ORDNANCE SURVEY GB

OS OPEN BUILT UP AREAS™ – TECHNICAL SPECIFICATION

Version history

Version	Date	Description
1.0	11/2022	Initial version.

Purpose of this document

This document provides information about and insight into the OS Open Built Up Areas product and its potential applications. For more information on the contents and structure of OS Open Built Up Areas, please refer to the Overview.

The terms and conditions on which OS Open Built Up Areas is made available to you and your organisation are contained within Open Government Licence (https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/), which allows you to use OS Open Built Up Areas in any way and for any purpose. We simply ask that you acknowledge us with: "Contains OS data © Crown Copyright and database rights (2022)".

We may change the information in this document at any time. We do not accept responsibility for the content of any third-party websites referenced or accessed in or through this document.

This document has been screened according to Ordnance Survey's Equality Scheme. If you have difficulty reading this information in its current format and would like to find out how to access it in a different format (braille, large print, computer disk or in another language), please contact us on:
+44 (0)3456 05 05 05.

Copyright in this document

Trade marks

Ordnance Survey, OS and the OS Logos are registered trade marks and OS Open Built Up Areas and OS OpenData are trade marks of Ordnance Survey, Britain's mapping agency.

OGC is a registered trademark of the Open Geospatial Consortium, Inc.

Contact details

OS website 'Contact us' page (https://www.ordnancesurvey.co.uk/contact-us).

Contents

1.   Introduction

OS Open Built Up Areas is a dataset representing the built-up areas of Great Britain. It was designed primarily to underpin the statistical analysis that contributes to policy enablement across the public sector.

The Office for National Statistics (ONS) and Scottish Government made significant contributions in the design of the dataset to ensure it is fit for purpose and focussed on the needs of the wider public sector.

OS Open Built Up Areas is released and maintained as an OS OpenData product with Open Government Licensing (OGL), which means that anyone can use this data for a wide range of purposes.
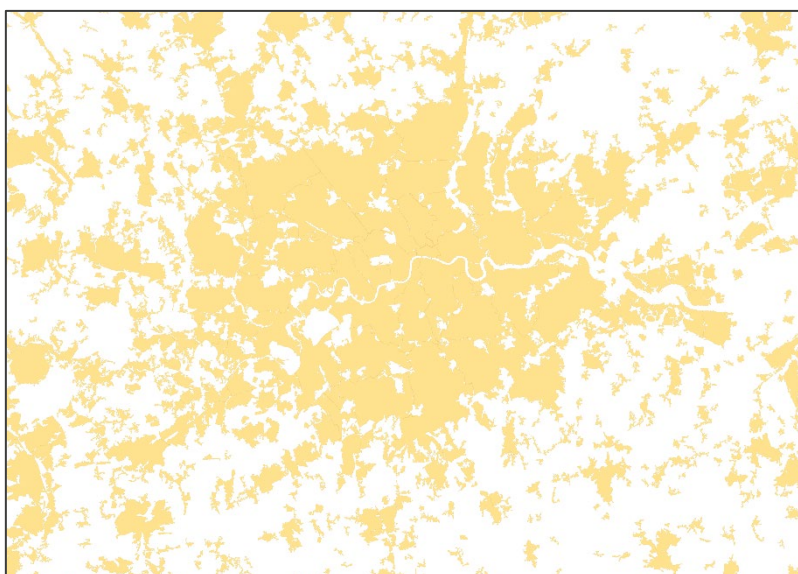

Figure 1: OS Open Built Up Areas.

1.1   Feature types

OS Open Built Up Areas was co-designed with ONS and Scottish Government and is based on their requirements. This collaboration enabled OS to create a customer-focussed product that underpins statistical analysis and supports policy development.

OS Open Built Up Areas is classified into three feature types:

• **Built Up Areas** is the aggregation of Built Up Extents and Non Built Up Extents.
 This is a multipart area feature with a single Government Statistical Service (GSS) code and name or names.

• **Built Up Extents** is the representation of built-up areas only within the extent of Built Up Areas. This is the area that remains after Non Built Up Extents have been removed from Built Up Areas.
 This is a multipart area feature with the same GSS code as the related Built Up Areas with the same name or names.

• **Non Built Up Extents** is the representation of non-built-up areas within the extent of Built Up Areas.
 This is a multipart area feature with the same GSS code as the related Built Up Areas with the same name or names.

Each feature type has associated attribution that is detailed in this document.

Figure 2: Built Up Areas Feature Type.
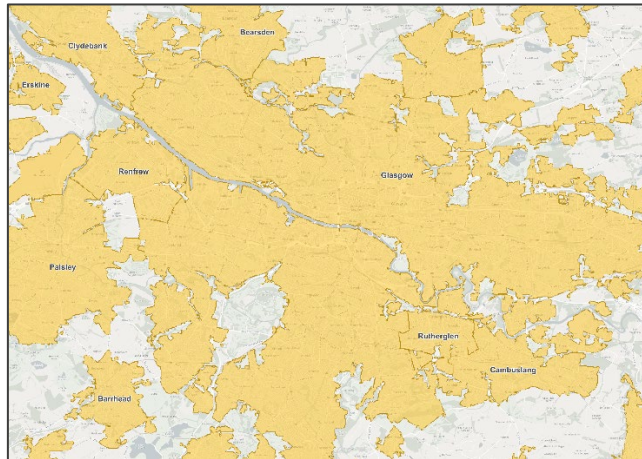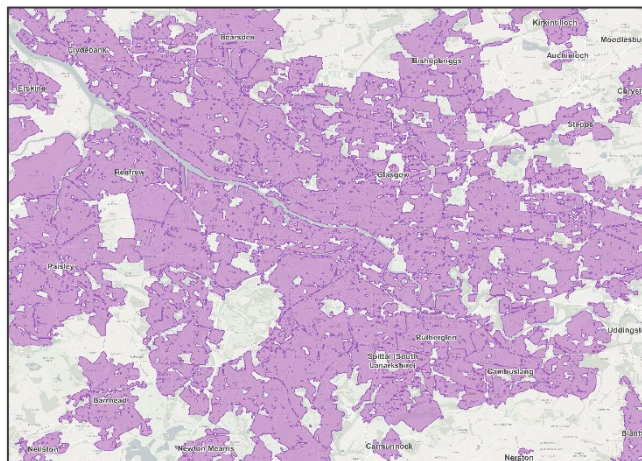

Figure 3: Built Up Extents Feature Type.

1.2   Identifiers

The persistent managed identifier in OS Open Built Up Areas is the GSS code provided by ONS and Scottish Government. The GSS code is consistent across the three feature types and is prefixed with:

- *E63* for England.

- *S45* for Scotland.

- *W45* for Wales.

- *K08* for Built Up Areas that cross the Wales-England border.

1.3 Data sources

OS Open Built Up Areas is created from the following Ordnance Survey data content stores:

• Topographic Buildings, Roads, Residential Gardens and Made Surfaces.

• Land Use Sites.

• Settlement Named Areas.

1.4 Methodology

The OS Open Built Up Areas dataset is created by following the methodology detailed below:

1. **Identify built-up cells**: 25m by 25m cells are flagged as built up based on the proportion of the cell that is classified as Building, Road, Residential Garden, Made Surface, or built-up types of Land Use Site.

2. **Assign names**: Each cell is assigned a name from the Settlement Named Area dataset.

3. **Create initial built-up area polygons**:

   a. Adjacent, identically-named 25m x 25m cells are merged together.

   b. Adjacent built-up areas that are not Cities or Towns are merged together.

   c. Smaller built-up areas are merged into nearby larger built-up areas.

4. **Filter to create final geometry**:

   a. Built-up areas >=200 000m^2 are retained and become Built Up Extents.

   b. Built Up Areas and Non Built Up Extents are then generated from the Built Up Extents.

5. **Refine names**: The name or names of the most significant constituent settlements are applied to Built Up Areas that are not Cities or Towns.

   • If two or more Built Up Areas have the same name, the name of the lowest-tier local authority in which the Built Up Area is located is appended in brackets.

   • If two or more Built Up Areas have the same name within a single local authority, the name of the parish in which the Built Up Areas is located is appended in brackets.

   • If two or more Built Up Areas have the same name within a single parish, the name of the nearest Built Up Area is appended in brackets.

6. **Assign GSS codes**: GSS codes, supplied by ONS (for England and Wales) and Scottish Government, are applied to the three feature types.

1.5   Available formats

OS Open Built Up Areas is supplied in the following formats:

• CSV (comma-separated values).

• GeoPackage.

See Supply formats overview for more information.

1.6   Currency

The first release of OS Open Built Up Areas was in December 2022. The second release is scheduled for April 2024. After this, the product is scheduled to be released every two years.

2.   Product structure

OS Open Built Up Areas is classified into three feature types: Built Up Areas, Built Up Extents, and Non Built Up Extents.
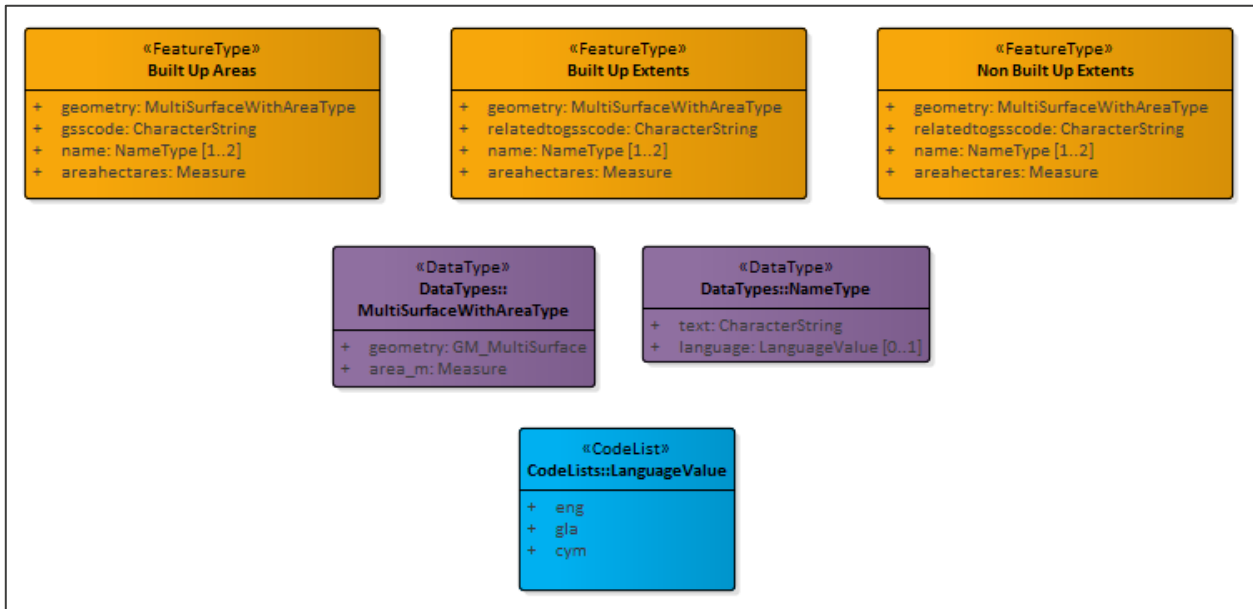

Figure 4: Unified Modeling Language (UML) diagram showing the data structure of OS Open Built Up Areas

2.1   Built Up Areas attributes

«featureType» Built Up Areas
Definition: Notional polygons representing built-up areas, derived from a 25m x 25m grid of topographic data, and based on the classification of OS topographic area and land use features.

Attribute: geometry	
Definition: Polygon geometry for the feature with area in square metres.	
Type: MultiSurfaceWithAreaType	Multiplicity: [1]

Attribute: gsscode	
Definition: A GSS identifier for Built Up Areas provided by ONS and Scottish Government.	
Type: CharacterString	Multiplicity: [1]

Attribute: name	
Definition: The name (including any language alternatives) of the highest classification of settlement that relates to a Built Up Area polygon or, if there are multiple settlements for the same classification, the name of the most significant settlement.	
Type: NameType	Multiplicity: [1..2]

Attribute: areahectares	
Definition: The area of each Built Up Area polygon in hectares.	
Type: Measure	Multiplicity: [1]

2.2   Built Up Extents attributes

«featureType» Built Up Extents	
Definition: Notional polygons representing built-up areas, excluding non-built up areas (for example, wooded strips, waterbodies, and roadside land), within the extent of a Built Up Area polygon, derived from a 25m x 25m grid of topographic data, and based on the classification of OS topographic area and land use features.	
Attribute: geometry	
Definition: Polygon geometry for the feature with area in square metres.	
Type: MultiSurfaceWithAreaType	Multiplicity: [1]
Attribute: relatedtogsscode	
Definition: A reference to the GSS code for Built Up Areas provided by ONS and Scottish Government.	
Type: CharacterString	Multiplicity: [1]
Attribute: name	
Definition: The name (including any language alternatives) of the highest classification of settlement that relates to a Built Up Area polygon or, if there are multiple settlements for the same classification, the name of the most significant settlement.	
Type: NameType	Multiplicity: [1..2]
Attribute: areahectares	
Definition: The area of each Built Up Extent polygon in hectares.	
Type: Measure	Multiplicity: [1]

2.3   Non Built Up Extents attributes

«featureType» Non Built Up Extents	
Definition: Notional polygons representing non-built up areas (for example, wooded strips, waterbodies, and roadside land) within the extent of a Built Up Area polygon, derived from a 25m x 25m grid of topographic data, and based on the classification of OS topographic area and land use features.	
Attribute: geometry	
Definition: Polygon geometry for the feature with area in square metres.	
Type: MultiSurfaceWithAreaType	Multiplicity: [1]
Attribute: relatedtogsscode	
Definition: A reference to the GSS code for Built Up Areas provided by ONS and Scottish Government.	
Type: CharacterString	Multiplicity: [1]
Attribute: name	
Definition: The name, including any language alternatives, of the highest classification of settlement that relates to a Built Up Area polygon or, if there are multiple settlements for the same classification, the names of the most significant settlements.	
Type: NameType	Multiplicity: [1..2]
Attribute: areahectares	
Definition: The area of each Non Built Up Extent polygon in hectares.	
Type: Measure	Multiplicity: [1]

2.4 Data types

«DataType» MultiSurfaceWithAreaType	
Definition: A spatial area object describing the geometry, extent and area of the feature.	
Attribute: geometry	
Definition: Geometry for the feature.	
Type: GM_MultiSurface	Multiplicity: [1]
Attribute: area	
Definition: The measured area of the geometry in square metres.	
Type: Measure	Multiplicity: [1]

«DataType» NameType		
Definition: An object that describes the proper nouns (and their language) that apply to the feature.		
Attribute: text		
Definition: Name assigned to identify the feature.		
Type: CharacterString	Size: 254	Multiplicity: [1]
Attribute: language		
Definition: The language type associated with the name. The valid values are defined in the LanguageValue code list.		
Type: CharacterString	Size: 3	Multiplicity: [0..1]

2.5 Code list

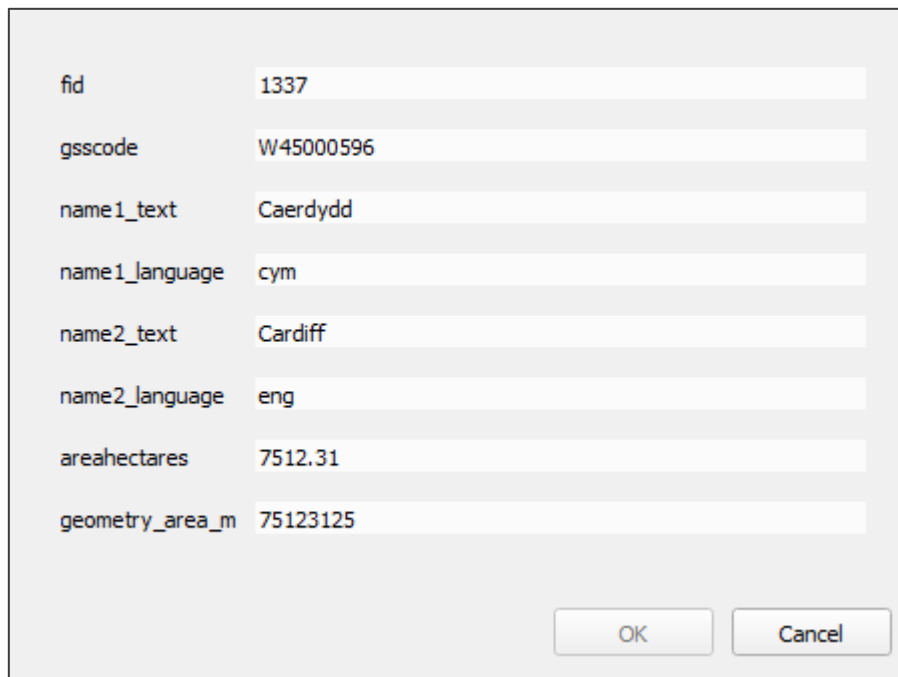| Code list: LanguageValue	
Value	Description
eng	ISO 639-2 code for English language.
gla	ISO 639-2 code for Gaelic; Scottish Gaelic language.
cym	ISO 639-2 code for Welsh language.

3. Supply formats overview

3.1 GeoPackage

OS Open Built Up Areas is supplied as a single GeoPackage for Great Britain. GeoPackage (*.gpkg) is an open, non-proprietary, platform-independent, standards-based data format for geographic information systems (GIS), as defined by the Open Geospatial Consortium (OGC). It is designed to be a lightweight format that can contain large amounts of varied and complex data in a single, easy-to-distribute and ready-to-use file. GeoPackage is natively supported by numerous software applications.

GeoPackage offers users the following benefits:

• The single file is easy to transfer and offers the end user a rich experience.

• Attribute names are not limited in length, making the format user friendly.

• The file size limit is large at 140 TB.
  Note: A file size limit could be imposed by the file system to which the file is written.

• It supports raster, vector and database formats, making it a highly versatile solution.

• It is an OGC standard.

• In most cases, it is a plug-and-play format.

For information on how to open, use and understand a GeoPackage dataset, please refer to our 'Getting Started with GeoPackage' guide (https://www.ordnancesurvey.co.uk/documents/getting-started-with-geopackage.pdf), which is available on the OS website. For further information on GeoPackage, please see the GeoPackage website (https://www.geopackage.org/).

fid	1337
gsscode	W45000596
name1_text	Caerdydd
name1_language	cym
name2_text	Cardiff
name2_language	eng
areahectares	7512.31
geometry_area_m	75123125

Figure 5: GeoPackage attributes for the Built Up Areas Feature Type.

3.2   CSV

A comma-separated values (CSV) file is a common interchange format for spreadsheets and databases that facilitates the simplistic use of data. Each field is either textual or numeric. Within the CSV, each field is separated from the next by a comma. CSV file format is universally supported for easy ingestion into all major database products.

Note: CSV files are designed to be opened in a database or GIS application and opening them in other software applications might corrupt the data. In particular, Excel has a row limit that is easily exceeded by large CSV files. We recommend that you load CSV files directly into a database or GIS, rather than trying to open these files in Excel.

The CSV supply has the following features:

• Geometry is provided as Well-Known Text (WKT).

• Header rows are included in each file.

• There is one record per line in each file.

• Fields are separated by commas.

• Where string fields contain commas, they are delimited by double quotes.

• Double quotes inside strings are escaped by doubling.

• Records are terminated by carriage returns and line feeds.

• Files use UTF-8 character encoding.

gsscode	name1_text	name1_language	name2_text	name2_language	areahectares	geometry_area_m	geometry
W45000596	Caerdydd	cym	Cardiff	eng	7512.31	75123125	MULTIPOLYGON (((322850 179125, 322850 179150, 322850 179175, 322875

Figure 6: CSV attributes for the Built Up Areas Feature Type.

3.3　　Attribute naming format comparison

The names of attributes in CSV and GeoPackage format are very similar. The following table maps CSV attribute names to GeoPackage attribute names:

CSV attributes	GeoPackage attribute
*	fid[1]
gsscode[2]	gsscode[2]
relatedtogsscode[3]	relatedtogsscode[3]
name1_text	name1_text
name1_language	name1_language
name2_text	name2_text
name2_language	name2_language
areahectares	areahectares
geometry_area_m	geometry_area_m
geometry[4]	*[5]

* Indicates that the attribute is not mapped between the formats.

[1] fid is an additional attribute. It is an INTEGER NOT NULL column that acts as a primary key. This attribute is a requirement of the OGC GeoPackage specification.

[2] Present in the Built Up Areas Feature Type only.

[3] Present in the Built Up Extents and Non-Built Up Extents Feature Types only.

[4] CSV contains the Well-Known Text (WKT) data type that represents the vector geometry objects of the feature.

[5] The geometry column is always the second column, but the attribute (or its value) is typically not visible in GIS software. GeoPackage is based on Well-Known Binary (WKB) as defined in ISO/IEC 13249-3:2011.