# Quality Assessment of Semantic Tags in OpenStreetMap

**E J Hoffmann[1], M Werner[2] and X X Zhu[1,2]**

[1]Signal Processing in Earth Observation (SiPEO), Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

[2]Remote Sensing Institute (IMF), German Aerospace Center (DLR), Mnchener Str. 20, 82234 Weling, Germany

E-mail: eike.jens.hoffmann@tum.de, martin.werner@dlr.de, xiaoxiang.zhu@dlr.de

**Abstract.** A comprehensive understanding of urban areas includes deep knowledge about locations and functions of buildings within. In most developed countries cadastral data is available for this research, but from a global perspective the only free and comprehensive data source is OpenStreetMap (OSM).

For this study we selected 42 cities across the globe covering a wide range of climate zones as well as cultures and assess the accuracy of building function labels in OSM indirectly by comparing them to Google Places. We state that points-of-interest (POIs) are reasonably covered in Google Places due to a large number of users and a business perspective driving its development. We study how many semantic building tags are in accordance with the proposed scheme by OSM. In this regard Los Angeles has the best coverage, followed by Amsterdam and Cologne. In Melbourne, Paris, and Sydney we find the most matching building functions of OSM and Google Places.

In summary, we conclude that OSM is not ready to provide ground truth labels for each place on the globe, but can serve as a powerful and rich label source in selected study areas. Our study gives a first insight where obtaining training labels from OSM is a valid and reliable approach.

## 1. Introduction and Related Work

Since official data on urban land use is not available on a global scale the most comprehensive data source for this information is OpenStreetMap (OSM), a volunteered geo-information project founded in 2004 [1]. Since then it has become a widely used source for researchers of all areas and its data quality has been analyzed in several ways, e.g. [2]. These studies showed that the completeness of OSM varies a lot: in the western hemisphere it is much more comprehensive than is other ares [3]. A fact that is strongly related to the activity of contributing users mapping their surrounding area.

Additionally, OSM focuses on street topology rather than build-up structures. The latter are covered well in terms of geometries, but not with respect to semantic tags. A lot of contributors draw accurate shapes of building polygons but semantic information is rare due to ambiguities in reality, which are hard to cover in a formal naming schema. Nevertheless, OSM proposes a XML-tag *building*[1] together with a naming schema.

[1] https://wiki.openstreetmap.org/wiki/Key:building

## 2. Result

In this study we compared semantic building information from OSM with data retrieved from Google Places, a commercial database covering points-of-interest (POI) on a global scale. We selected 42 metropolitan areas from six continents: Asia (17), Africa (3), North America (5), South America (3), Europe (12) and Australia (2) for a diverse sampling. Our definition of the area shapes follows roughly the administrative boundaries in OSM. Figure 1 summarises our results.

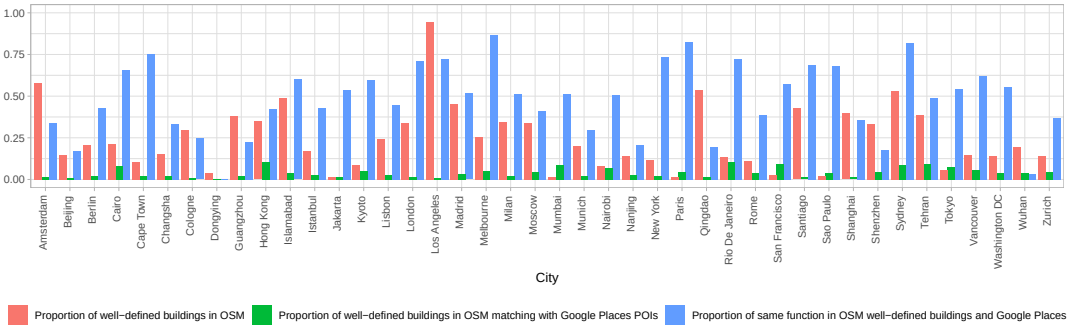


**Figure 1.** Relative numbers of well-defined buildings according to OSM schema compared to all OSM buildings in a city (red), relative number of well-defined OSM buildings matching a Google Places entry (green), and relative number of agreement on function between matched buildings (blue)

## 3. Conclusion and Outlook

In this work, we discuss the 2018 status of semantic information in OSM related to building use by comparing this to a sample from the reliable and up-to-date Google Places database. While the coverage of this study is limited as we need to restrict to those places for which we have both OSM polygons and a Google Places entry, it clearly leads to the following general conclusions that have already been drawn in other contexts:

First, the OSM data is very incomplete and while the non-standardized tagging of OSM objects made it a very flexible map, the tag use is also a very fragmented.

Second, the OSM is in general not ready to be used as a ground truth for urban land use analysis. It needs manual completion and validation on a region-by-region basis.

On the positive side, we can, however, select well-mapped cities for pre-studies. The best city to start with might be Los Angeles, in which most buildings are clearly labeled residential or commercial.


## Acknowledgements

This work is jointly supported by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-StG-714087], Acronym: So2Sat), Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018, www.sipeo.bgu.tum.de), and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg.


## References


[1] Jokar Arsanjani J, Zipf A, Mooney P, Helbich M. An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications; 2015. p. 1–15.

[2] Jokar Arsanjani J, Mooney P, Zipf A, Schauss A. Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets; 2015. p. 37–58.

[3] Neis P, Zipf A. Analyzing the Contributor Activity of a Volunteered Geographic Information Project The Case of OpenStreetMap. ISPRS International Journal of Geo-Information. 2012;1(2):146–165.