A model of European buildings
(https://doi.org/10.5880/GFZ.2.6.2023.011)

Danijel Schorlemmer, Tara Evaz Zadeh, Lingner, Laurens J.N. Oostwegel

GFZ German Research Centre for Geosciences, Potsdam, Germany

1. Licence

2. Citation

When using the data please cite:

Schorlemmer, D.; Evaz Zadeh, T.; Lingner, L.; Oostwegel, L. J.N. (2023): A model of European buildings. GFZ Data Services. https://doi.org/10.5880/fidgeo.2023.011

3. Data Description

This data publication provides a European assessment of building exposure, organized country-by-country. The dataset provides information about the number of buildings; the number of occupants; structural information and structural costs of buildings per geographical area. The main purpose of this data collection is risk assessment for natural hazards, however it can be used by anyone in need of a building exposure dataset.

The data holds information about single buildings, with global estimates of built-up area on 10m x 10m pixels and exposure information per district. All OpenStreetMap (OSM) buildings existing in an OSM excerpt from 1 July 2023, 00:00 UTC (OpenStreetMap contributors, 2023), all buildings from the Global ML Building Footprint (GMLBF, Microsoft, 2023) dataset 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 2022A Layer has been used as initial distribution of built area (Pesaresi, 2022). Aggregated exposure information, including the structural information and the number of occupants, stems the ESRM20 (Crowley et al., 2020).

The resulting dataset is distributed per country as an SQLite/SpatiaLite database. Each database contains three tables and one view. The database is organized around three key concepts, that each have their own table. An *Entity* is a geographical unit that contains exposure. In this dataset, the *entities* are tiles in a multi-resolution grid, according to the Quad tree structure (Finkel & Bentley, 1974), with

the tiles projected using the Web Mercator projection (EPSG:3857). The zoom-level of the Quadkeys inside the grid varies from level-15 to level-18, depending on the number of buildings inside each tile to preserve privacy-sensitive information. Practically, the size of the tiles varies between around 100m x 100m and 1km x 1km. Each entity consists of one or more *Assets*, defining the number of buildings of a particular structural type and their population and structural value. The structural type is described using a taxonomy string, describing for example structural properties, occupancy type and the expected number of stories. The exact definition of a taxonomy that is used in this dataset is described in the GEM Building Taxonomy v2.0 (Brzev et al., 2013). On top of the tables, one key view has been defined too. A view is essentially a query on the table that give some insights into the data. The `key_values_per_tile` provides the total number of buildings, total number of occupants at night and total structural costs summed over all assets in one tile entity.

4. File description

The exposure model is organized per country or region. *2023-011_Schorlemmer_et_al* prefixes each file, followed by the ISO 3166-1 alpha-3 code and the name of the country or region. The source of the exposure attributes is ESRM20, if it exists. Some regions do not exist in ESRM20 (e.g. Faroe Islands). The population value for these regions is averaged by the total population and the structural properties and value are not considered. The building footprints are sourced from OSM and the GMLBF. If the country is considered complete in OSM, the GMLBF dataset is excluded.

File name **Prefix for all filenames: 2023-011_Schorlemmer-et-al_**	**Country / Region**	**Source data**
ALB.Albania.db	Albania	ESRM20, OSM, GMLBF, GHSL
AND.Andorra.db	Andorra	ESRM20, OSM, GMLBF, GHSL
AUT.Austria.db	Austria	ESRM20, OSM, GHSL
BEL.Belgium.db	Belgium	ESRM20, OSM, GMLBF, GHSL
BGR.Bulgaria.db	Bulgaria	ESRM20, OSM, GMLBF, GHSL
BIH.Bosnia_and_Herzegovina.db	Bosnia and Herzegovina	ESRM20, OSM, GMLBF, GHSL
BSA.British_Sovereign_Areas.db	British Sovereign Areas	ESRM20, OSM, GMLBF, GHSL
CHE.Switzerland.db	Switzerland	ESRM20, OSM, GHSL
CYP.Cyprus.db	Cyprus	ESRM20, OSM, GMLBF, GHSL
CZE.Czechia.db	Czechia	ESRM20, OSM, GHSL
DEU.Germany.db	Germany	ESRM20, OSM, GHSL
DNK.Denmark.db	Denmark	ESRM20, OSM, GHSL
ESP.Spain.db	Spain	ESRM20, OSM, GMLBF, GHSL
EST.Estonia.db	Estonia	ESRM20, OSM, GHSL
FIN.Finland.db	Finland	ESRM20, OSM, GMLBF, GHSL
FRA.France.db	France	ESRM20, OSM, GHSL
FRO.Faroe_Islands.db	Faroe Islands	OSM, GHSL
GBR.United_Kingdom.db	United Kingdom	ESRM20, OSM, GMLBF, GHSL

GGY.Guernsey.db	Guernsey	OSM, GMLBF, GHSL
GIB.Gibraltar.db	Gibraltar	ESRM20, OSM, GMLBF, GHSL
GRC.Greece.db	Greece	ESRM20, OSM, GMLBF, GHSL
HRV.Croatia.db	Croatia	ESRM20, OSM, GMLBF, GHSL
HUN.Hungary.db	Hungary	ESRM20, OSM, GMLBF, GHSL
IMN.Isle_of_Man.db	Isle of Man	ESRM20, OSM, GMLBF, GHSL
IRL.Ireland.db	Ireland	ESRM20, OSM, GMLBF, GHSL
ISL.Iceland.db	Iceland	ESRM20, OSM, GHSL
ITA.Italy.db	Italy	ESRM20, OSM, GMLBF, GHSL
JEY.Jersey.db	Jersey	OSM, GMLBF, GHSL
LIE.Liechtenstein.db	Liechtenstein	ESRM20, OSM, GHSL
LTU.Lithuania.db	Lithuania	ESRM20, OSM, GMLBF, GHSL
LUX.Luxembourg.db	Luxembourg	ESRM20, OSM, GHSL
LVA.Latvia.db	Latvia	ESRM20, OSM, GMLBF, GHSL
MCO.Monaco.db	Monaco	ESRM20, OSM, GMLBF, GHSL
MDA.Moldova.db	Moldova	ESRM20, OSM, GMLBF, GHSL
MKD.North_Macedonia.db	North Macedonia	ESRM20, OSM, GMLBF, GHSL
MLT.Malta.db	Malta	ESRM20, OSM, GMLBF, GHSL
MNE.Montenegro.db	Montenegro	ESRM20, OSM, GMLBF, GHSL
NLD.The_Netherlands.db	The Netherlands	ESRM20, OSM, GHSL
NOR.Norway.db	Norway	ESRM20, OSM, GMLBF, GHSL
POL.Poland.db	Poland	ESRM20, OSM, GHSL
PRT.Portugal.db	Portugal	ESRM20, OSM, GMLBF, GHSL
ROU.Romania.db	Romania	ESRM20, OSM, GMLBF, GHSL
SJM.Svalbard_and_Jan_Mayen.db	Svalbard & Jan Mayen	OSM, GHSL
SMR_San_Marino.db	San Marino	OSM, GHSL
SRB_Serbia.db	Serbia	ESRM20, OSM, GMLBF, GHSL
SVK_Slovakia.db	Slovakia	ESRM20, OSM, GHSL
SVN_Slovenia.db	Slovenia	ESRM20, OSM, GMLBF, GHSL
SWE_Sweden.db	Sweden	ESRM20, OSM, GMLBF, GHSL
TCY_Northern_Cyprus.db	Northern Cyprus	ESRM20, OSM, GHSL
TUR_Turkey.db	Turkey	ESRM20, OSM, GMLBF, GHSL
VAT_Vatican_City.db	Vatican City	OSM, GHSL
XKO_Kosovo.db	Kosovo	ESRM20, OSM, GMLBF, GHSL

4.1.Description of data tables

The data are provided as individual files (.db) for each country. Each file has the following components:

4.1.1. Entity table

The Entity features denote the location of the physical exposure objects. The table has four columns: an Entity identifier (id), the geometry (geom), the ISO 3166-1 alpha-3 code (country_iso_code) and the Quadkey (quadkey). While the original exposure comes from individual buildings, the data in this publication is aggregated to the Quadkey tiles to preserve privacy-sensitive information about individual buildings. The highest resolution of the Quadkeys is level-18 and the lowest is level-15, depending on how many buildings are inside the tile. The exposure attributes are found in the Asset table (see section 4.1.3) and are linked through the Entity id.

Column header	unit	Description
id		Entity Identifier
geom	POLYGON	Geometry of tile entity in WGS84
country_iso_code	Char(3)	
quadkey	Char(15-18)	A 15 to 18 characters long string identifying the Quadkey of the tile

4.1.2. Taxonomy table

The Taxonomy features define all structural types existing in the database. All features have an identifier (id) and a taxonomy description (taxonomy_string). The taxonomy is according to the GEM Taxonomy v.3, that is described on https://github.com/gem/gem_taxonomy. While there is no formal description of this version, there is one of the second version (Brzev et al., 2013). More information can also be found on https://taxonomy.openquake.org/.

Column header	Description
id	Taxonomy Identifier
taxonomy_string	Taxonomy description according to GEM Taxonomy v.3

4.1.3. Asset table

The Asset features link the entities with the structural definition of the Taxonomy table. Each feature contains a link to the entity via the identifier (entity_id) and one to a taxonomy (taxonomy_id). The combination between the two is unique. The features also have semantic values: the number of buildings (number) defines the amount of buildings of the taxonomy description can be found in the entity. Because the distribution of building types over the entities is probabilistic, this value can be less than zero. The sum of all number attribute values over all assets inside one entity results in the total number of buildings inside the tile entity. The structural costs of the assets (structural) and the number of people inside the asset at night (night) are defined too. Again, to find the total structural cost or total population in one tile, all the assets in the tile need to be summed up together.

Column header	unit	Description
entity_id		Entity Identifier
taxonomy_id		Taxonomy Identifier
number		Number of buildings
structural	Euro	Structural costs of the asset
night	Persons	Number of people inside the asset at night

4.1.4. Key values per tile view

In the key_values_per_tile view, the number of buildings, structural costs and number of people at night over all assets are summed up in each tile.

Column header	unit	Description
idx		Entity Identifier
geometry	POLYGON	Geometry of tile entity in WGS84
quadkey	Char(15-18)	A 15 to 18 characters long string identifying the Quadkey of the tile
number_buildings		Total number of buildings inside the entity
sum_structural	Euro	Total structural costs of the entity
number_night	Persons	Number of people inside the entity at night

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

Crowley, H., Despotaki, V., Rodrigues, D., Silva, V., Costa, C., Toma-Danila, D., Riga, E., Karatzetzou, A., Fotopoulou, S., Sousa, L., Ozcebe, S., Gamba, P., Dabbeek, J., Romão, X., Pereira, N., Castro, J.M, Daniell, J., Veliu, E., Bilgin, H., … Hancilar, U.,. (2021). European Exposure Model Data Repository (v1.0) [dataset]. Zenodo. https://doi.org/10.5281/ZENODO.4062044

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 2023, 00:00 UTC]. Retrieved from https://planet.openstreetmap.org

Pesaresi, M. (2022). GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030) [dataset]. European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/D07D81B4-7680-4D28-B896-583745C27085