RDA 9th Plenary Meeting *Data Infrastructures for Open Science* | 5-7 April 2017, Barcelo Sants Hotel - Barcelona, Spain

Volunteered Geographic Information and OpenStreetMap: Experimentations and Perspectives

POLITECNICO MILANO 1863

Marco Minghini

Politecnico di Milano, Dept. of Civil and Environmental Engineering
GEOlab | Piazza Leonardo da Vinci 32, 20133 Milan, Italy

RDA RESEARCH DATA ALLIANCE

Abstract

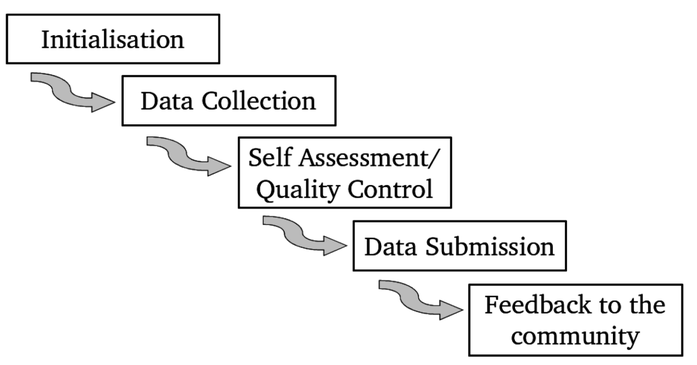
Volunteered Geographic Information (VGI) is the most successful term to mark the shift (happened almost a decade ago) according to which citizens, in addition to being data consumers, have also become precious providers of geographic information. The increasingly large volumes of datasets generated by VGI projects have attracted a huge academic interest in a number of topics, ranging from data quality assessment to data mining and data conflation, societal studies as well as ethical and privacy issues. The most popular VGI project, **OpenStreetMap (OSM)**, which aims at the generation of a crowdsourced map of the world, can be nowadays considered as a research topic on its own (Mooney & Minghini 2017).

In the poster a wide range of the author's current research works on VGI and OSM is shown. These works partially derive from the author's involvement in three (current or just ended) EU COST Actions: TD1202 "Mapping and the Citizen Sensor", IC1203 "ENERGIC", and CA15212 "Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe".

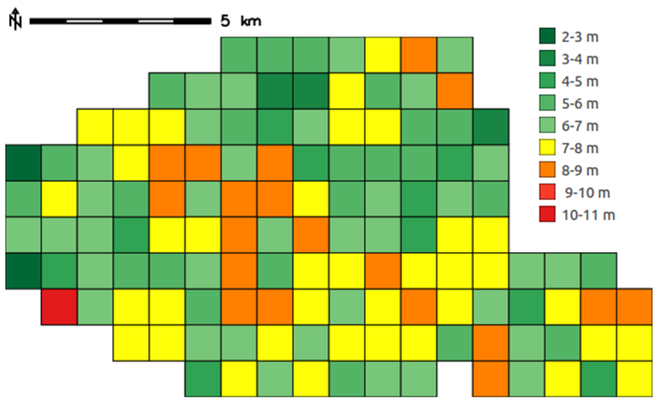
Development of Protocols for VGI Data Collection

Typically, VGI projects offer simple guidelines and recommendations instead of rigorous **data collection protocols**, and this results into **variable data quality**. Protocols ensure standardization in data collection and impact on VGI quality and usability. They must be designed to balance the needs of providing detailed instructions while **maintaining contributors' motivation** (Minghini et al. 2017).

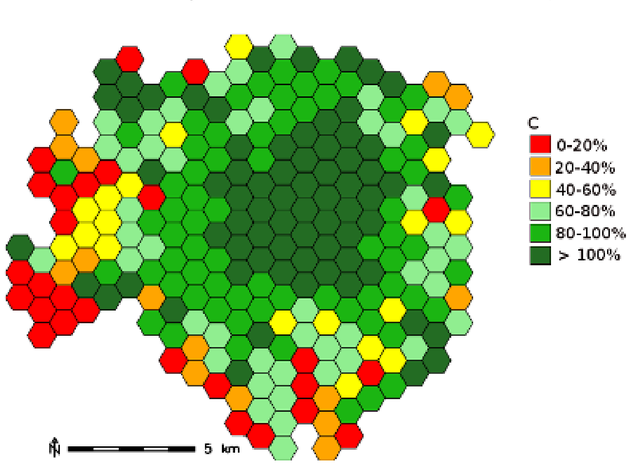
A generic protocol is developed to guide data collection in new and existing VGI projects (Mooney et al. 2016). It is suitable to novice and experience contributors and is based on existing quality standards and successful practices. The protocol is initially specified for **VGI vector data** and covers the specifications of data model (geometry and attributes), collection methods (dos and don'ts, demos and good practices) and characteristics (CRS, topology, scale, etc.). It addresses three types of data collection: **digitization** from imagery, **field survey** (e.g. through GPS-enabled devices) and **bulk import**. Future work should stimulate the adoption of the protocol in terms of future software implementations for VGI data collection and verify/assess its goodness on real VGI projects.



Sequential stages for implementing the VGI data collection protocol



Positional accuracy assessment of OSM roads in Paris, France



Completeness assessment of OSM buildings in Milan, Italy

Quality Assessment of OpenStreetMap road network and building datasets

In contrast to authoritative data, VGI is produced without any indication of its quality. **VGI quality assessment** is an extremely hot topic of research and most of the available analyses are focused on OSM. Two procedures are presented in the following, which extrinsically assess the **quality of OSM road networks and OSM buildings** through the comparison against their authoritative counterparts.

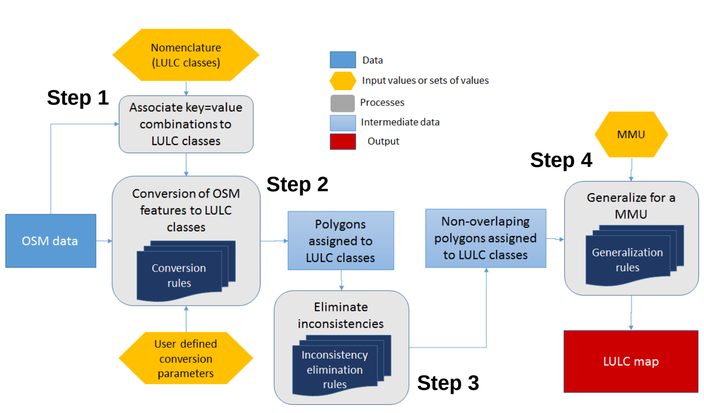
1. 3 independent modules for the open source GRASS GIS software, written in Python and available with a Graphical User Interface (GUI), are developed to assess the **completeness** and **positional accuracy** of OSM road network datasets (Brovelli et al. 2016a):
   - **v.osm.precomp**: performs a **preliminary comparison** of the OSM and reference road network datasets, computes some global statistics, and evaluates their geometrical similarity.
   - **v.osm.preproc**: performs a **geometric preprocessing** of the OSM road network dataset, which only extracts its subset of roads having a correspondence in the reference dataset.
   - **v.osm.acc**: evaluates the **positional accuracy** of the OSM road network dataset using a **grid-based approach**. For each grid cell, the module: 1) computes the maximum deviation of the OSM dataset from the reference dataset; and 2) evaluates the positional accuracy of the OSM dataset against one or more thresholds specified by the user.

   The procedure was applied in two research works to assess the quality of the OSM road networks of: **Coimbra, Portugal** through comparison against the dataset provided by the Coimbra City Hall (Antunes et al. 2015); and **Paris, France** through comparison against the French official dataset provided by IGN (Brovelli et al. 2016b).
2. Quality assessment of OSM buildings in Milan, Italy is done by comparison with the official vector cartography (Brovelli et al. 2016c):
   - **completeness**, assessed using the area-ratio unit based method, is shown to **decrease from the city center to the outskirts**.
   - **spatial accuracy** is assessed using an advanced algorithm allowing a quasi-automated detection of homologous points based on geometrical, topological and semantic analyses. A total of about 100k homologous points are found, which show the same **very high positional accuracy in both the city center and the outskirts** (mean value < 1m). Through the algorithm, a range of least squares transformations (including affine transformations and transformations based on multi-resolution splines) can be also applied to the OSM building dataset in order to find a new, optimal version to be then included in the OSM database.

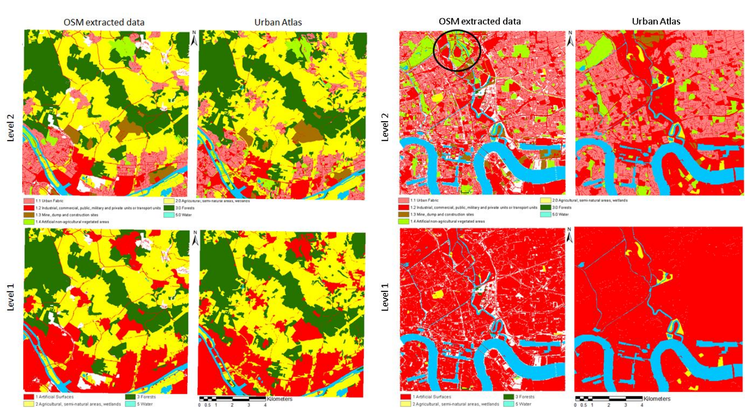
Creation of Land Use/Land Cover maps from OpenStreetMap

LULC maps, produced from the classification of satellite imagery and on-field validation, are crucial for many applications, however they have high production costs and are updated only every few years. In contrast, OSM is openly available and is extremely **detailed, rich and up-to-date** in terms of LULC information included.

This research aims at delivering an **automated procedure** to convert OSM data into LULC maps having the nomenclature of official LULC maps: Urban Atlas (UA), Corine Land Cover (CLC) and GlobeLand30 (GL30) (Fonte et al. 2016). The procedure is written in Python and makes full use of **open source software**: GRASS GIS, PostgreSQL/PostGIS and GDAL/OGR. It is composed of 4 steps (Fonte et al. 2017): association of OSM tags to the corresponding LULC classes, according to the chosen nomenclature; conversion of OSM objects into polygon objects for each LULC target class; **elimination of inconsistencies** like overlapping regions assigned to different LULC classes; **generalization** according to a predefined Minimum Mapping Unit (MMU).



Sequential steps of the procedure to convert OSM into LULC maps



OSM-derived LULC maps in UA nomenclature: Paris (left), London (right)

Geotagged photos for LC purposes

Geotagged photos, available from online repositories such as Panoramio, Flickr, Geograph, Instagram, etc. are an increasingly used source of VGI. The purpose of this work is to evaluate **protocols for the collection of geotagged photos** and study the usability of geotagged photos for LC applications: **calibration**, **validation** and **verification**.

There are few protocols for **social networking sites** (Facebook, Twitter, Google+, Foursquare, etc.) and **photo sharing sites** (Panoramio, Picasa, Flickr, Instagram, etc.) compared to **sites for landscape documentation** (Degree Confluence Project, Geograph, Geo-Wiki, etc.). Metadata show as well a high variation, and in social networking and photo sharing sites photos can even lack location information.

A total of **1000 photos** collected in London in May 2015 are randomly extracted from each of Flick, Panoramio and Geograph websites. **Photos usefulness for LC applications** is evaluated by 7 experts by assigning to each – based on some rules – one of the following:

- **YES**: only one LC type could be clearly seen in the photo;
- **MAYBE**: more than one LC type could be identified in the photo;
- **NO**: no useful evidence of LC is available.

Results show that, for all the three sites, **more than 50% of the photos are considered to be useful**; for Geograph, this increases to 72%. Thus, despite protocols for collecting geo-tagged photos can be improved, there is already clearly potential for using them for LC applications.

References

- Antoniou V., Fonte C.C., See L., Estima J., Jokar Arsanjani J., Lupia F., Minghini M., Foody G. & Fritz S. (2016) Investigating the feasibility of geo-tagged photographs as sources of land cover input data. *ISPRS International Journal of Geo-Information* 5(5), 64. doi:10.3390/ijgi5050064.
- Antunes F., Fonte C.C., Brovelli M.A., Minghini M., Molinari M.E. & Mooney P. (2015) Assessing OSM Road Positional Quality with Authoritative Data. *Proceedings of the VIII Conferência Nacional de Cartografia e Geodesia*, Lisbon, October 29-30, 2015.
- Brovelli M.A., Minghini M., Molinari M.E. & Mooney P. (2016a) Towards an automated comparison of OpenStreetMap with authoritative road datasets. *Transactions in GIS*. doi:10.1111/tgis.12182.
- Brovelli M.A., Minghini M. & Molinari M.E. (2016b) An automated GRASS-based procedure to assess the geometrical accuracy of the OpenStreetMap Paris road network. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B7, 919-925. doi:10.5194/isprsarchives-XLI-B7-919-2016.
- Brovelli M.A., Minghini M., Molinari M.E. & Zamboni G. (2016) Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection: the method and a case study. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B2, 615-620. doi:10.5194/isprsarchives-XLI-B2-615-2016.
- Fonte C.C., Minghini M., Antoniou V., See L., Patriarca J., Brovelli M.A. & Milcinski G. (2016) An automated methodology for converting OSM data into a land use/cover map. *Proceedings of the 6th International Conference on Cartography & GIS*, Vol. 1, 462-473, Albena (Bulgaria), June 13-17, 2016, Bulgarian Cartographic Association.
- Fonte C.C., Patriarca J., Minghini M., Antoniou V., See L. & Brovelli M.A. (2017) Using OpenStreetMap to Create Land Use and Land Cover Maps: Development of an Application. In: *Volunteered Geographic Information and the Future of Geospatial Data*. IGI Global, 113-137. doi:10.4018/978-1-5225-2446-5.ch007.
- Minghini M., Antoniou V., Fonte C.C., Estima J., Olteanu-Raimond A.-M., See L., Laakso M., Skopeliti A., Mooney P., Arsanjani J.J. & Lupia F. (2017) The relevance of protocols for VGI collection. In: *Mapping and the Citizen Sensor*. Ubiquity Press.
- Mooney P. & Minghini M. (2017) A review of OpenStreetMap data. In: Foody G., See L., Fritz S., Mooney P., Olteanu-Raimond A.-M., Fonte C.C. and Antoniou V. (Eds) *Mapping and the Citizen Sensor*. Ubiquity Press.
- Mooney P., Minghini M., Laakso M., Antoniou V., Olteanu-Raimond A.-M. and Skopeliti A. (2016) Towards a Protocol for the Collection of VGI Vector Data. *ISPRS International Journal of Geo-Information* 5(11), 217. doi:10.3390/ijgi5110217.