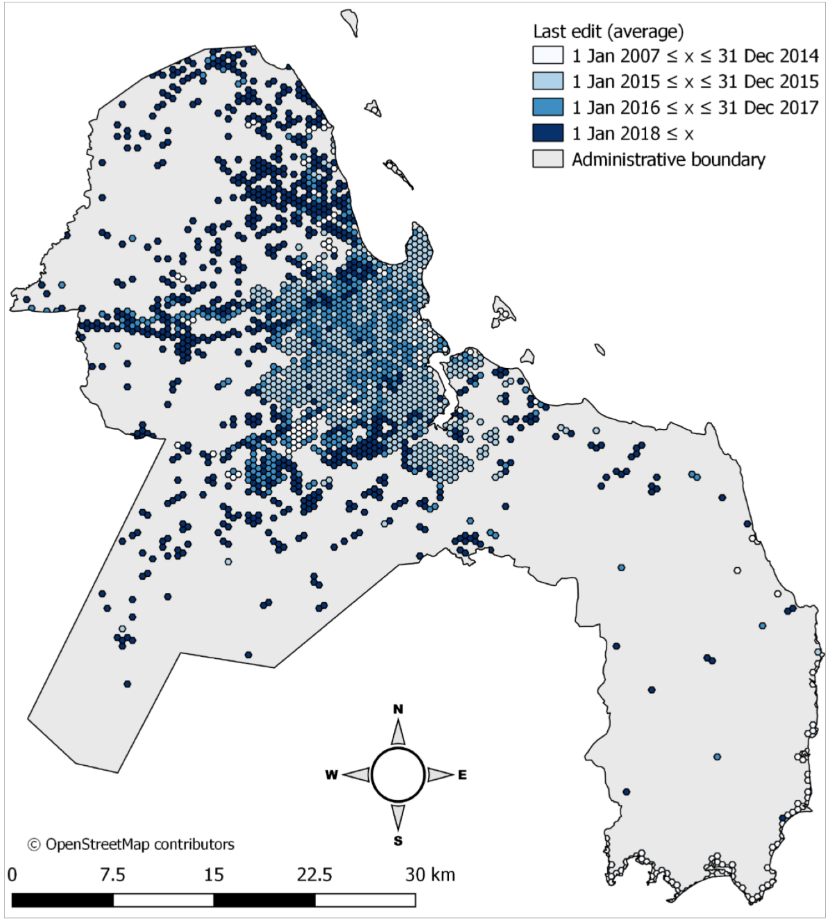
Results

- Average date of creation of OSM nodes:
  - most of the nodes in the city center created in 2015
  - attention gradually moved to the peripheral areas in 2016, 2017 and 2018



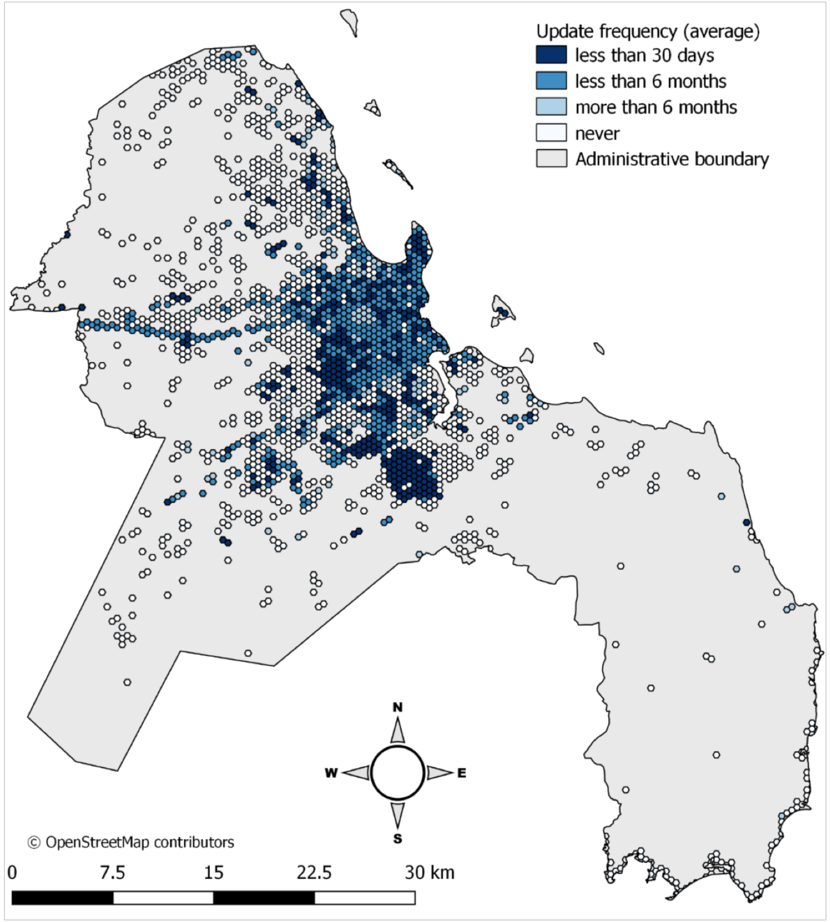
Results

- Average date of last edit of OSM nodes:
  - few of the nodes created in 2014-2015 were later updated
  - mapping in 2018 focused on peripheral areas



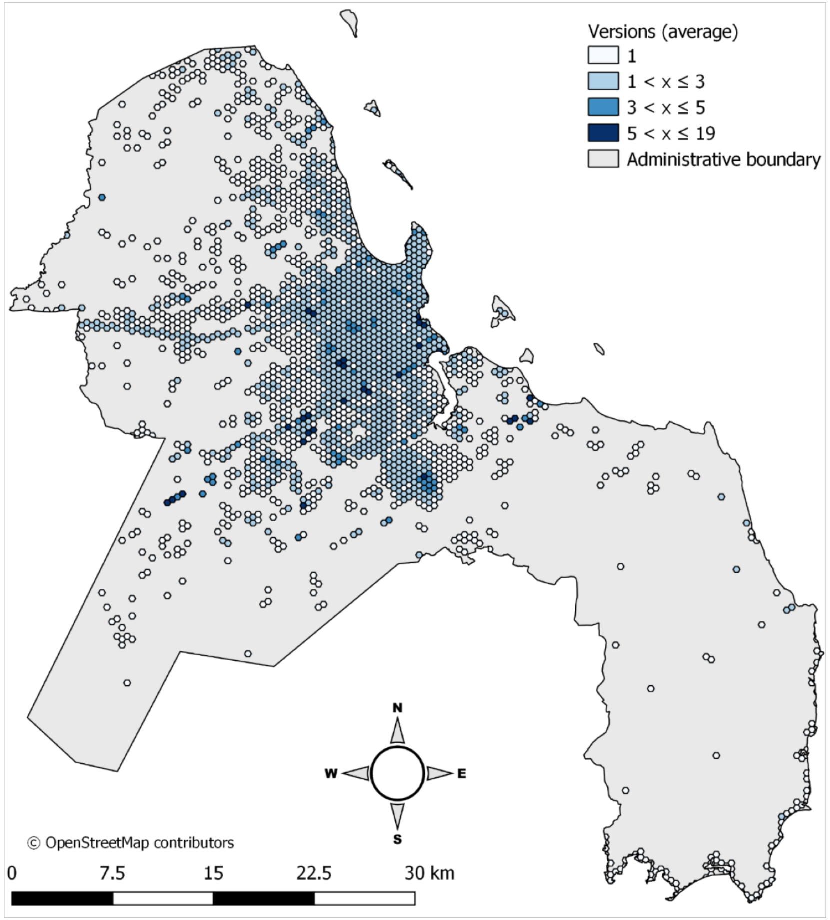
Results

- Average update frequency of OSM nodes:
    - highest update frequencies in the city center
    - most of the nodes created in 2018 have not yet been updated



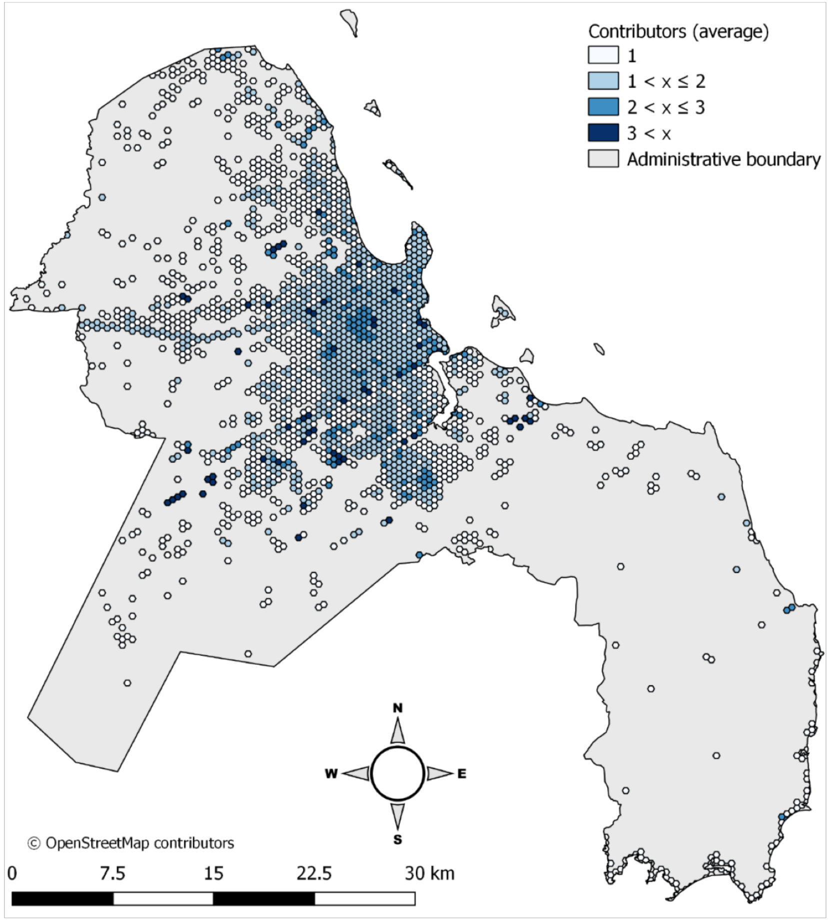
Results

- Average number of versions of OSM nodes:
  - most of recently created nodes not (yet) updated
  - increase in the number of version when moving to the city center



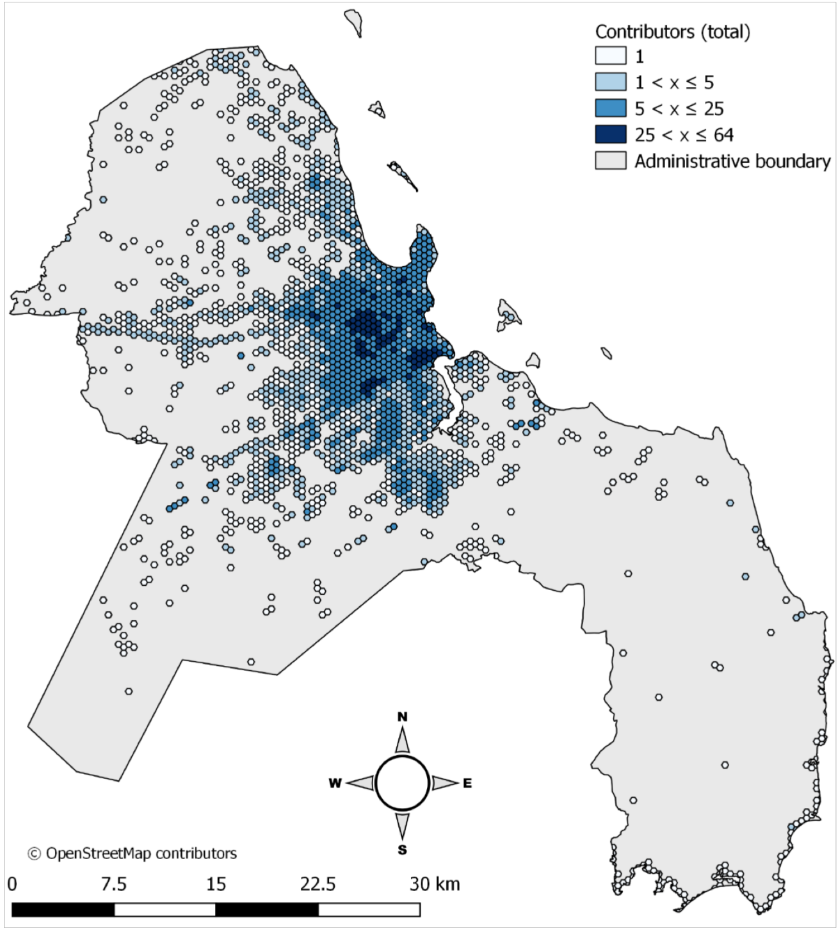
Results

- Average number of different contributors on OSM nodes:
    - equal to 1 for 53% of the cells, mainly in the outskirts
    - increases towards the city center



Results

- Total number of different contributors on OSM nodes:
  - generally equal to 1 in the periphery, increases towards the city center



Reference

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-4/W8, 2018
FOSS4G 2018 – Academic Track, 29–31 August 2018, Dar es Salaam, Tanzania

AN OPEN SOURCE APPROACH FOR THE INTRINSIC ASSESSMENT OF THE TEMPORAL ACCURACY, UP-TO-DATENESS AND LINEAGE OF OPENSTREETMAP

M. Minghini *, M. A. Brovelli, F. Frassinelli

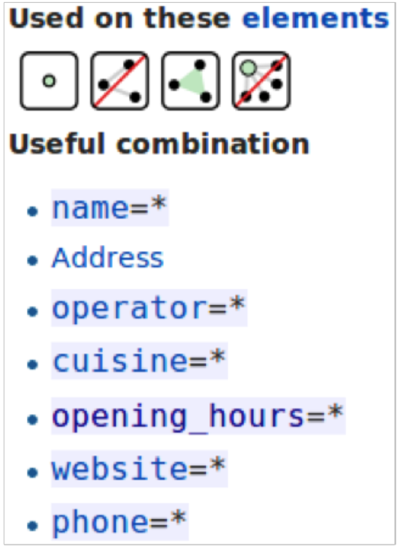
European Commission

3

Analysis of OSM contribution patterns

Tagging in OSM

- OSM applies a folksonomy approach to tagging with no formal rules forced:
    - tagging rule-book is the OSM Map Features wiki page
        - guidance on which tags and combinations of tags to use



Tagging in OSM

- OSM applies a folksonomy approach to tagging with no formal rules forced:
  - tagging rule-book is the OSM Map Features wiki page
    - guidance on which tags and combinations of tags to use
  - taginfo shows that this guidance may not be universally adopted!

Used on these elements

Useful combination

- name=*
- Address
- operator=*
- cuisine=*
- opening_hours=*
- website=*
- phone=*

amenity=restaurant

Overview | Combinations | Map | Wiki | Projects

Combinations

This table shows only the most common combinations of the most common tags.

Page 1 of 13 JSON Displaying 1 to 11 of 135 items

Count →		Other tags
687 829	91.36%	name=*
329 005	43.70%	cuisine=*
246 939	32.80%	addr:street=*
204 893	27.22%	addr:housenumber=*
180 841	24.02%	addr:city=*
168 643	22.40%	addr:postcode=*
140 042	18.60%	building=*
127 409	16.92%	building=yes
113 375	15.06%	phone=*
111 769	14.85%	website=*
93 607	12.43%	source=*

Analysis of OSM tagging practices

- Research questions:
  - do OSM contributors comply to the suggested combinations of tags?
  - does this compliance vary spatially?
- Selection of 10 among the most frequently occurring tags in OSM

Target Tag	TagInfo Ranking	Number of Objects
highway=residential	2	34,688,039
natural=tree	17	7,019,552
highway=footway	18	6,126,861
highway=path	24	4,506,593
highway=tertiary	25	4,328,513
amenity=parking	52	2,061,012
highway=primary	59	1,869,021
highway=bus_stop	66	1,677,724
railway=rail	69	1,584,142
leisure=pitch	93	977,983

Analysis of OSM tagging practices

- Research questions:
  - do OSM contributors comply to the suggested combinations of tags?
  - does this compliance vary spatially?
- Selection of 10 among the most frequently occurring tags in OSM
- Selection of 40 world cities

Methodology

- For each city, for each target tag and for each of the suggested tags to be used in combination:
  - computation of the fraction of objects containing both the target tag and the suggested tag
  - mapping of the fraction to a 5 part Likert Scale
    - 0-20% – POOR
    - 20-40% – FAIR
    - 40-60% – AVERAGE
    - 60-80% – GOOD
    - 80-100% – EXCELLENT
- Example: Christchurch (New Zealand), tag *leisure=pitch*

Report for Tag: `leisure=pitch`			
Total number of objects: 470			
sport	364	77.5%	GOOD
surface	42	9.0%	POOR
Total number of different tags used: 26			

Results

- *highway=residential*

KEY	Poor	Fair	Average	Good	Excellent
name	8	1	7	5	19
oneway	34	5	1	0	0

- *natural=tree*

KEY	Poor	Fair	Average	Good	Excellent
circumference	38	0	1	0	1
taxon	38	0	0	0	2
leaf_type	34	2	2	1	1
start_date	39	0	0	1	0
height	37	0	1	0	2
denotation	36	1	2	0	1
genus	38	1	1	0	0
species	35	1	2	0	2

- *highway=primary*

KEY	Poor	Fair	Average	Good	Excellent
lanes	10	10	6	6	8
ref	8	10	6	2	14
name	0	2	4	10	24

Results

- *highway=bus_stop*

KEY	Poor	Fair	Average	Good	Excellent
operator	28	4	2	3	3
public_transport	21	7	5	3	4
name	3	4	3	9	21

- *leisure=pitch*

KEY	Poor	Fair	Average	Good	Excellent
sport	0	2	7	16	15
surface	40	0	0	0	0

- Summary:

Tag	Keys	Poor	Fair	Average	Good	Excellent
highway=primary	3	15.00	18.33	13.33	15.00	38.33
highway=tertiary	4	40.00	20.00	13.75	14.38	11.88
highway=bus-stop	3	43.33	12.50	8.33	12.50	23.33
railway=rail	9	46.39	18.61	12.78	11.67	10.56
leisure=pitch	2	50.00	2.50	8.75	20.00	18.75
highway=residential	2	52.50	7.50	10.00	6.25	23.75
amenity=parking	6	90.83	6.67	2.50	0.00	0.00
highway=path	7	91.78	5.71	2.50	0.00	0.00
natural=tree	8	92.19	1.56	2.81	0.62	2.81
highway=footway	6	94.58	4.58	0.83	0.00	0.00

4

Production of Land Use/Land Cover (LULC) maps from OSM

European Commission

Land Use/Land Cover (LULC) maps

- LULC maps are crucial products for multiple areas of application:
  - modeling climate and biochemistry of the Earth
  - natural resources management
  - planning/urban studies
  - many others
- LULC maps are created through the classification of satellite imagery and validated using reference data:
  - the creation and updating process is costly and time-consuming - insufficient to describe rapidly-changing environments
  - level of detail and spatial coverage inadequate for many applications

OSM as a source of LULC maps

- Exploiting OSM as a source for LULC maps has a number of advantages:
    - OSM full spatial coverage in the world
    - OSM richness
    - OSM non-stop updating
    - OSM open license
- Exploiting OSM as a source for LULC maps has some disadvantages:
    - OSM uneven spatial coverage
    - OSM positional accuracy & geometrical inconsistencies
    - OSM semantic inconsistencies
- Purpose: creating an automated procedure which converts OSM data in a specific area into a LULC map
    - reference nomenclatures of current EU LULC maps

Urban Atlas (UA)

- Former project of the European Environment Agency (EEA), now moved under the Copernicus Land Monitoring Service
- Aims to provide high resolution LULC maps for Pan-European regions
  - with more than 50000 inhabitants
- Freely available in vector format
- Scale: 1:10000
- Minimum Mapping Units (MMU):
  - 100m for linear features
  - 0,25ha (0.0025 km^2) for urban area features, 1 ha for rural area features
- Positional accuracy: ±5m, overall thematic accuracy higher than 80-85%
- The classification of UA uses a nomenclature separated into levels

Corine Land Cover (CLC)

- Project of the Copernicus Land Monitoring Service
- Provides a consistent, comparable, pan-European land cover product
- Freely available in vector & raster formats
- Scale: 1:100000
- Minimum Mapping Units (MMU):
    - 100m for linear features
    - 0,25km^2 for area features
- The positional accuracy is 100m and the overall thematic accuracy is greater than 85%
- The CLC classification is made considering a nomenclature separated into levels (there are 44 land cover classes in the most detailed level)

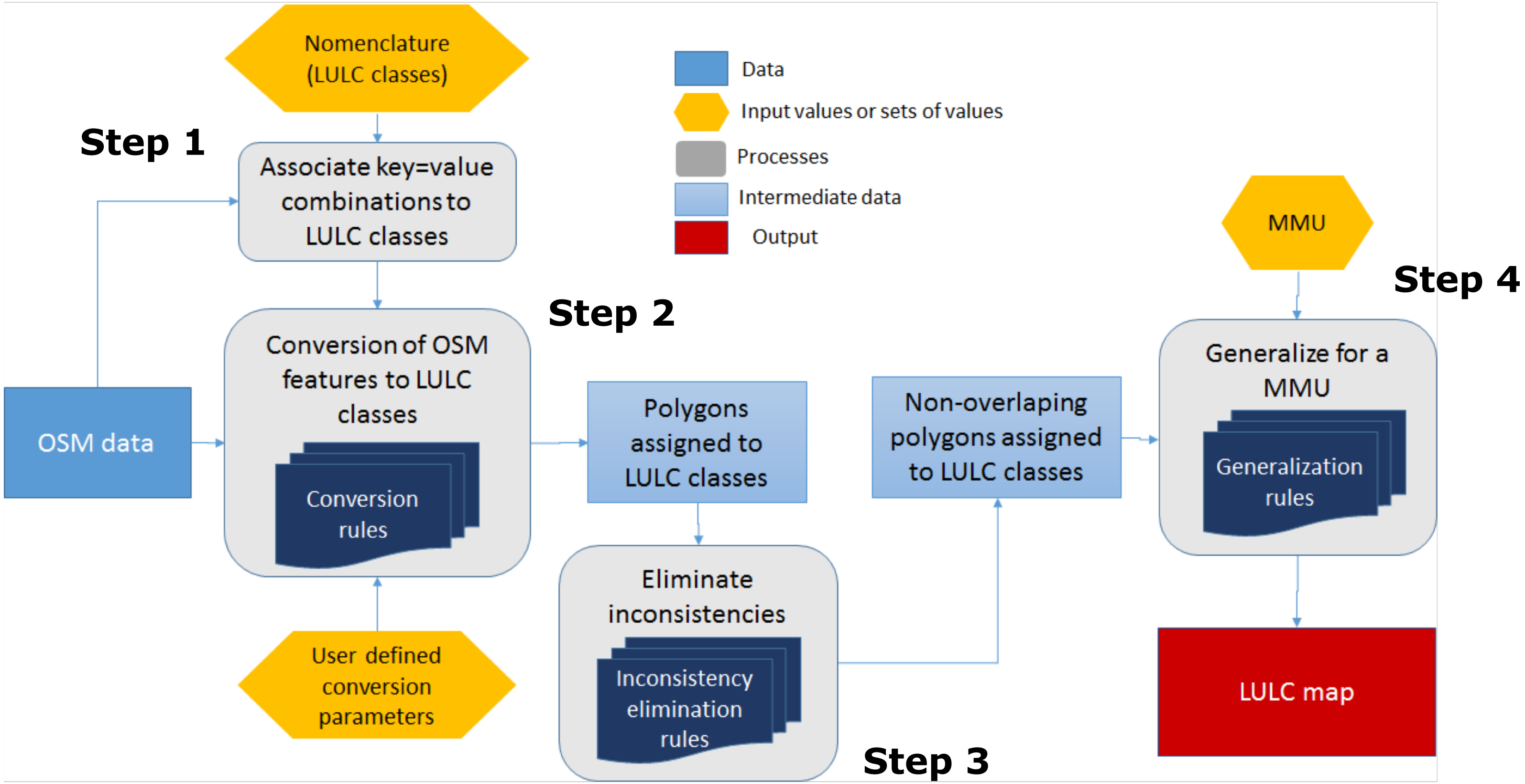
UA & CLC nomenclatures

- The nomenclatures of UA and CLC are compatible. The main difference is that more detailed urban classes can be found in UA and other land cover classes in CLC are more detailed than in UA, which reflects differences in their overall purpose.

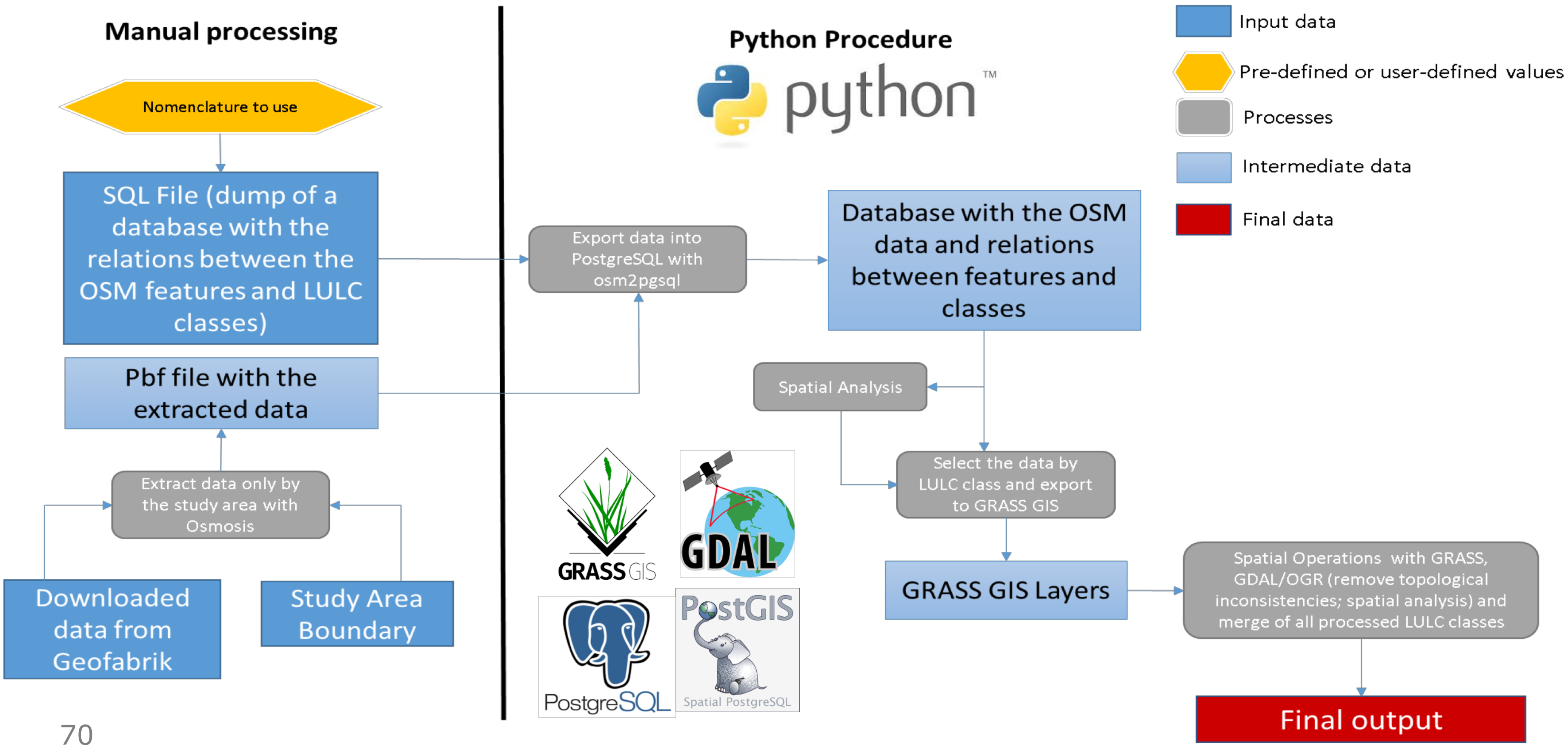
Urban Atlas nomenclature			**Corine Land Cover nomenclature**		
Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
1.Artificial Surfaces	1.1 Urban Fabric	1.1.1 Continuous urban fabric 1.1.2 Discontinuous urban fabric 1.1.3 Isolated Structures	1.Artificial Surfaces	1.1 Urban Fabric	1.1.1 Continuous urban fabric 1.1.2 Discontinuous urban fabric
	1.2 Industrial, commercial, public, military, private and transport units	1.2.1 Industrial, commercial, public, military and private units 1.2.2 Road and rail network and associated land 1.2.3 Port areas 1.2.4 Airports		1.2 Industrial, commercial, public, military, private and transport units	1.2.1 Industrial or commercial units 1.2.2 Road and rail network and associated land 1.2.3 Port areas 1.2.4 Airports
	1.3 Mine, dump and construction sites	1.3.1 Mineral extraction and dump sites 1.3.3 Construction sites 1.3.4 Land without current use		1.3 Mine, dump and construction sites	1.3.1 Mineral extraction 1.3.2 Dump sites 1.3.3 Construction sites
	1.4 Artificial non-agricultural vegetated areas	1.4.1 Green urban areas 1.4.2 Sports and leisure facilities		1.4 Artificial non-agricultural vegetated areas	1.4.1 Green urban areas 1.4.2 Sports and leisure facilities

Urban Atlas nomenclature			Corine Land Cover nomenclature		
Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
2. Agricultural, semi-natural areas, wetlands			2.Agricultural areas	2.1 Arable land	2.1.1 Non-irrigated arable land
					2.1.2 Permanently irrigated land
					2.1.3 Rice fields
				2.2 Permanent crops	2.2.1 Vineyards
					2.2.2 Fruit trees and berry plantations
					2.2.3 Olive groves
				2.3 Pastures	2.3.1 Pastures
				2.4 Heterogeneous agricultural areas	2.4.1 Annual crops associated with permanent crops
					2.4.2 Complex cultivation patterns
					2.4.3 Land principally occupied by agriculture, with significant areas of natural vegetation
					2.4.4 Agro-forestry areas
3. Forests			3. Forest and semi natural areas	3.1 Forests	3.1.1 Broad-leaved forest
					3.1.2 Coniferous forest
					3.1.3 Mixed forest
				3.2 Scrub and/or herbaceous vegetation associations	3.2.1 Natural grasslands
					3.2.2 Moors and heathland
					3.2.3 Sclerophyllous vegetation
					3.2.4 Transitional woodland-shrub
				3.3 Open spaces with little or no vegetation	3.3.1 Beaches, dunes, sands
					3.3.2 Bare rocks
					3.3.3 Sparsely vegetated areas
					3.3.4 Burnt areas
					3.3.5 Glaciers and perpetual snow
______			4. Wetlands	4.1 Inland wetlands	4.1.1 Inland marshes
					4.1.2 Peat bogs
				4.2 Maritime wetlands	4.2.1 Salt marshes
					4.2.2 Salines
					4.2.3 Intertidal flats
5. Water			5. Water	5.1 Inland waters	5.1.1 Water courses
					5.1.2 Water bodies
				5.2 Marine waters	5.2.1 Coastal lagoons
					5.2.2 Estuaries
					5.2.3 Sea and ocean

Methodology to convert OSM data into UA & CLC nomenclature

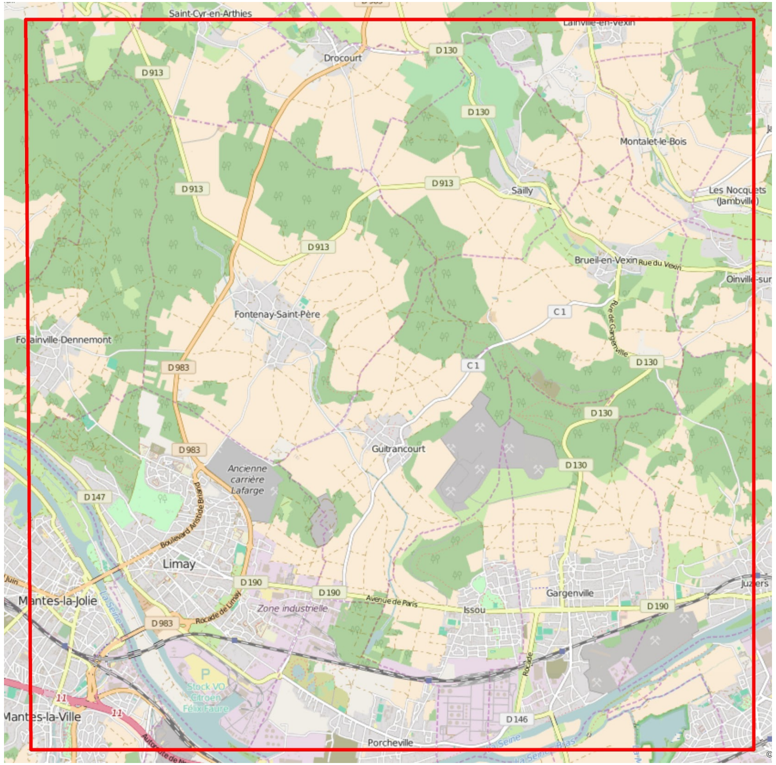


Methodology (application workflow)

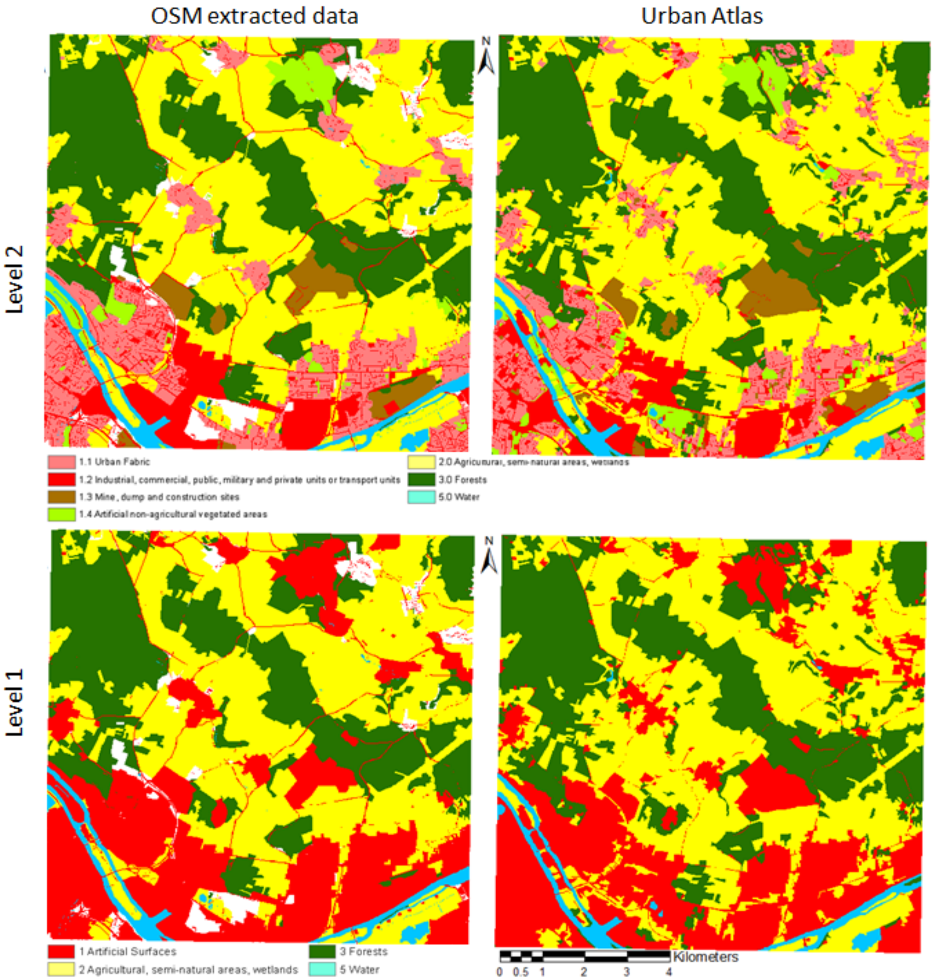


Case studies

- Paris area

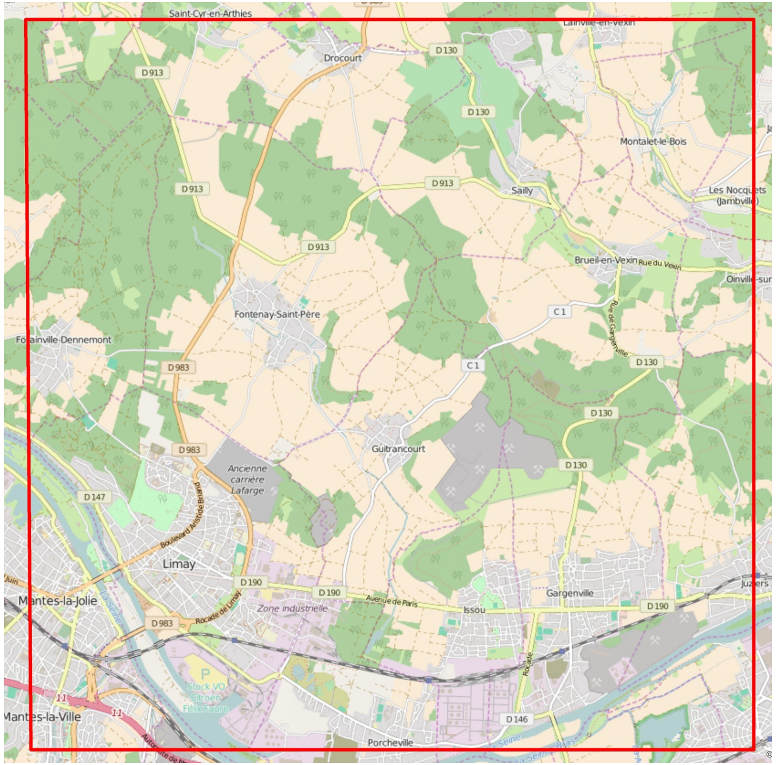


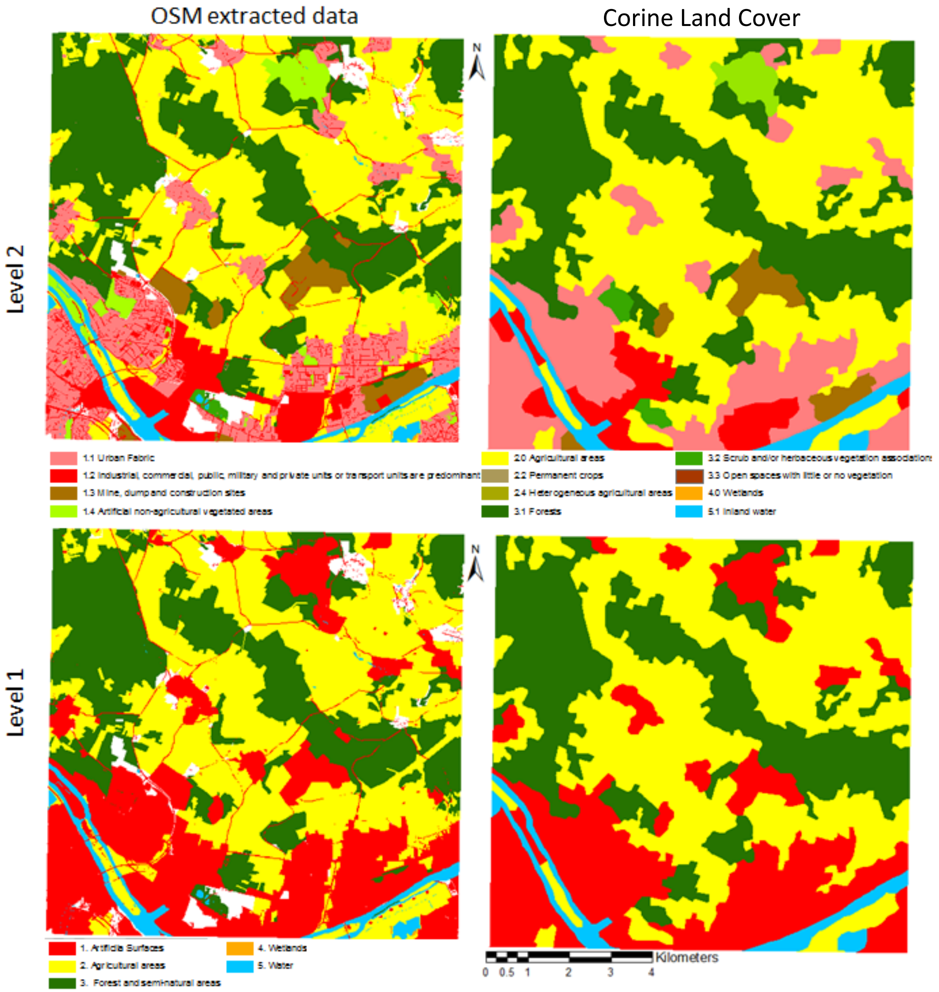
© OpenStreetMap contributors



Case studies

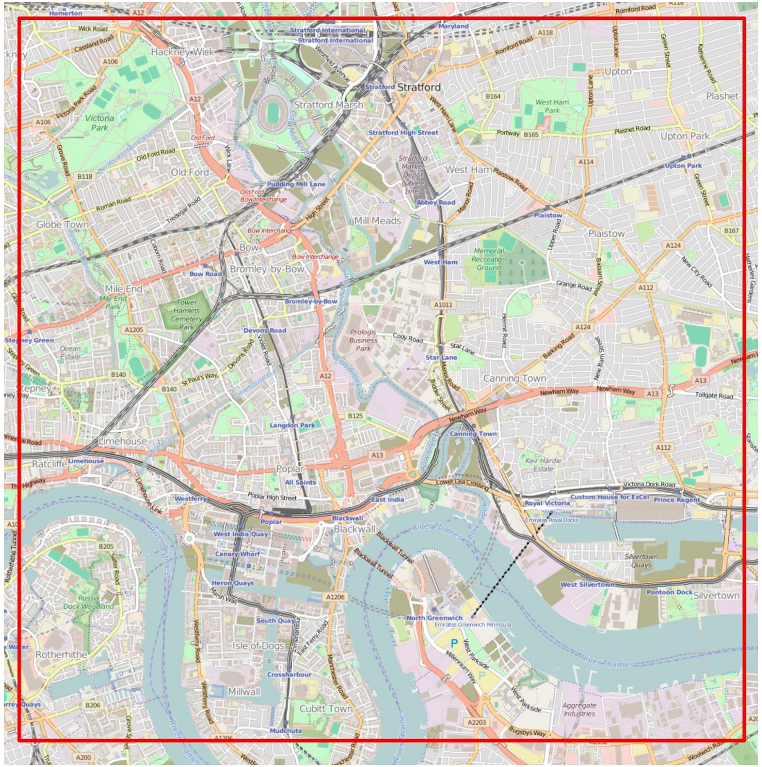
- Paris area



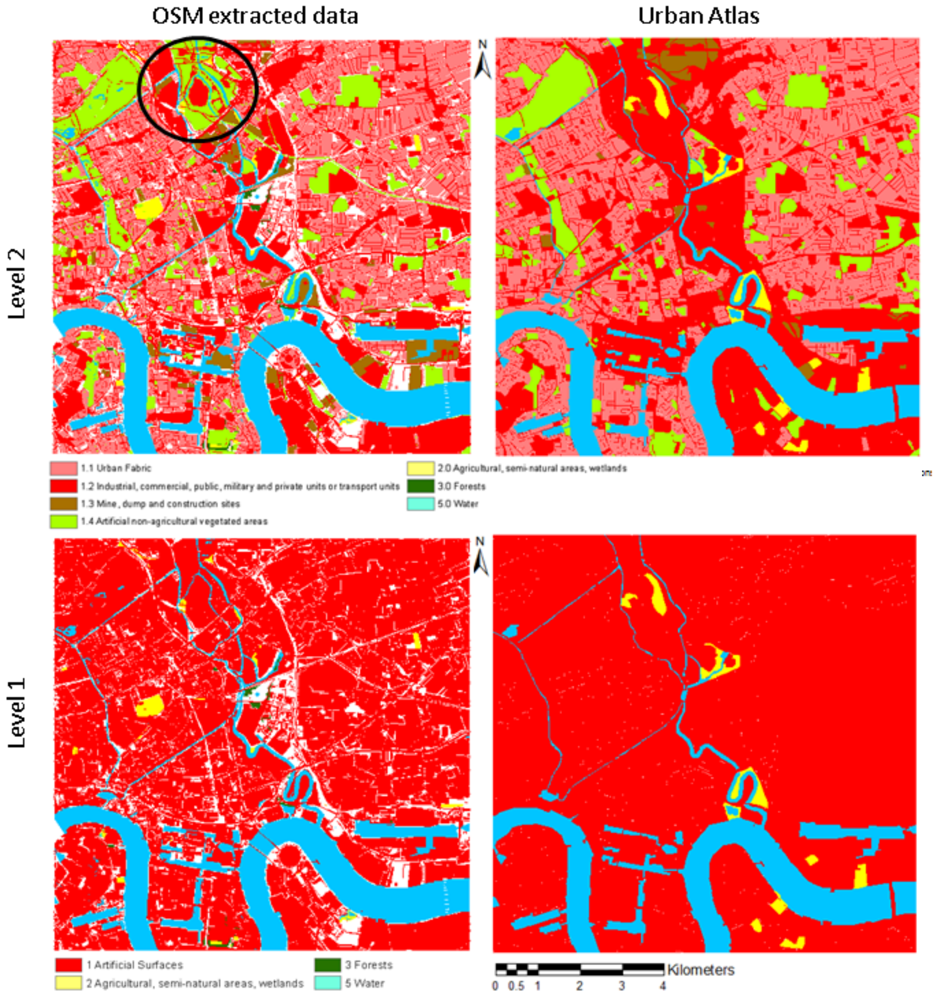


Case studies

- London area

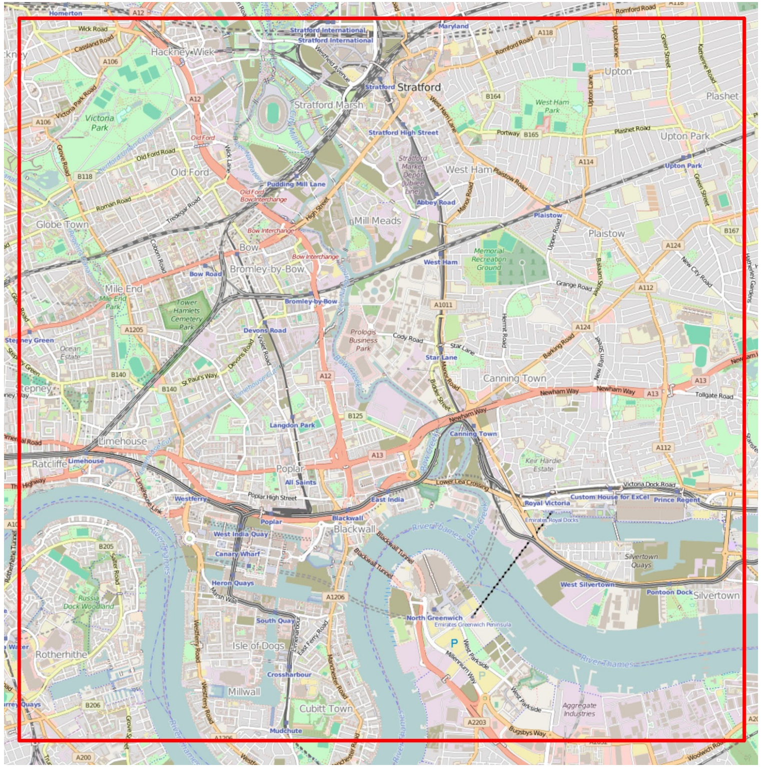


© OpenStreetMap contributors

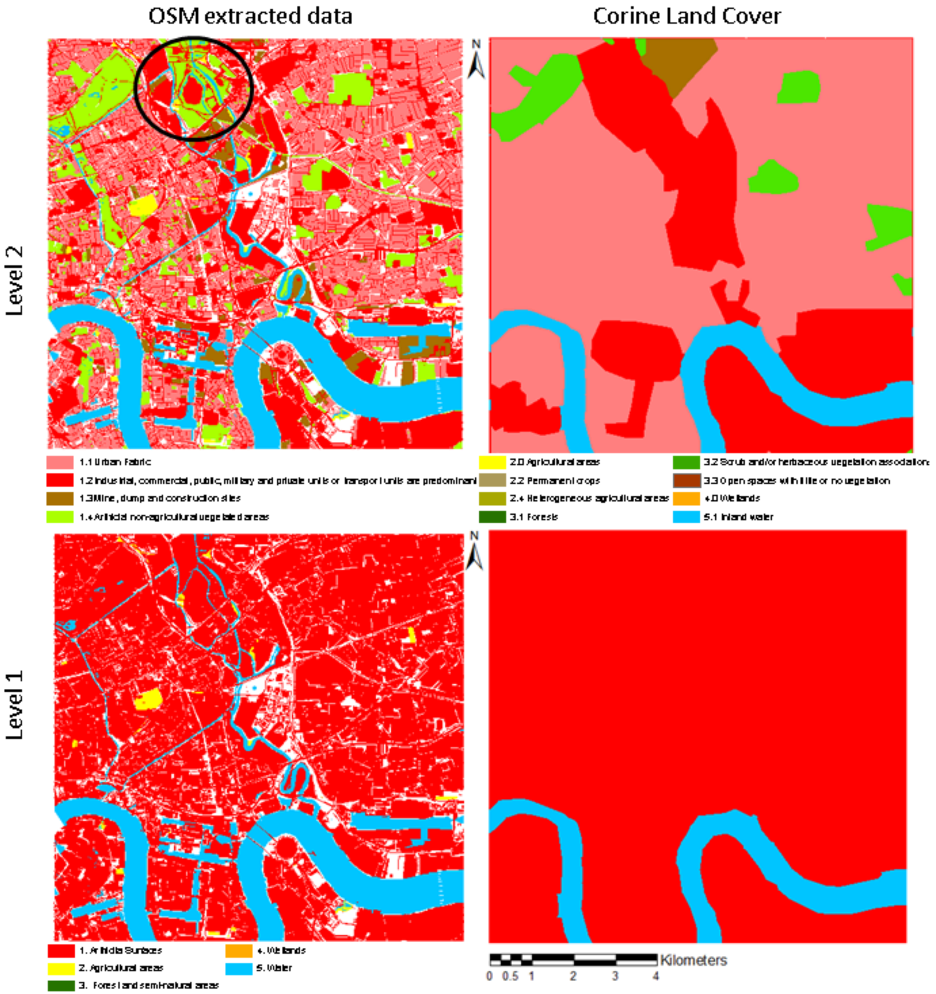


Case studies

- London area



© OpenStreetMap contributors



Case studies

- Areas [ha] occupied by level 2 classes associated to the overlapping regions in the **UA** and the map extracted from OSM, for the Paris and London study areas.

PARIS		Classes assigned to the overlapping regions in the OSM derived map							Area in UA (ha)	Match/Row Sum (%)	Match/Area in UA (%)
		11	12	13	14	20	30	50			
Classes assigned to the overlapping regions in UA	11	967	106	1	11	50	24	1	1226	83	79
	12	186	640	37	20	50	13	3	972	67	66
	13	19	24	227	0	45	7	0	330	71	69
	14	56	26	0	161	57	6	5	341	52	47
	20	108	148	33	43	3545	124	10	4163	88	85
	30	21	28	11	44	138	2425	5	2724	91	89
	50	3	4	1	1	6	5	221	242	92	91
Match/Column Sum (%)		71	66	73	57	91	93	90		84.6	

LONDON		Classes assigned to the overlapping regions in the OSM derived map							Area in UA (ha)	Match/Row Sum (%)	Match/Area in UA (%)
		11	12	13	14	20	30	50			
Classes assigned to the overlapping regions in UA	11	2346	796	16	86	8	2	21	3596	72	65
	12	525	2323	214	174	32	8	86	4023	69	58
	13	25	51	18	26	5	3	7	161	14	11
	14	19	111	5	644	17	5	18	891	79	72
	20	5	18	41	23	3	3	9	129	3	2
	30	0	0	0	0	0	0	0	0	0	0
	50	12	22	8	5	0	0	1107	1201	96	92
Match/Column Sum (%)		80	70	6	67	4	0	89		72.8	

Case studies

- Areas [ha] occupied by level 2 classes associated to the overlapping regions in the **CLC** and the map extracted from OSM, for the Paris and London study areas.

PARIS		Classes assigned to the overlapping regions in the OSM derived map								Area in CLC (ha)	Match/Row Sum (%)	Match/Area in CLC (%)
		11	12	13	14	20	31	32	51			
Classes assigned to the overlapping regions in CLC	11	1238	309	20	48	97	7	1	3	1762	72	70
	12	31	457	4	16	26	5	0	4	548	84	84
	13	0	16	167	0	31	9	0	0	224	75	75
	14	5	2	0	131	0	1	0	0	139	94	94
	20	52	132	9	37	3480	128	1	9	4109	90	85
	31	19	43	310	282	1	0	0	0	2763	0	0
	32	5	5	95	0	13	2	23	4	150	16	15
	51	9	11	1	7	45	3	0	223	303	75	74
Match/Column Sum (%)		91	47	54	46	90	0	88	91		75.5	

LONDON		Classes assigned to the overlapping regions in the OSM derived map							Area in CLC (ha)	Match/Row Sum (%)	Match/Area in CLC (%)
		11	12	13	14	20	30	51			
Classes assigned to the overlapping regions in CLC	11	2685	1861	64	448	45	8	120	5880	51	46
	12	218	1258	228	171	17	13	374	2726	55	46
	13	12	77	11	29	2	2	2	156	8	7
	14	16	120	0	312	0	0	18	478	67	65
	20	0	0	0	0	0	0	0	0	0	0
	30	0	0	0	0	0	0	0	0	0	0
	51	2	6	0	1	0	0	740	760	99	97
Match/Column Sum (%)		92	38	4	32	0	0	59		56.5	

References

Proceedings, 6th International Conference on Cartography and GIS, 13-17 June 2016, Albena, Bulgaria
ISSN: 1314-0604, Eds: Bandrova T., Konecny M.

AN AUTOMATED METHODOLOGY FOR CONVERTING OSM DATA INTO A LAND USE/COVER MAP

Cidália Fonte[(1,2)], Marco Minghini[(3)], Vyron Antoniou[(4)], Linda See[(5)], Joaquim Patriarca[(2)], Maria Antonia Brovelli[(3)], Grega Milcinski[(6)]

Using OpenStreetMap to Create Land Use and Land Cover Maps:
Development of an Application

Cidália C. Fonte[1,2], Joaquim Patriarca[2], Marco Minghini[3], Vyron Antoniou[4], Linda See[5], Maria Antonia Brovelli[3]
[1] *University of Coimbra. Portugal*
[2] *INESC Coimbra, Portugal*
[3] *Politecnico di Milano, Italy*
[4] *Hellenic Military Geographical Service, Greece*
[5] *International Institute for Applied Systems Analysis, Austria*

Open problems of VGI

- Data privacy:
  - it must be considered from the initial design of VGI systems
  - VGI must not provide linkable private data about individuals
  - is geographic information itself private data?
  - privacy does not only apply to users
- Ethical issues:
  - VGI must ensure respect of individuals, commit to nondisclosure of participants’ identities, and minimization of potential harm
  - participants should be informed of the full nature of the research
  - participants have the ethical responsibility of collecting true data
- Legal issues:
  - for producers: intellectual property, ownership and licenses
  - liability, i.e. who is liable and under which circumstances

Open problems of VGI

- Example of privacy violation from Strava (January 2018):

Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Latest: Strava suggests military users 'opt out' of heatmap as row deepens

A military base in Helmand Province, Afghanistan with route taken by joggers highlighted by Strava. Photograph: Strava Heatmap

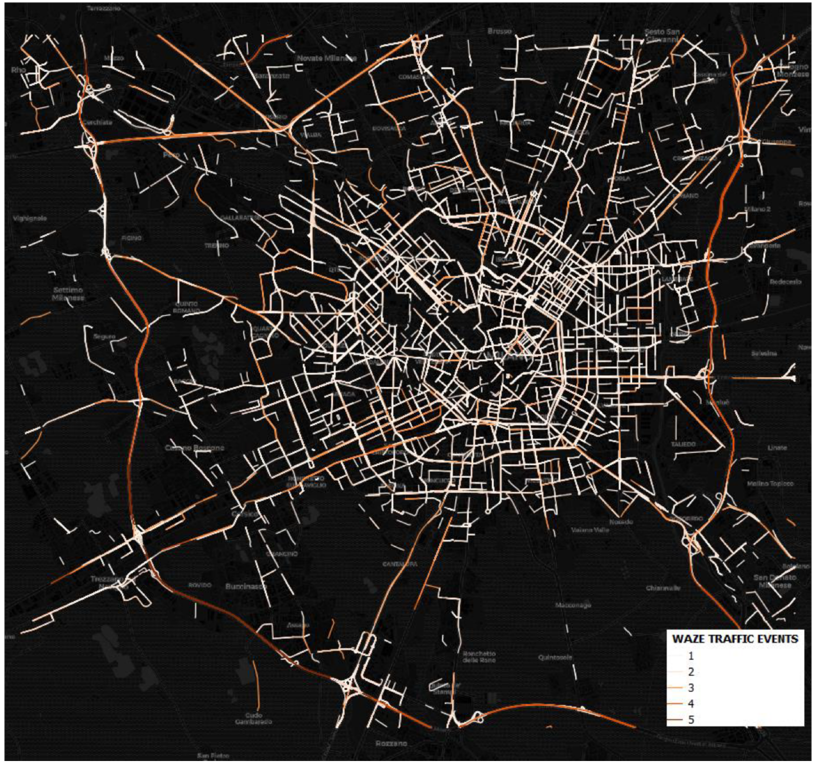
Source: https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

Open problems of VGI

- A typical solution to protect privacy of individuals is to blur personal data, e.g. car plates ad human faces in photographs (example: Mapillary).

Open problems of VGI

- A typical solution to protect privacy of individuals is to anonimyze and aggregate data, e.g. for traffic data (example: Waze).



Thank you!

Questions?

marco.minghini@ec.europa.eu

Stay in touch

EU Science Hub: ***ec.europa.eu/jrc***

Twitter: ***@EU_ScienceHub***

Facebook: ***EU Science Hub - Joint Research Centre***

LinkedIn: ***Joint Research Centre***

YouTube: ***EU Science Hub***

European Commission