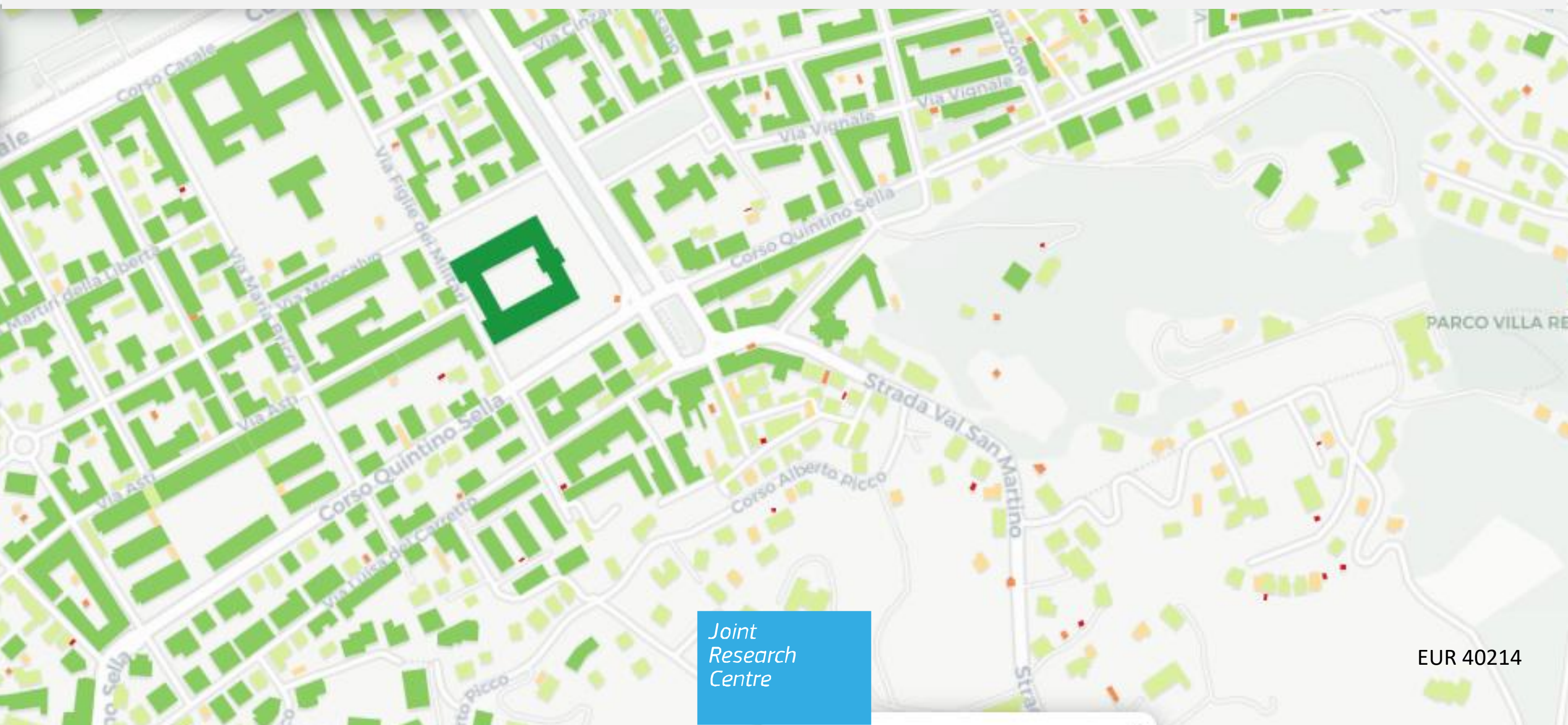
ISSN 1831-9424

European Commission

GHS-OBAT: Global Human Settlement - Open Buildings Attribute Table

Florio, P., Politis, P., Krasnodebska, K., Uhl, J.H., Melchiorri, M., Martinez, A., Kakoulaki, G., Pesaresi, M., Kemper, T.

2025



Joint Research Centre

EUR 40214

Contact information
Name: Thomas Kemper
Address: via Enrico Fermi 2749, 21027 Ispra (VA), Italy
Email: JRC-GHSL-DATA@ec.europa.eu
Tel.: +39 0332 78-5576

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC140737

EUR 40214

PDF ISBN 978-92-68-24173-8 ISSN 1831-9424 doi:10.2760/4854352 KJ-01-25-062-EN-N

Luxembourg: Publications Office of the European Union, 2025

How to cite this report: European Commission: Joint Research Centre, Florio, P., Politis, P., Krasnodebska, K., Uhl, J.H., Melchiorri, M. et al., *GHS-OBAT: Global Human Settlement - Open Buildings Attribute Table,* Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/4854352, JRC140737

Contents

Abstract

The Global Human Settlement Layer (GHSL) project produces new global spatial information, evidence-based analytics and knowledge describing the human presence on Earth. It operates in a fully open and free data and methods access policy. The knowledge generated with the GHSL is supporting the definition, the public discussion and the implementation of European policies and the monitoring of international frameworks such as the 2030 Development Agenda. GHSL data continue to support the GEO Human Planet Initiative that is committed to developing a new generation of measurements and information products providing new scientific evidence and a comprehensive understanding of the human presence on the planet and that can support global policy processes with agreed, actionable and goal-driven metrics.

This document presents the Global Human Settlement – Open Building Attribute Table (GHS-OBAT). The dataset is based on GHSL Data Package 2023, which constitute its source information, here brought to the level of individual building footprints. Easily accessible building footprint vector datasets with global coverage are on the rise, yet they primarily consist of footprint polygons with minimal thematic details. In this context, we describe a comprehensive data integration initiative aimed at associating thematic attributes with buildings, derived from the dataset of 2.3 billion building geometries that are freely available (Overture maps). Specifically, we gather data regarding the building construction epoch, building height, building function (i.e., residential/non-residential), and building compactness (expressed as shape factor) from open-source datasets included in the Global Human Settlement Layer data suite. The attributes at the footprint level were acquired through vector-raster integration, specifically utilizing zonal statistics at the building feature level. The comprehensive global coverage of this dataset offers significant insights across various fields, including disaster risk management, urbanisation, energy transition and climate mitigation, affordable housing and renovation, urban design, real estate, and much more.

Authors

Pietro Florio, JRC E.1
Panagiotis Politis, European Dynamics Belgium S.A.
Katarzyna Krasnodębska, Polish Academy of Sciences
Johannes H. Uhl, JRC E.1
Michele Melchiorri, JRC E.1
Ana M. Martinez, JRC C.2
Georgia Kakoulaki, JRC C.2
Martino Pesaresi, JRC E.1
Thomas Kemper, JRC E.1 (team and project leader)

Introduction

1.1 Overview

The Global Human Settlement Layer (GHSL) project produces global spatial information, evidence-based analytics, and knowledge describing the human presence on the planet. The GHSL relies on the design and implementation of spatial data processing technologies that allow automatic data analytics and information extraction from large amounts of heterogeneous geospatial data including global, fine-scale satellite image data streams, census data, and crowd sourced or volunteered geographic information sources.

GHS-OBAT (Global Human Settlement – Open Buildings Attribute Table) leverages on this interlinkage between open built-up information produced by GHSL at global level based on remote sensing, with other global initiatives gathering efforts from private and public institutions vowed to disclosing building data to the public. Specifically, GHS-OBAT combines open source geometry from the Overture project with GHSL raster grids issued from its latest release, overlays them spatially and generates a list of attributes allocated to each individual building footprint.

GHS-OBAT (Global Human Settlement – Open Buildings Attribute Table) includes the following dataset distributions:

- Attribute tables per country / administrative unit, with latitude and longitude of building footprint centroids: Comma Separated Value files (CSV)
- Attribute tables per country / administrative unit, with georeferenced building footprint centroids: GeoPackage files (GPKG)
- Country and sub-country level statistics: MS Excel files (XLSX)
- Open Python code on GitHub: Python files (PY)

1.2 Background

Architects, urban planners, city designers and professionals in the field frequently utilize building footprints, in vector polygonal format. The popularity of such data format is steadily increasing among policymakers and analysts, attributed to their detailed spatial representation and their structure, adapted to object-based attribute storage and querying, which is well suited for multi-thematic analyses. Whereas most cadastral data report building footprints in a strict sense (building intersection with the ground), remotely sensed data usually collect building “roofprints” (roof projection to the ground): in this work, we refer generically to building footprints as polygons representing the shape of the building on a planar view to the ground, without distinction.

Globally available building footprint vector datasets are on the rise, yet they often consist solely of vector geometries accompanied by limited or incomplete thematic attributes (e.g., building height when accessible). Other open data initiatives, such as the Global Human Settlement Layer [1], [2], generate consistent, multi-temporal gridded datasets with global coverage, which are valuable for extensive analyses and visualization objectives, but they fall short in providing detailed spatial information at the level of individual buildings.

Consequently, the combination of detailed data on building footprints with thematically rich information on human settlements sourced from remote-sensing initiatives significantly increases the value of building footprint data and paves the way for new research opportunities in building-stock analysis, particularly in transdisciplinary studies encompassing various thematic areas.

Input data sources

GHS-OBAT is a dataset that consists of a table, with building footprint features as rows and attributes as columns, split by country and administrative areas available in the Global ADMinistrative layer version 4.1 (GADM4.1). The attributes include unique identifiers to link to Overture buildings 2024-07-22.0, their centroid coordinates, country and administrative unit location, and building characteristics such as height, shape factor (compactness), functional use, construction year, area, and perimeter. These characteristics are derived from the Global Human Settlement Layer (GHSL) global datasets GHS R2023 and R2024. The dataset is available both as CSV and as SQLite-based geopackages, which incorporate the centroids of building footprints as geometry features.

Table 1. Attributes of the GHS-OBAT and their sources. See section 0 for further details on each attribute.

Attribute	**Source**	**Release / Version**
Geometry (in gpkg)	Overture buildings (geoprocessed)	2024-07-22.0
Id	Overture buildings	2024-07-22.0
Lon	Overture buildings (geoprocessed)	2024-07-22.0
Lat	Overture buildings (geoprocessed)	2024-07-22.0
Country	GADM	4.1 (modified)
Adm1	GADM	4.1 (modified)
Height	GHS-BUILT-H	R2023 V1.0
Shapefactor	GHS-BUILT-H, Overture buildings (geoprocessed)	R2023 V1.0, 2024-07-22.0
Use	GHS-BUILT-C	R2023 V1.0
Epoch	GHS-AGE	R2024 V1.0
Area	Overture buildings (geoprocessed)	2024-07-22.0
Perimeter	Overture buildings (geoprocessed)	2024-07-22.0

Source: JRC analysis.

2.1 GHSL: GHS-BUILT

Most attributes computed at building level rely on several components of the GHS-BUILT (Built-Up) layer of the Global Human Settlement Layer (GHSL)[1]. In particular, attributes rely on the products described herewith. For detailed descriptions on such data, see their distribution report [3] or the dedicated scientific literature [2].

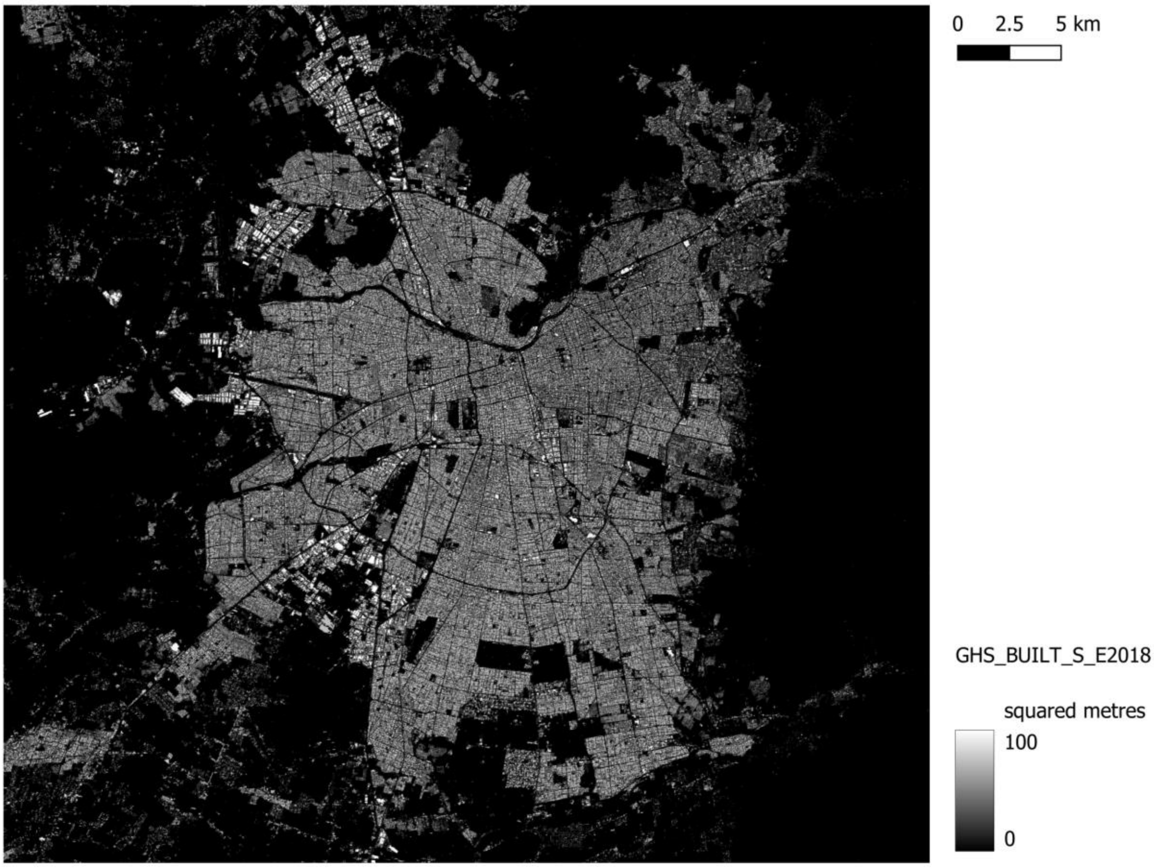
[1] https://human-settlement.emergency.copernicus.eu/

2.1.1 GHS-BUILT-S

The GHS-BUILT-S spatial raster dataset [4] depicts the distribution of the built-up (BU) surfaces estimates between 1975 and 2030 in 5 year intervals (**Figure 1**). The data is made by spatial-temporal interpolation of five observed collections of multiple-sensor, multiple-platform satellite imageries: Landsat (MSS, TM, ETM sensor) data supports the 1975, 1990, 2000, and 2014 epochs, while a Sentinel-2 (S2) image composite (GHS-composite-S2 R2020A) supports the 2018 epoch.

The built-up surface is the gross surface (including the thickness of the walls) bounded by the building wall perimeter with a spatial generalization matching the 1:10K topographic map specifications, that it also informally called “building footprint”.

Figure 1. Santiago de Chile. GHS-BUILT-S_GLOBE_R2023A for the epoch 2018 from Sentinel-2 image data with a continuous 10m-resolution method.



Source: JRC analysis.

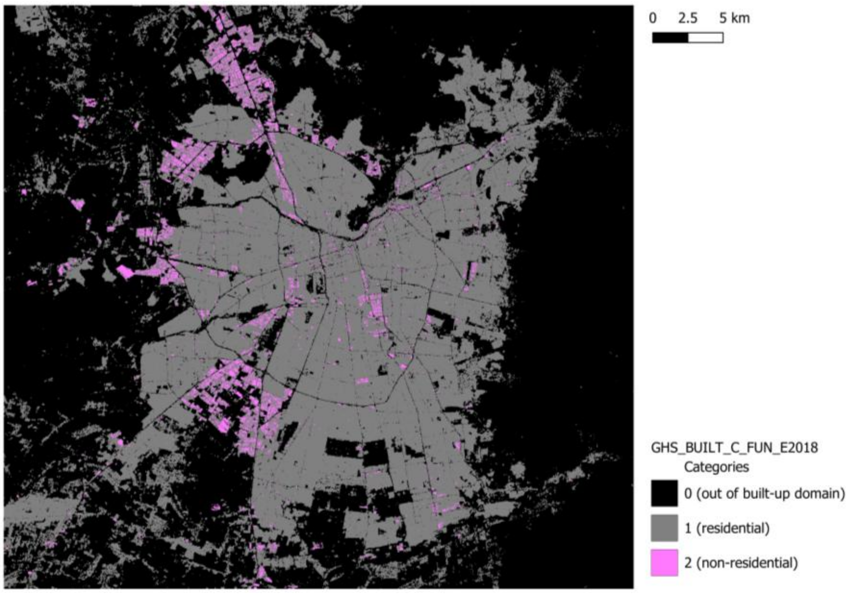
2.1.2 GHS-BUILT-C (FUN)

The GHS-BUILT-C spatial raster datasets [5] delineate the boundaries of the human settlements at 10m resolution (**Figure 2**), and describe their inner characteristics in terms of the morphology of the built environment and the functional use. The FUN module classifies the built-up domain into residential and non-residential.

The definition of built-up surface adopted in GHSL R2023 is in line with the INSPIRE designation[2], and encompasses aboveground buildings and constructions, including those with temporary use, for sheltering humans or goods under any roofed structure. Based on remote sensing, this definition of built-up necessarily excludes underground structures, while it includes temporary settlements as associated to slums, rapid migratory patterns, or displaced people because of natural disasters or crises. Similarly, the residential use of built-up areas is defined as dominantly for housing of people, including a mix with other non-conflicting uses (e.g. in mixed-use buildings, with offices or shops occupying part of the floor space.

In compliance with such definition, the residential label in GHSL R2023 is not excluding other non-residential uses. On the other hand, the non-residential label indicates industrial or commercial facilities, warehouses, and infrastructural nodes etc., which are commonly considered not suitable for residential use. In the perspective of territorial development, non-residential built-up constitutes a proxy for economic wealth, unveiling commercial investments and capital flows to that specific site.

Figure 2. Santiago de Chile. GHS-BUILT-C_GLOBE_R2023A (FUN) for the epoch 2018.



Source: JRC analysis.

2.1.3 GHS-BUILT-H

This GHS-BUILT-H spatial raster dataset [6] depicts the distribution of the building heights generalized at the resolution of 100m, and referred to the year 2018 (**Figure 3**). The input data used to predict the building heights are the ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)[3], the NASA Shuttle Radar Topographic Mission data - 30m (SRTM30)[4] , and the Sentinel-2

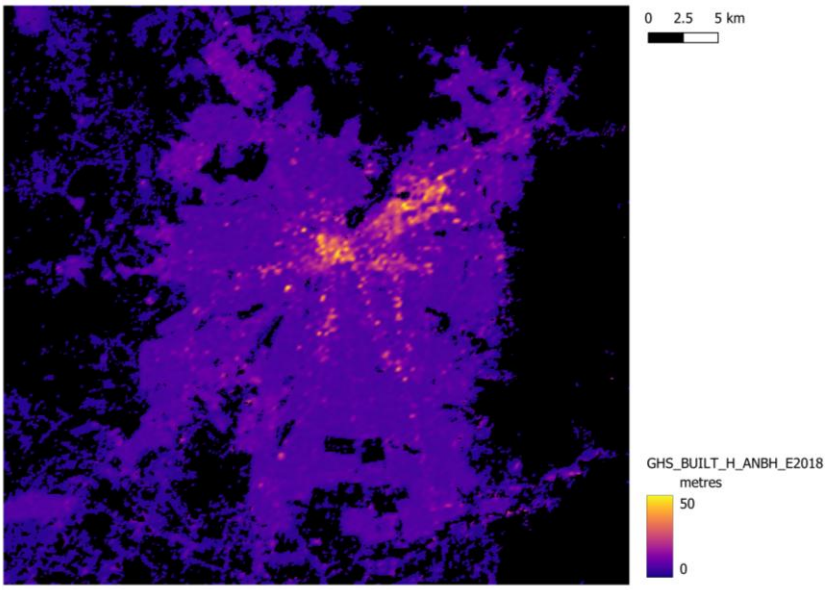
[2] https://inspire.ec.europa.eu/id/document/tg/bu

[3] https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm

[4] https://cmr.earthdata.nasa.gov/search/concepts/C1000000240-LPDAAC_ECS.html

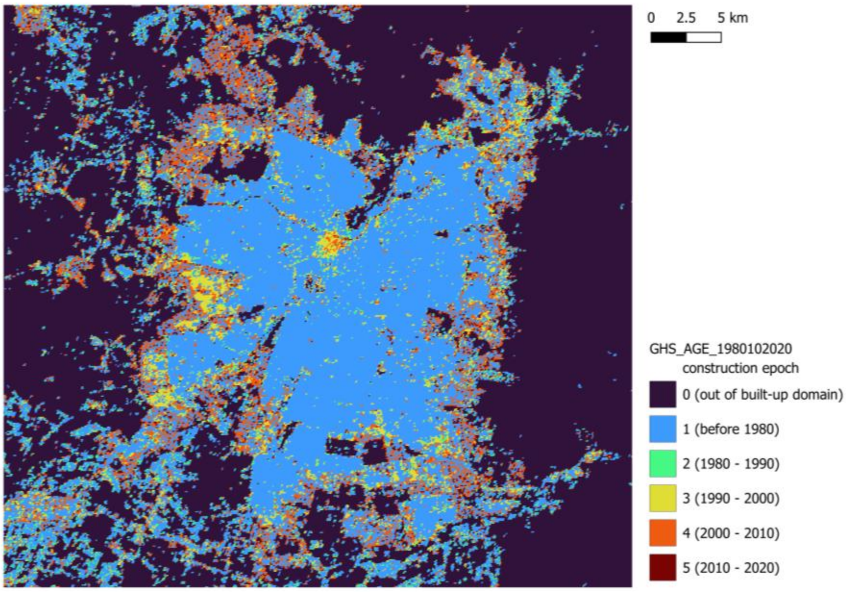
global pixel based image composite from L1C data for the period 2017-2018[5] that is the support of the year 2018 in the GHS-BUILT-S R2023A dataset.

Figure 3. Average building height (ANBH 100m) estimates in Santiago de Chile.



Source: JRC analysis.

Figure 4. GHS-AGE building construction epoch in decadal range between 1980 and 2020, mapped in Santiago de Chile.



Source: JRC analysis.

[5] GHS-composite-S2 R2020A https://ghsl.jrc.ec.europa.eu/download.php?ds=compositeS2

2.1.4 GHS-AGE

GHS-AGE [7] maps the earliest epoch, with 5 or 10 years' time step in the range 1980-2020, in which at least 50% of the contemporary built-up surface existed (**Figure 4**). This epoch constitutes the construction year of the majority of built-up surface in a given 100m resolution grid cell, and our best estimate for the construction year of the buildings predominantly located within that grid cell. The epochs limits are designed based on the availability of satellite imagery.

2.2 Overture

The Overture Maps Foundation[6] is an initiative that brings together a consortium of esteemed partners, including Amazon, Meta, Microsoft, and TomTom, to create a comprehensive, open-source mapping platform. The primary objective of this collaborative effort is to develop a unified, accurate, and continuously updated set of map data, which can be utilized by various industries, organizations, and individuals to drive innovation and improve decision-making.

At its core, the Overture Maps Foundation aims to provide a robust and scalable mapping ecosystem, comprising several key components, including:

- **Road Networks**: A detailed and up-to-date representation of the world's roads, highways, and other transportation infrastructure.
- **Points of Interest**: A vast database of notable locations, such as businesses, landmarks, and public services.
- **Buildings**: A comprehensive catalogue of structures, including residential, commercial, and industrial buildings, which is a primary focus of the project. This component is designed to provide detailed information about building footprints, heights, and other relevant attributes, enabling a wide range of applications, from urban planning and disaster response to logistics and navigation.
- **Geographical Features**: A detailed representation of natural and man-made geographical features, such as rivers, lakes, and parks.

The Overture Maps Foundation has made significant strides in achieving its objectives, with several notable achievements, including:

- **The creation of a unified data model**, which enables seamless integration and exchange of map data across different platforms and applications.
- **The development of a robust data pipeline**, which ensures the continuous updating and validation of map data.
- **The establishment of a collaborative community**, which brings together experts from various fields to contribute to the project and provide feedback.

The Overture Maps Foundation aims to unlock new possibilities for urban planning, emergency response, and other applications that rely on accurate and up-to-date information about structures.

---

[6] https://overturemaps.org/

As the project continues to evolve, it is expected to have a profound impact on various industries and aspects of society, from logistics and transportation to public safety and environmental sustainability.

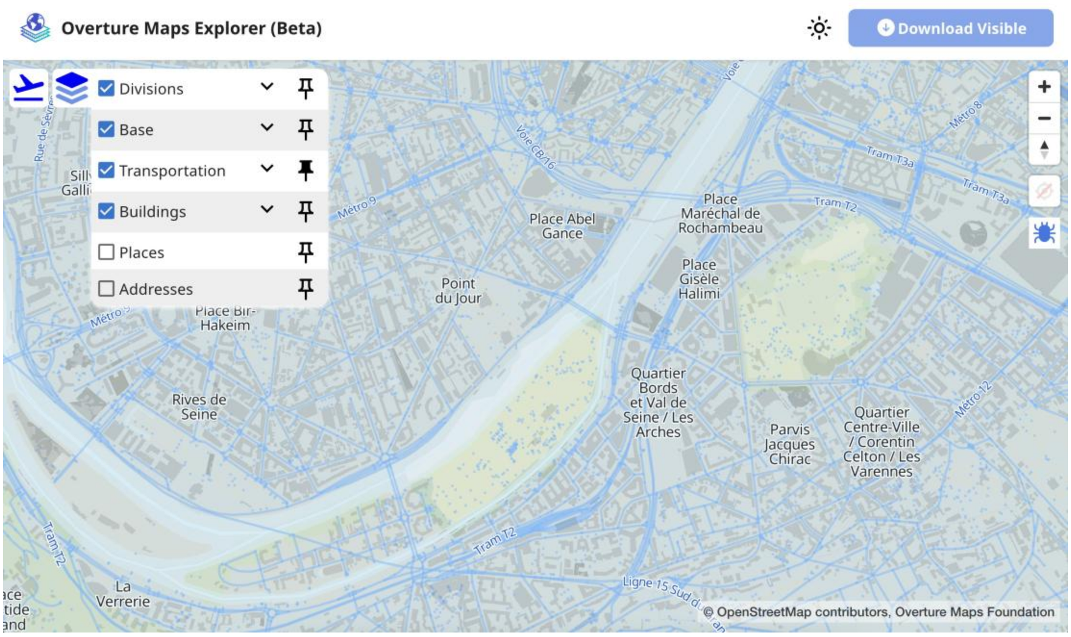
In the framework of GHS-OBAT, the **Overture Buildings** [8] were used. Overture buildings is a collection of conflated building features from different open datasets as shown in **Table 2**.

Table 2. Overture buildings components and their conflation priority.

Source	Type	Conflation Priority	Count
OpenStreetMap (OSM)	Community-contributed	1	~620 Million
Esri Community Maps	Community-contributed	2	~15 Million
Google Open Buildings	ML-derived roofprints (>90% precision)	3	~400 Million
Microsoft	ML-derived roofprints	4	~736 Million
Google Open Buildings	ML-derived roofprints (<90% precision)	5	~700 Million

Source: Overture Maps Foundation[7].

Figure 5. A snapshot of the Overture Maps Explorer[8], which highlights the several components available in the dataset.



Source: Overture Maps Foundation © [8].

[7] Table issued from https://docs.overturemaps.org/guides/buildings/

[8] https://explore.overturemaps.org/

2.3 GADM

The GADM 4.1 dataset [9] is a comprehensive geographic database of global administrative areas, providing a standardized and harmonized representation of administrative boundaries worldwide. Developed by the Database of Global Administrative Areas (GADM), this dataset is widely used in various fields, including geography, demography, economics, and environmental studies.

The GADM 4.1 dataset contains administrative boundaries for 342 countries and territories, with a total of 3,835,249 polygons. The dataset is available in most common geospatial formats (shapefile, geoJSON, geopackage, CSV) and is organized into a hierarchical structure, with six levels of administrative divisions:

1. Country
2. First-level administrative division (e.g., state, province)
3. Second-level administrative division (e.g., county, municipality)
4. Third-level administrative division (e.g., city, town)
5. Fourth-level administrative division (e.g., village, neighbourhood)
6. Fifth-level administrative division (e.g., local council, ward)

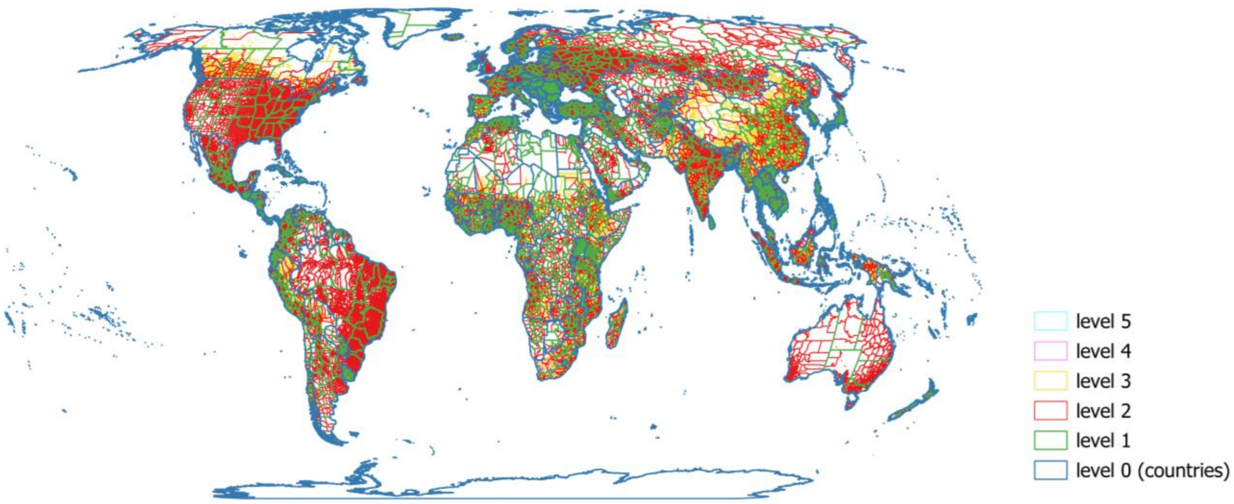
In terms of attributes, the GADM 4.1 dataset includes the following features:

- Administrative boundaries: Polygons representing administrative areas at various levels.
- Geographic coordinates: Latitude and longitude values for each polygon's centroid.
- Names and codes: Standardized names and codes for each administrative area.
- Hierarchy: Relationships between administrative areas at different levels.
- Population data: Estimated population figures for each administrative area (optional).

The GADM 4.1 dataset has undergone rigorous quality control and validation procedures to ensure accuracy and consistency, including checking of the data sources, geometric validation, and attribute validation.

In GHS-OBAT, level 0 (countries) and level 1 territorial units are used to split the attribute table and to calculate summary statistics (refer to **Table 3**).

Figure 6. Map view of the GADM 4.1 dataset.



Source: JRC analysis.

Table 3. GADM entities with their names and the territorial unit available at level 1.

GADM ISO	GADM NAME	Level 1
ABW	Aruba	-
AFG	Afghanistan	Province
AGO	Angola	Province
AIA	Anguilla	District
ALA	Aland	Sub-Region
ALB	Albania	County
AND	Andorra	Parish
ARE	United Arab Emirates	Emirate
ARG	Argentina	Province
ARM	Armenia	Province
ASM	American Samoa	District
ATA	Antarctica	-
ATF	French Southern Territories	District
ATG	Antigua and Barbuda	Dependency
AUS	Australia	Territory
AUT	Austria	State
AZE	Azerbaijan	Region
BDI	Burundi	Province
BEL	Belgium	Region
BEN	Benin	Department
BES	Bonaire, Sint Eustatius and Saba	Municipality
BFA	Burkina Faso	Region
BGD	Bangladesh	Division
BGR	Bulgaria	Province
BHR	Bahrain	Governorate
BHS	Bahamas	District
BIH	Bosnia and Herzegovina	District
BLM	Saint-Barthelemy	Parish
BLR	Belarus	Region
BLZ	Belize	District
BMU	Bermuda	Parish
BOL	Bolivia	Department
BRA	Brazil	State
BRB	Barbados	Parish
BRN	Brunei	District
BTN	Bhutan	District
BVT	Bouvet Island	-
BWA	Botswana	District
CAF	Central African Republic	Prefecture
CAN	Canada	Province
CCK	Cocos Islands	-
CHE	Switzerland	Canton
CHL	Chile	Region
CHN	China	Province
CIV	Côte d'Ivoire	Autonomous district
CMR	Cameroon	Region
COD	Democratic Republic of the Congo	Province
COG	Republic of the Congo	Region
COK	Cook Islands	Island Council
COL	Colombia	Commissiary
COM	Comoros	Autonomous Island
CPV	Cabo Verde	County
CRI	Costa Rica	Province
CUB	Cuba	Province
CUW	Curaçao	-
CXR	Christmas Island	-
CYM	Cayman Islands	District
CYP	Cyprus	District
CZE	Czechia	Region
DEU	Germany	State
DJI	Djibouti	Region
DMA	Dominica	Parish
DNK	Denmark	Region
DOM	Dominican Republic	Province
DZA	Algeria	Province
ECU	Ecuador	Province
EGY	Egypt	Governorate
ERI	Eritrea	Region
ESH	Western Sahara	Province
ESP	Spain	Autonomous Community
EST	Estonia	County
ETH	Ethiopia	City
FIN	Finland	Province
FJI	Fiji	Division
FLK	Falkland Islands	-
FRA	France	Region
FRO	Faroe Islands	Region
FSM	Micronesia	State
GAB	Gabon	Province
GBR	United Kingdom	Constituent Country
GEO	Georgia	Autonomous Republic
GGY	Guernsey	Parish
GHA	Ghana	Region
GIB	Gibraltar	-
GIN	Guinea	Region
GLP	Guadeloupe	District
GMB	Gambia	Independent City
GNB	Guinea-Bissau	Region
GNQ	Equatorial Guinea	Province
GRC	Greece	Decentralized administration
GRD	Grenada	Dependency
GRL	Greenland	Commune
GTM	Guatemala	Department
GUF	French Guiana	Arrondissement
GUM	Guam	Municipality

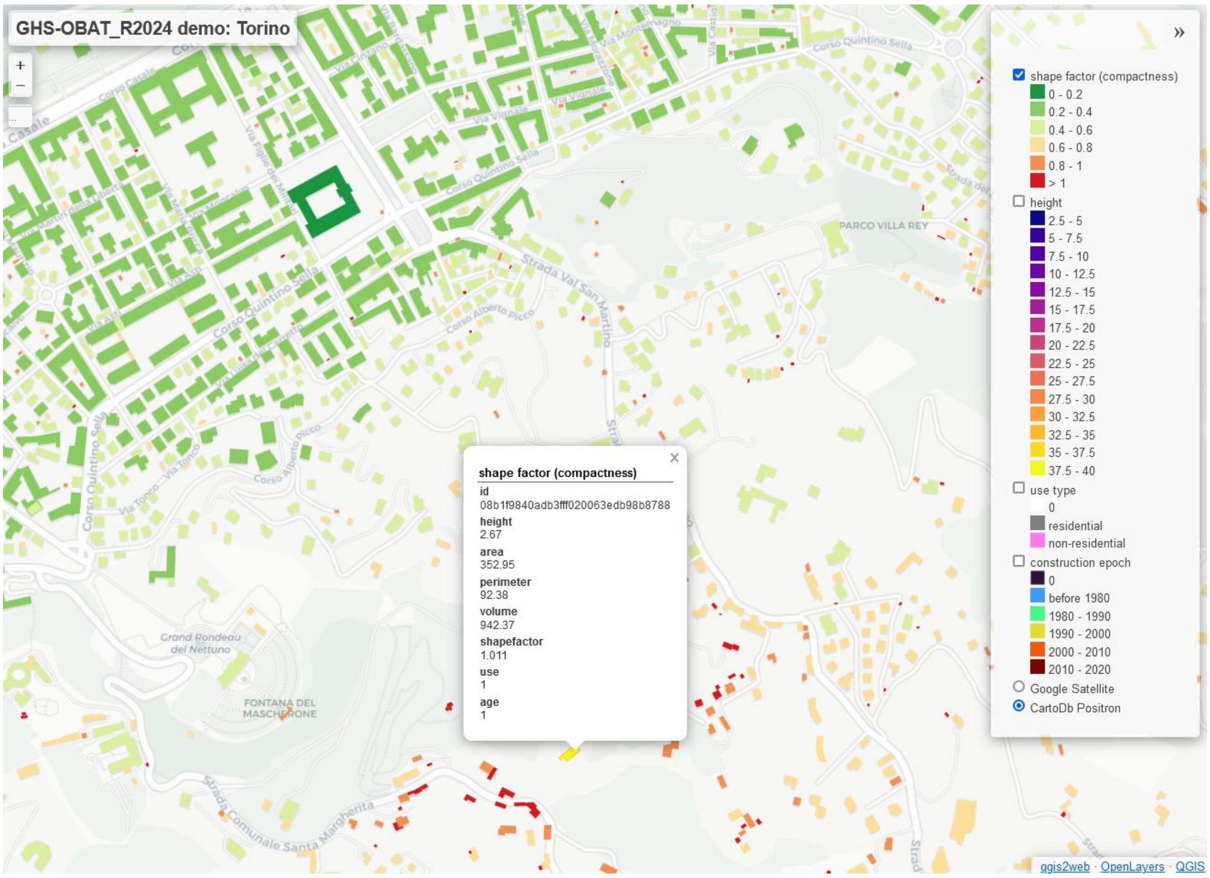
GUY	Guyana	Region
HMD	Heard Island and McDonald Island	-
HND	Honduras	Department
HRV	Croatia	County
HTI	Haiti	Department
HUN	Hungary	County
IDN	Indonesia	Province
IMN	Isle of Man	Parish District
IND	India	Union Territory
IOT	British Indian Ocean Territory	-
IRL	Ireland	County
IRN	Iran	Province
IRQ	Iraq	Province
ISL	Iceland	Region
ISR	Israel	District
ITA	Italy	Region
JAM	Jamaica	Parish
JEY	Jersey	Parish
JOR	Jordan	Province
JPN	Japan	Prefecture
KAZ	Kazakhstan	Region
KEN	Kenya	County
KGZ	Kyrgyzstan	Province
KHM	Cambodia	Province
KIR	Kiribati	-
KNA	Saint Kitts and Nevis	Parish
KOR	South Korea	Metropolitan C
KWT	Kuwait	Province
LAO	Laos	Province
LBN	Lebanon	Governorate
LBR	Liberia	County
LBY	Libya	District
LCA	Saint Lucia	Quarter
LIE	Liechtenstein	Commune
LKA	Sri Lanka	District
LSO	Lesotho	District
LTU	Lithuania	County
LUX	Luxembourg	District
LVA	Latvia	Province
MAF	Saint-Martin	-
MAR	Morocco	Region
MCO	Monaco	-
MDA	Moldova	District
MDG	Madagascar	NA
MDV	Maldives	-
MEX	Mexico	State
MHL	Marshall Islands	Atol
MKD	North Macedonia	Municipality
MLI	Mali	District
MLT	Malta	Region
MMR	Myanmar	Division
MNE	Montenegro	Municipality
MNG	Mongolia	Province
MNP	Northern Mariana Islands	Municipality
MOZ	Mozambique	Province
MRT	Mauritania	Region
MSR	Montserrat	Parish
MTQ	Martinique	Arrondissemen
MUS	Mauritius	Region
MWI	Malawi	District
MYS	Malaysia	State
MYT	Mayotte	Commune
NAM	Namibia	Region
NCL	New Caledonia	Province
NER	Niger	Department
NFK	Norfolk Island	-
NGA	Nigeria	State
NIC	Nicaragua	Autonomous Region
NIU	Niue	-
NLD	Netherlands	Province
NOR	Norway	County
NPL	Nepal	Development Region
NRU	Nauru	District
NZL	New Zealand	Region
OMN	Oman	Region
PAK	Pakistan	Province
PAN	Panama	Province
PCN	Pitcairn Islands	-
PER	Peru	Region
PHL	Philippines	Province
PLW	Palau	State
PNG	Papua New Guinea	Autonomous Region
POL	Poland	Voivodeship
PRI	Puerto Rico	Municipality
PRK	North Korea	Province
PRT	Portugal	District
PRY	Paraguay	Department
PSE	Palestine	District
PYF	French Polynesia	Administrative subdivisions
QAT	Qatar	Municipality
REU	Reunion	Arrondissemen
ROU	Romania	County
RUS	Russia	Republic
RWA	Rwanda	Province
SAU	Saudi Arabia	Province
SDN	Sudan	State
SEN	Senegal	Region
SGP	Singapore	Region
SGS	South Georgia and the South Sand	-
SHN	Saint Helena, Ascension and Tris	Administrative Area
SJM	Svalbard and Jan Mayen	Territory
SLB	Solomon Islands	Province
SLE	Sierra Leone	Province
SLV	El Salvador	Department
SMR	San Marino	Municipality
SOM	Somalia	Region
SPM	Saint Pierre and Miquelon	Commune
SRB	Serbia	District
SSD	South Sudan	State
STP	Sao Tome and Principe	Municipality
SUR	Suriname	District
SVK	Slovakia	Region
SVN	Slovenia	Statistical Region
SWE	Sweden	County
SWZ	Swaziland	District
SXM	Sint Maarten	-
SYC	Seychelles	District
SYR	Syria	Governorate
TCA	Turks and Caicos Islands	District
TCD	Chad	Region
TGO	Togo	Region
THA	Thailand	Province
TJK	Tajikistan	Districts of Republican Subordin
TKL	Tokelau	Atoll
TKM	Turkmenistan	Province
TLS	Timor-Leste	District
TON	Tonga	Island Group
TTO	Trinidad and Tobago	Borough
TUN	Tunisia	Governorate
TUR	Turkey	Province
TUV	Tuvalu	Town Council
TWN	Taiwan	Province
TZA	Tanzania	Region
UGA	Uganda	District
UKR	Ukraine	?
UMI	United States Minor Outlying Isl	Island
URY	Uruguay	Department
USA	United States	State
UZB	Uzbekistan	Region
VAT	Vatican City	-
VCT	Saint Vincent and the Grenadines	Parish
VEN	Venezuela	State
VGB	British Virgin Islands	District
VIR	Virgin Islands, U.S.	District
VNM	Vietnam	Province
VUT	Vanuatu	Province
WLF	Wallis and Futuna	Kingdom
WSM	Samoa	District
XAD	Akrotiri and Dhekelia	Sovereign Base Area
XCA	Caspian Sea	-
XCL	Clipperton Island	-
XKO	Kosovo	District
XPI	Paracel Islands	-
XSP	Spratly Islands	-
YEM	Yemen	Governorate
ZAF	South Africa	Province
ZMB	Zambia	Province
ZNC	Northern Cyprus	District
ZWE	Zimbabwe	City
GADM ISO	GADM NAME	Level 1
ABW	Aruba	-

Source: JRC analysis.

Data description

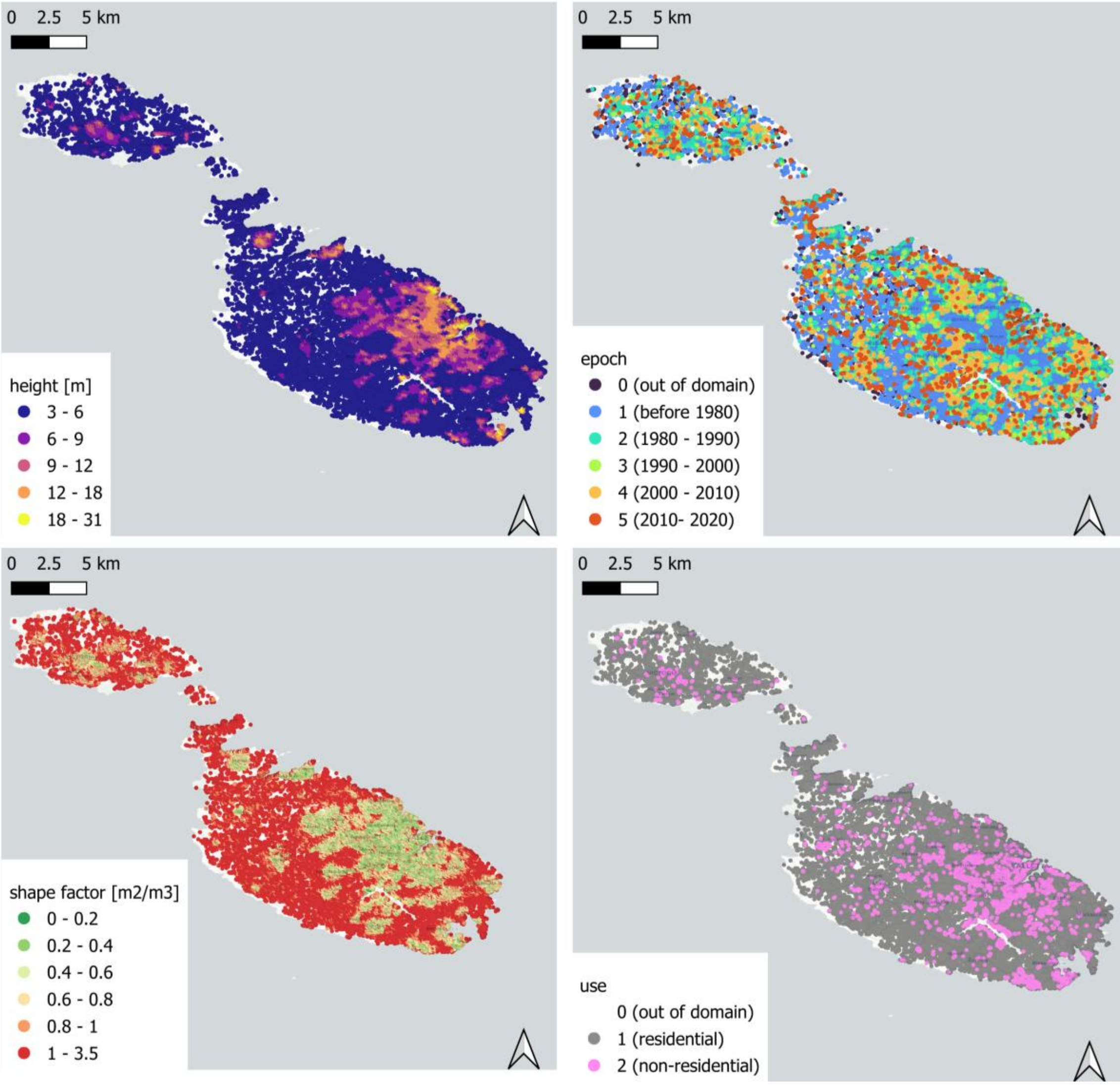
The main component of the dataset consists of an attribute table in CSV format split by country (and by administrative areas level 1 for larger countries), each file with a maximum of circa 50 million building footprint features as rows and attributes as columns. Such table incorporates unique identifiers to link each set of attributes to the correct footprint polygon in the original geometry files (**Figure 7**). The original geometry files can be downloaded from the Overture website [8], following the instructions provided there. Alternatively, the dataset is also released as a set of SQLite-based geopackages that incorporate the centroids of building footprints as geometry features, holding the same attributes of the CSV files, but much handier to use for further geospatial analyses (**Figure 8**).

Figure 7. Example of a possible web-interface for the GHS-OBAT dataset in association with building footprint geometries from Overture, featuring the city of Torino (Italy).



Source: JRC analysis.

Figure 8. Preview of the GHS-OBAT dataset for the country of Malta (MLT) in a GIS environment. The four main attributes are represented, namely height, construction epoch, shape factor and use. In the bottom, the attributes of an example record are shown.



Feature	Value
fid	59022
id	08b3f3042a530fff0200d80ce488be6d
country	MLT
adm1	Tramuntana
height	2.5
shapefactor	1.61169
use	2
epoch	1
area	27.52
perimeter	22.33

Source: JRC analysis.

3.1 Technical details

Product name: GHS-OBAT_GLOBE_R2024A (Global Human Settlement – Open Buildings Attribute Table)

Spatial extent: Global

Temporal extent: 2020

Coordinate system: WGS84 (EPSG:4326)

Spatial resolution available: - (vector data)

Records: attributes for 2,282,514,036 building footprints

Data organisation:

- **GHS-OBAT_GLOBE_R2024A_CSV**: Attribute tables per country / administrative unit, with latitude and longitude of building footprint centroids: Comma Separated Value files (CSV)
- **GHS-OBAT_GLOBE_R2024A_GPKG**: Attribute tables per country / administrative unit, with georeferenced building footprint centroids: GeoPackage files (GPKG)
- **GHS-OBAT_COUNTRYSTATS_GLOBE_R2024A**: Country and sub-country level statistics: MS Excel files (XLSX)
- Open Python code on GitHub: Python files (PY)

3.1.1 GHS-OBAT_GLOBE_R2024A: Dataset attributes

The attribute table includes the following attributes (**Table 4**):

Table 4. Attributes description

Attribute	Data Type	Unit	Encoding	Description
geometry[9]	Geometry	-	Point	centroid of the building footprint
id	string	-	32 HEX string	Unique identifier that constitutes the stable link to the geometry in Overture (the so-called GERS ID[10]). The Global Entity Reference System (GERS) is a framework for structuring, encoding, and matching map data to a shared universal reference. All features in Overture have a unique ID, a mechanism to match features across datasets, track data stability, and detect errors in the data. It is a 128 bit string of 32 hexadecimal characters.
lon[11]	decimal	deg	WGS84 (EPSG:4326)	X-coordinate (longitude) of the centroid of the building footprint in WGS84 projection (EPSG:4326), expressed in decimal degrees.

[9] Only in the Geopackage distribution

[10] https://docs.overturemaps.org/gers/

[11] Only in the CSV distribution

lat[11]	decimal	deg	WGS84 (EPSG:4326)	Y-coordinate (latitude) of the centroid of the building footprint in WGS84 projection (EPSG:4326), expressed in decimal degrees.
country	string	-	ISO 3166-1 alpha-3 three-letter country code	ISO 3166-1 alpha-3 three-letter country code of the country the building footprint belongs to, according to an adjusted version of GADM 4.1 [9], made to fill gaps and overlaps and fix most unassigned territories. Building footprints not falling in any GADM level 0 country entity are marked “XXX”.
adm1	string	-	text	text string holding the name of the administrative unit (subnational level 1), according to an adjusted version of GADM 4.1 [9], made to fill gaps and overlaps and fix most unassigned territories. Building footprints not falling in any GADM level 1 entity are marked “XXX”.
height	decimal	m	-	Mean building height calculated from the 100 metres resolution grid cells of the Average of the Net Building Height (ANBH) raster from the GHS-BUILT-H dataset [2], [6, Sec. 4.3.1], using zonal statistics. The ExactExtract library computes the mean height of the grid cells intersecting with each building footprint polygon, excluding any NoData values (e.g. value 0 in the ANBH raster). The height is expressed as floating point number in metres [m], and is adjusted to a lower limit of 2.5 [m]. The ANBH raster is generated by extracting building height information from a composite of global Digital Elevation Models: the optical-based Advanced Land Observing Satellite World 3D [10] and the radar-based Shuttle Radar Topography Mission SRTM30 [11], both in ~30 [m] resolution, and updating the building height estimations to year 2018 using shadow markers derived from Sentinel-2 imagery composite [12].
shapefactor	decimal	m^2/m^3 (1/m)	-	Building shape factor: S/V [13], with S being the building surface (summing the roof, the floor and the façades) and V the volume of the building. The shape-factor can be estimated at feature level, approximating the building solid at LOD1 (simple extrusion). The unit is [m^2/m^3] and it is inversely proportional to the size of the building.
use	integer	-	0-2	Building functional use (residential / non-residential) derived from the mode (majority) of the intersecting 10-meter resolution grid cells, using the original 10-meter resolution raster functional classification of the built domain (FUN) from the GHS-BUILT-C dataset [5]. This classification is further extended by incorporating the 100-meter functional classification from the BUTYPE raster grid, a derived product of the GHSL R2023A dataset [2, Sec. 3.62 and 4.4.2]. Building functional use is expressed as an integer number, with value 0 for footprints outside of the built-up domain (defined as the union of GHS-BUILT-C and BUTYPE), 1 for

				"residential" and 2 for "non-residential" built-up surface. The definition of residential built-up areas is aligned to the INSPIRE definition[12] and includes areas dedicated prevalently for residential use, including mixed with other non-conflicting uses. The non-residential areas are dedicated exclusively to non-residential use, and predicted by observation of radiometric, textural and morphological features derived from Sentinel-2 imagery satellite imagery [2, Sec. 4.2.3].
epoch	integer	-	0-5	The building construction year is derived using the mode (majority rule) of the intersecting 100-meter resolution grid cells from the original raster of building age (1975–2020), a derived product of the GHS-BUILT-S dataset [4], [7]. Specifically, the spatio-temporal data cube of the GHS-BUILT-S dataset, measuring sub-pixel built-up surface at 100m resolution, globally from 1975 to 2020, was queried along the temporal domain, in order to identify the year in which 50% or more of the contemporary built-up surface was present in each grid cell. This year represents the most likely construction epoch of the built stock within each grid cell, in the absence of more detailed building-level construction year information. The estimated building construction epoch estimation (GHS-AGE) ranges between 1980 and 2020, with a time step of 10 years: it is expressed as an integer number ranging from 1 to 5 (respectively before 1980, 1980-1990, 1990-2000, 2000-2010, 2010-2020), with 0 being footprints outside of the built-up domain of GHS-BUILT-S.
area	decimal	m^2	-	Surface of the Overture building footprint expressed as floating point number in metres squared [m^2], computed in the local UTM projection where the building footprint lies.
perimeter	decimal	m	-	Perimeter of the Overture building footprint expressed as floating point number in metres [m], computed in the local UTM projection where the building footprint lies.

Source: JRC analysis.

[12] https://inspire.ec.europa.eu/codelist/HILUCSValue/5_ResidentialUse

3.1.2 GHS-OBAT_COUNTRYSTATS_GLOBE_R2024A: Summary statistics

The summary statistics reported below in **Table 5** are available at country and administrative level 1.

Table 5. Summary statistics available at country and administrative level 1 for the attributes contained in the GHS-OBAT dataset.

Summary field	Suffixes	Unit	Description
Surface_*	* = min, q1, median, average, q3, max, sum	m^2	Building footprints surface aggregation in the country of interest (or administrative unit level 1): respectively minimum, 1st quartile, median, average, 3rd quartile, maximum, sum.
Height_*	* = min, q1, median, average, q3, max	m	Buildings height aggregation in the country of interest (or administrative unit level 1): respectively minimum, 1st quartile, median, average, 3rd quartile, maximum.
Shapefactor_*	* = min, q1, median, average, q3, max	m^2/m^3 (1/m)	Buildings shape factor aggregation in the country of interest (or administrative unit level 1): respectively minimum, 1st quartile, median, average 3rd quartile, maximum.
Footprint_count_Use_*	* = 0, 1, 2	-	Number of building footprints in each of the functional use categories: respectively 1 for residential and 2 for non-residential. 0 is for building footprints outside of the built-up raster domain.
Footprint_surface_ha_Use_*	* = 0, 1, 2	hectares	Building footprints surface sum in each of the functional use categories: respectively 1 for residential and 2 for non-residential. 0 is for building footprints outside of the built-up raster domain.
Footprint_count_Epoch_*	* = 0, 1, 2, 3, 4, 5	-	Number of building footprints in each of the construction epochs: respectively 1 for before 1980, 2 for 1980-1990, 3 for 1990-2000, 4 for 2000-2010, 5 for 2010-2020). 0 is for building footprints out of the raster domain.
Footprint_surface_ha_Epoch_*	* = 0, 1, 2, 3, 4, 5	hectares	Building footprints surface sum in each of the construction epochs: respectively 1 for before 1980, 2 for 1980-1990, 3 for 1990-2000, 4 for 2000-2010, 5 for 2010-2020). 0 is for building footprints outside of the built-up raster domain.

Source: JRC analysis.

3.2 Methodological note

The full methodological details are provided in the GHS-OBAT accompanying scientific paper [14], but the overarching workflow integrates building footprint vectors and raster grids providing thematic information using geospatial operations in Python v3.11 (**Figure 9**).

- The workflow involves pre-processing steps such as splitting the data into a default tiling system, and aligning the data to a common Coordinate Reference System (CRS). Country and sub-country level labels are assigned to building footprints by spatial join with GADM (described in section 2.3).
- The pre-processed data is then integrated using zonal statistics to extract raster statistics (from GHSL layers) within each polygonal feature (from Overture buildings) representing building footprints.

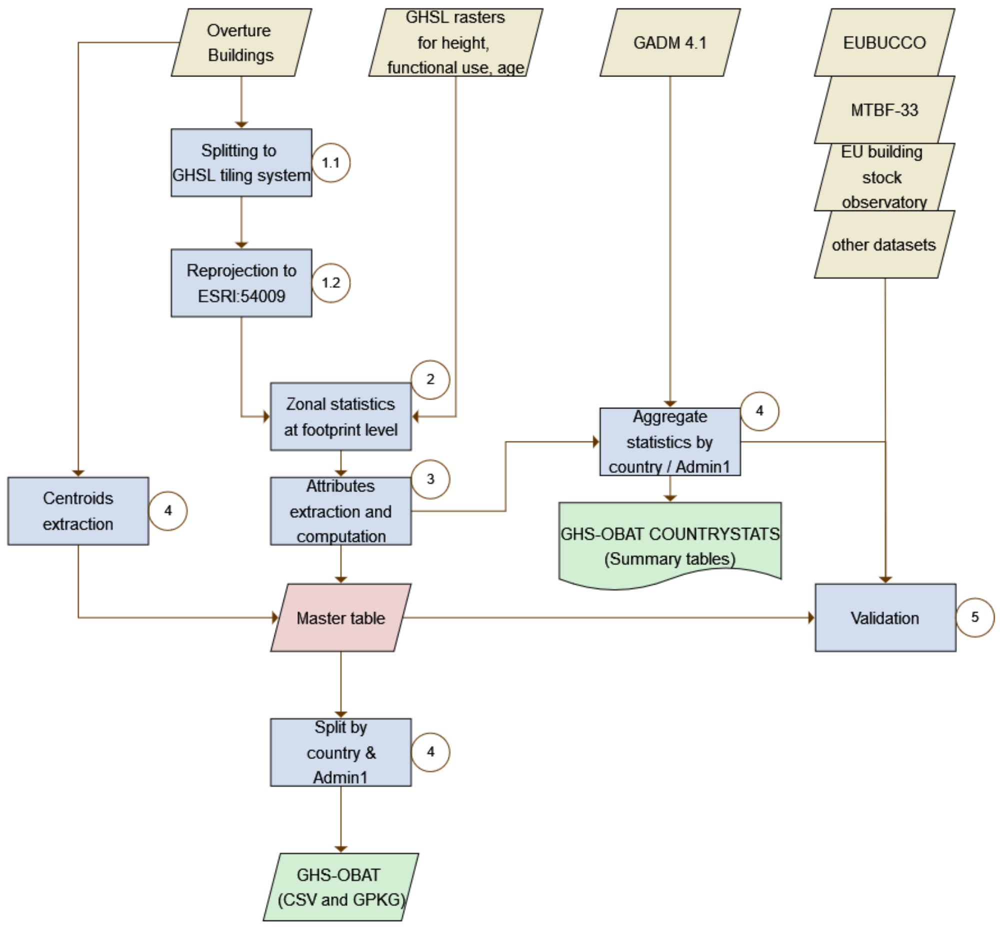
The various attributes are computed for each building footprint, including building height, shape factor, functional use class, and construction epoch.

- The building height is computed as the mean of intersecting grid cells, while the shape factor is calculated as:

  $$shape\ factor = \frac{S}{V} = \frac{2 \cdot A + P \cdot h}{A \cdot h} = \frac{2A}{A \cdot h} + \frac{P \cdot h}{A \cdot h} = \frac{2}{h} + \frac{P}{A}$$

  with A footprint area, P footprint perimeter and h mean building height; it approximates the ratio of the building outer surface (S) to the building volume (V), when simplifying the building volume to an extruded “shoebox” (Level Of Detail - LOD1).
- The functional use class is derived from the majority of grid cells within the footprint's boundary.
- The construction epoch is obtained from the gridded GHS-AGE data product, by majority of grid cells within the footprint's boundary.

Figure 9. illustration of the methodological workflow. Datasets are shown as parallelograms, statistics reports as a rectangle with wavy boundaries, processing steps as rectangles. Input data is in yellow, intermediate (unpublished) data is in red, output is in green. Blue colour refers to processing steps, with a circled number pointing to the section describing each into detail.



Source: JRC analysis.

3.2.1 Open code

The code used for processing has been re-adapted for extracting GHSL attributes from any building footprint dataset, and made available at the following link on the EU code repository:

https://code.europa.eu/jrc-ghsl/building-data-integrator

3.3 Validation

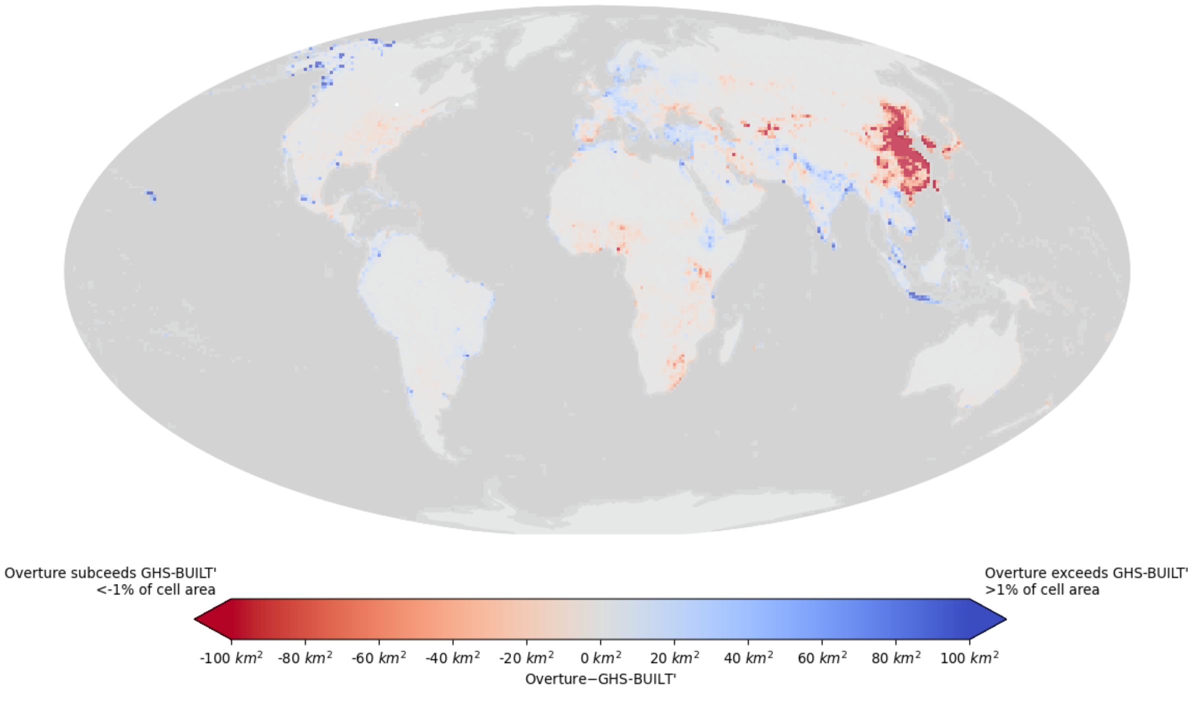
In addition to the main data sources mentioned in section 0 (Overture buildings for footprints polygons and GHSL for thematic attributes), the EUBUCCO database [15] provides authoritative attributes for validation purposes. EUBUCCO is a EU-wide dataset comprising building footprints and attributes on building heights and building construction epoch for a limited subset of buildings; it includes data from National Mapping and Cadastral Agencies (NMCAs) and fills the gaps with the community-based dataset Open Street Map (OSM).

The use of EUBUCCO data as reference for validation purposes is limited to authoritative sources only. The results show that the estimated attributes generally agree with the authoritative attributes, but there are some discrepancies. In particular, the building height is underestimated (Mean Absolute Error at feature level equal to 2.89 m, decreasing if weighted by the area of the buildings) and small non-residential buildings are not effectively detected. Buildings labelled as older than 1980 in GHS-OBAT have a 40% chance of being correct, whereas this likelihood increases to more than 65% for more recent buildings (after 1980). Compared against aggregated figures from the EU Building Stock Observatory [16], one third of EU countries have more than 10% difference in the number of non-residential buildings, while the overrepresentation of oldest buildings (before 1980) involves 16% buildings at country level on average.

3.4 Coverage

The coverage of GHS-OBAT is limited to the one of the Overture buildings dataset. In comparison against the GHS-BUILT-S product (section 2.1.1), weighted by appropriate adjustment factors [2, Tbl. 17], Overture shows non-negligible differences. The results (**Figure 10**) show that there are severe discrepancies in China and small regions in Central Asia, with probable underrepresentation of buildings in Overture. Other mismatches in Northern India, Malaysia and Alaska might be explained by wrong detections in Overture or less probable underestimations in GHS-BUILT-S.

Figure 10. Overture building surface benchmarking against adjusted GHS-BUILT-S (described in section 2.1.1).



Source: JRC analysis.

3.5 How to cite

Dataset:

Florio, P., Politis, P., Goch, K., Uhl, J.H., Melchiorri, M. et al*., 'GHS-OBAT R2024A - GHS Building Attributes at footprint level, with age, function and morphological information (2020)'.* European Commission, Joint Research Centre (JRC), 16 December 2024, accessed 13 June 2025, https://doi.org/10.2905/f41a22f1-5741-4c41-86eb-6384654f6927.

Concept & Methodology:

Florio, P., Politis, P., Krasnodębska, K., Uhl, J.H., Melchiorri, M. et al., *GHS-OBAT: Global, Open Building Attribute Data Reporting Age, Function, Height and Compactness at Footprint Level*, Data in Brief, June 2025, 11751, https://doi.org/10.1016/j.dib.2025.111751, JRC140775.

Code:

Politis P*.,* *Building Data Integrator*, https://code.europa.eu/jrc-ghsl/building-data-intergrator.

Conclusions

The Global Human Settlement – Open Building Attribute Table (GHS-OBAT) enhances 2.3 billion Overture building footprints available in the open domain with remotely sensed thematic attributes extracted from the publicly accessible Global Human Settlement Layer data. The attributes provided, such as estimates of the building construction epoch, building height, type of building functional use (residential / non-residential), and building shape factor, are particularly valuable for energy planning, especially considering the transition towards renewable energy, and for evaluating the built environment's effects on climate change. This integrated dataset is pertinent for disaster risk management, as it improves the evaluation of the vulnerability of the building stock and aids in the allocation of resources during emergencies. Policymakers, urban planners and designers, demographers, educators, and urban geographers gain insights from this dataset, which delves into the historical evolution of the building stock and the geographical distribution of building use types (residential / non-residential), facilitating studies of urban morphology that encompass both vertical and temporal aspects. The use of the GHS-OBAT dataset is recommended for querying and performing cross-statistics reported at aggregated level, or to facilitate international comparison on homogeneous attributes.

References

[1] M. Pesaresi *et al.*, “The Global Human Settlement Layer,” in *Proc. of Mapping Urban Areas from Space (MUAS) conference*, ESA, Nov. 2015.

[2] M. Pesaresi *et al.*, “Advances on the global human settlement layer by joint assessment of earth observation and population survey data,” *International Journal of Digital Earth*, vol. 17, no. 1, p. 2390454, Dec. 2024, doi: 10.1080/17538947.2024.2390454.

[3] European Commission. Joint Research Centre., *GHSL data package 2023.* LU: Publications Office, 2023. Accessed: Apr. 17, 2024. [Online]. Available: https://data.europa.eu/doi/10.2760/098587

[4] M. Pesaresi and P. Politis, “GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030).” European Commission, Joint Research Centre (JRC), Apr. 25, 2023. doi: 10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA.

[5] M. Pesaresi and P. Politis, “GHS-BUILT-C R2023A - GHS Settlement Characteristics, derived from Sentinel2 composite (2018) and other GHS R2023A data.” European Commission, Joint Research Centre (JRC), Apr. 25, 2023. doi: 10.2905/3C60DDF6-0586-4190-854B-F6AA0EDC2A30.

[6] M. Pesaresi and P. Politis, “GHS-BUILT-H R2023A - GHS building height, derived from AW3D30, SRTM30, and Sentinel2 composite (2018).” European Commission, Joint Research Centre (JRC), Apr. 25, 2023. doi: 10.2905/85005901-3A49-48DD-9D19-6261354F56FE.

[7] J. H. Uhl, M. Pesaresi, P. Politis, K. Krasnodębska, P. Florio, and T. Kemper, “GHS-AGE: Characterizing settlement age at planetary scale using the Global Human Settlement Layer,” 2024.

[8] Overture Maps Foundation, “Overture Maps - Buildings.” 2024. [Online]. Available: https://overturemaps.org/

[9] “GADM database of Global Administrative Areas, version 4.1.” 2022. [Online]. Available: www.gadm.org

[10] T. Tadono, H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto, “Precise Global DEM Generation by ALOS PRISM,” *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. II–4, pp. 71–76, Apr. 2014, doi: 10.5194/isprsannals-II-4-71-2014.

[11] B. Rabus, M. Eineder, A. Roth, and R. Bamler, “The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 57, no. 4, pp. 241–262, Feb. 2003, doi: 10.1016/S0924-2716(02)00124-7.

[12] C. Corbane *et al.*, “A global cloud free pixel- based image composite from Sentinel-2 data,” *Data in Brief*, vol. 31, p. 105737, Aug. 2020, doi: 10.1016/j.dib.2020.105737.

[13] P. Depecker, C. Menezo, J. Virgone, and S. Lepers, “Design of buildings shape and energetic consumption,” *Building and Environment*, vol. 36, no. 5, pp. 627–635, Jun. 2001, doi: 10.1016/S0360-1323(00)00044-5.

[14] P. Florio *et al.*, “GHS-OBAT: Global, Open Building Attributes at footprint level including age, function, height and compactness,” 2024.

[15] N. Milojevic-Dupont *et al.*, “EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings,” *Sci Data*, vol. 10, no. 1, p. 147, Mar. 2023, doi: 10.1038/s41597-023-02040-2.

[16] European Commission, “EU Building Stock Observatory.” 2024. [Online]. Available: https://building-stock-observatory.energy.ec.europa.eu/

List of abbreviations and definitions

Abbreviations	Definitions
ANBH	Average Net Building Height
EU	European Union
FUN	Functional use classification in the built-up domain, a sub-product of GHS-BUILT-C
GADM	Global database of Administrative areas
GHSL	Global Human Settlement Layer
GHS-AGE	Global Human Settlement Age, data released under GHSL (GHSL R2024A)
GHS-BUILT-C	Global Human Settlement Built-up Characteristics, data released under GHSL (GHSL R2023A)
GHS-BUILT-S	Global Human Settlement Built-up Surface, data released under GHSL (GHSL R2023A)
LOD	Level Of Detail
OSM	Open Street Map

Countries and territories abbreviations are in compliance with ISO 3166-1 alpha-3 standard

List of figures

List of tables

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
Joint-research-centre.ec.europa.eu

Publications Office
of the European Union