

# The OpenBuildingMap dataset

([https://doi.org/10.5880.GFZ.LKUT.2025.002](https://doi.org/10.5880/GFZ.LKUT.2025.002))

---

Laurens J.N. Oostwegel<sup>1,2</sup>, Doren Çalliku<sup>1</sup>, Tara Evaz Zadeh<sup>1</sup>, Pablo de la Mora<sup>1</sup>, Chengzhi Rao<sup>1</sup>, Danijel Schorlemmer<sup>1</sup>

1. *GFZ Helmholtz Centre for Geosciences, Potsdam, Germany*
2. *ISTerre, Université Grenoble Alpes, Université Savoie Mont-Blanc, CNRS, IRD, Université Gustave Eiffel, CS40700 38058 Grenoble cedex 9, 1381 Rue de la Piscine, 38610 Gières, France*

## 1. Licence

Open Data Commons Open Database License (ODbL) v1.0

## 2. Citation

**When using the data please cite:**

Oostwegel, L. J. N.; Schorlemmer, D.; Lingner, L.; Evaz Zadeh, T. (2025): OpenBuildingMap. GFZ Data Services. <https://doi.org/10.5880/GFZ.LKUT.2025.002>

**The data are supplementary material to:**

Oostwegel, L. J. N.; Schorlemmer, D.; Guéguen, P. (2025): From Footprints to Functions: A Comprehensive Global and Semantic Building Footprint Dataset. Scientific Data. <https://doi.org/10.1038/s41597-025-06132-z>

## Table of Contents

1. Licence.....	1
2. Citation.....	1
3. Data description.....	1
4. File description.....	2
4.1. File inventory.....	2
4.2. Description of data tables.....	3
4.2.1. Building table.....	3
4.2.2. Metadata table.....	3
5. References.....	4

## 3. Data description

This data publication contains the OpenBuildingMap dataset, a global assessment of building footprints, organized in a tiled grid. The dataset provides information about the geometry; occupancy type; height; and floorspace per building. It can be used for a wide range of purposes, like disaster risk assessment, urban planning or sustainability analysis. All OpenStreetMap (OSM) buildings existing in an OSM excerpt from 1 July 2024, 00:00 UTC (OpenStreetMap contributors, 2024), all buildings from the Global ML Building Footprint (Microsoft, 2024) dataset, and all buildings from the Google Open Buildings (Sirko et al., 2021) have been processed and for each building the occupancy type and number of stories have been identified based on data in OSM, such as land use and points of interest. The Global Human Settlement Built-up Characteristics 2023A Layer has been used for height estimation (Schiavina et al., 2023).

The resulting dataset is distributed per country as an GeoPackage, a widely used open file standard for geographical files. Each file contains two tables: Metadata and Building. In the Metadata file there is one row, that provides the license, the number of buildings in the tile, as well as the percentage with known occupancy types, height, floorspace and the share of buildings stemming from the OSM, Microsoft and Google datasets. The Building table contains all the building footprints, with the relevant attributes (building ID, floorspace, height, occupancy type, quadkey, building source dataset). The height and occupancy type are defined using the GEM Building Taxonomy v2.0 (Brzev et al., 2013). The Quadkey is a level-18 tile according to the Quad Tree structure (Finkel & Bentley, 1974), with the tiles projected using the Web Mercator projection (EPSG:3857). The method is described in more detail in Oostwegel et al. (2025).

## 4. File description

### 4.1. File inventory

The building model is organized per level-6 Quadkey tile. *2025-002\_Oostwegel-et-al\_data* includes 1271 files, where each is named `building`, followed by the Quadkey ID. The files are compressed with Bzip2. There are multiple ways to find out which Quadkey a region of interest belongs too.

- First, an overview of Quadkey tiles is given on the map in this data publication (<https://doi.org/10.5880.GFZ.LKUT.2025.002>).
- Second, on the webpage <https://www.openbuildingmap.org/download/>, a tile can be directly selected on a map and downloaded.
- Third, the website <https://labs.mapbox.com/what-the-tile/> can provide information on which Quadkey ID belongs to a region of interest.

All files are licensed under the Open Data Commons Open Database Licence (OdbL) v1.0.

File name
building.002202.gpkg.bz2
building.002213.gpkg.bz2
building.002220.gpkg.bz2
...

*List of file names.*

### Additional files

The folder *2025-002\_Oostwegel-et-al\_additional-files* includes five files that support the data publication. The first file (*A\_Osmium\_mapping.yaml*) contains the Osmium mapping that was used to create the initial database with OpenStreetMap data. The second file (*B\_buildings\_and\_Poi\_tags.csv*) contains a mapping between OpenStreetMap tag values and the resulting GEM Taxonomy tag. In the third file (*C\_occupancy\_types.csv*), the explanation of each of the occupancy types is found. The fourth file (*D\_overriding\_occupancies.csv*) contains the list of overriding occupancy types. The last file (*E\_KullbackLeibler\_cities.csv*) contains the completeness index and Kullback-Leibler divergence of the three biggest cities in each country, that have at least 100.000 inhabitants.

## 4.2. Description of data tables

### 4.2.1. Building table

The Building table contains all information about the building footprints.

Column header	Unit	Description
Id	Integer	Building ID. For OSM, this is the OpenStreetMap ID. For the Google / Microsoft datasets this ID is created by us
Floorspace	Real	
Occupancy	String	Occupancy type of the building, according to the GEM taxonomy
Height	String	Height and/or number of stories of the building, according to the GEM taxonomy
Quadkey	String	Level-18 Quadkey ID
Source ID	Integer	Source dataset identifier: - 0: OpenStreetMap - 1: Google - 2: Microsoft
Relation ID	Integer	Only relevant for OpenStreetMap buildings. If a building is part of another, the ID of the entire building is provided here.
Last update	Datetime	Time of processing the building footprint.

*Description of the Building table columns.*

### 4.2.2. Metadata table

The Metadata table provides some statistics about the buildings inside this tile, as well as general information.

Column header	Unit	Description
License	String	The license of the dataset. In all cases OdbL v1.0.
number_of_buildings	Integer	The total number of buildings in the file.
percentage_known_occupancy	Real	The percentage of buildings in the file where the occupancy type is not unknown (UNK).
percentage_known_height	Real	The percentage of buildings in the file where there is a height estimation.
percentage_known_floorspace	Real	The percentage of buildings in the file with a floorspace estimation.
percentage_source_openstreetmap	Real	The percentage of buildings sourced from OpenStreetMap.
percentage_source_google	Real	The percentage of buildings sourced from Google Open Buildings.
percentage_source_microsoft	Real	The percentage of buildings sourced from the Microsoft Global ML Building Footprints.

*Description of the Metadata table columns.*

## 5. References

Brzev, S., Scawthorn, C., Charleson, A. W., Allen, L., Greene, M., Jaiswal, K., & Silva, V. (2013). GEM Building Taxonomy Version 2.0. GEM Technical Report 2013-02. <https://doi.org/10.13117/GEM.EXP-MOD.TR2013.02>

Finkel, R. A., & Bentley, J. L. (1974). Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1), 1–9. <https://doi.org/10.1007/BF00288933>

Microsoft. (2023). Microsoft Global ML Building Footprints [dataset]. <https://github.com/microsoft/GlobalMLBuildingFootprints>

OpenStreetMap contributors. (2023) Planet dump [Data file from 1 July 2024, 00:00 UTC]. Retrieved from <https://planet.openstreetmap.org>

European Commission: Joint Research Centre (2023). GHSL Data Package 2023, Publications Office of the European Union, Luxembourg, 2023, JRC133256, <https://doi.org/10.2760/098587>, <https://op.europa.eu/s/z6PL>

Sirko, W. et al. (2021) Continental-Scale Building Detection from High Resolution Satellite Imagery, Version Number: 2, <https://doi.org/10.48550/arxiv.2107.493.12283>