

Buildings

3.1

Geoscape

07 December 23

Contents

1	Preface	3
1.1	Responsible Party	3
1.2	Copyright and disclaimer	3
1.3	Privacy	3
2	Product Version	5
3	Overview	7
4	Technical Description	9
4.1	Linkages	10
4.2	Attributes	12
4.3	Data Model	16
4.4	Data Dictionary	17
4.5	Domain Values	18
4.6	Data Maintenance	19
4.7	Data Quality	21
4.8	Extent/Geographic Description	26
4.9	Spatial Reference System	26
4.10	Delivery Format	27
4.11	Product Versioning	29

Disclaimer

Geoscape Australia believes this publication to be correct at the time of printing and does not accept responsibility for any consequences arising from the use of the information herein. Readers should rely on their own skill and judgement to apply information to particular issues.

This work is copyright. Apart from any use as permitted under the Copyright Act 1968, no part may be reproduced by any process without prior written permission of Geoscape Australia.

Preface

1.1 Responsible Party

PSMA Australia Limited trading as Geoscape Australia

ABN: 23 089 912 710

Unit 6, 113 Canberra Ave, GRIFFITH ACT 2603 Australia

T: +61 2 6260 9000

E: info@geoscape.com.au

URL: www.geoscape.com.au

1.2 Copyright and disclaimer

Please see geoscape.com.au/legal/data-copyright-and-disclaimer/

1.3 Privacy

Geoscape products and services should not contain any personal names or other personal information. Geoscape undertakes reasonable data cleansing steps as part of its production processes to ensure that is the case. If you think that personal information may have inadvertently been included in Geoscape products or services, please contact support@geoscape.com.au

Product Version

3.1

Overview

Buildings is a national digital dataset representing buildings for each State and Territory in Australia. The purpose of Buildings is to spatially represent building related features for use by industry and government in geographic information systems, visualisations, engineering modelling, and for the derivation of analytics.

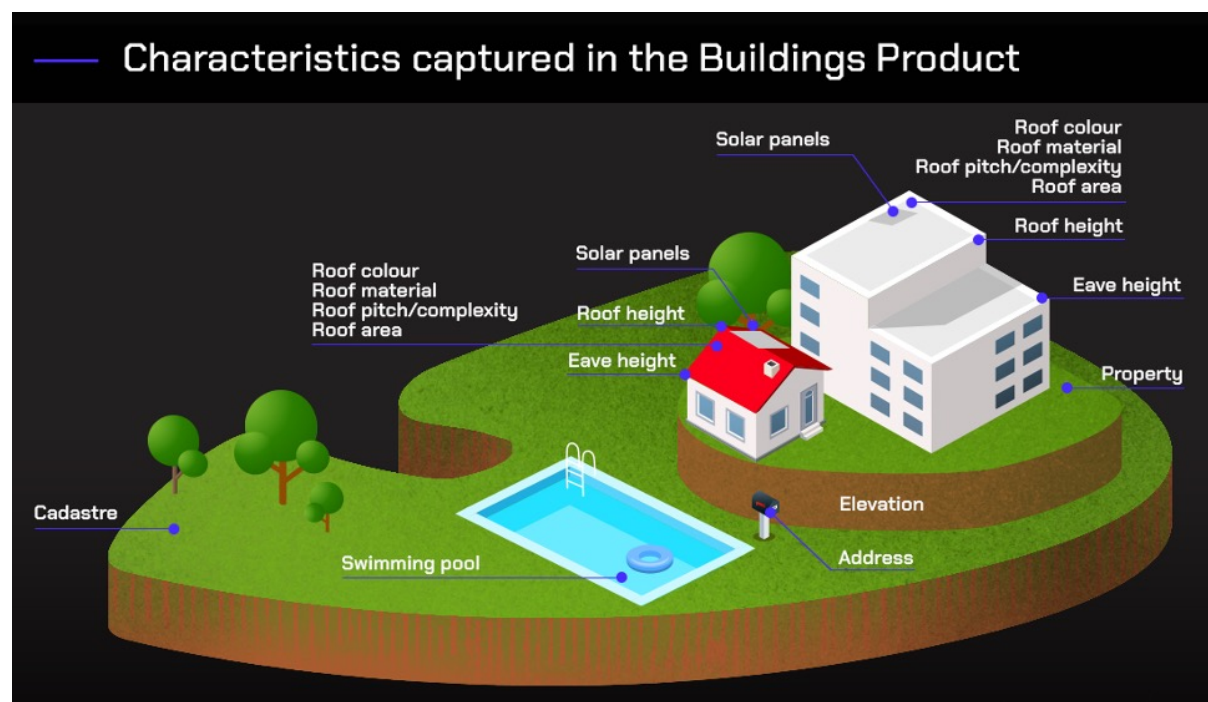


The Buildings product is comprised of four layers; buildings, building_cad, building_property and building_address. The buildings layer provides a spatial representation of the roof outline of a building as well as associated attribution including building height values, roof material and colour, and building area and volume. The aspatial building_cad, building_property and building_address tables provide additional information related to a building.

Geoscape Australia welcomes your feedback on our Buildings product. We also publish regular updates on the development of our products on the Geoscape website (www.geoscape.com.au).

Technical Description

The buildings theme (buildings layer) consists of digital representations of the roof outline of a building. Buildings have been digitised from remotely sensed imagery using a combination of automated and manual processes to identify, extract and orthogonalise objects resembling a building structure greater than 9m². The process also determines building attributes including roof materials, presence of solar panels and presence of swimming pools. Image quality factors including currency, capture geometry, and applicable weather conditions influence the specific image which can be utilised for further processing. A suitable high-quality Digital Surface Model is constructed to assist with feature extraction.



Data quality and potential capture timelines will vary across Australia based on three categories. Each category has been developed based on several factors defined by the population distribution (categorised based on population size), industrial/commercial activities, the probability of natural events (e.g. flooding) and the image source.

- **Urban (satellite source)** - areas with a population greater than 200, or with significant industrial/commercial activity in a visual assessment, digitised from satellite imagery
- **Urban (aerial source)** - areas with a population greater than 200, or with significant industrial/commercial activity in a visual assessment, digitised from aerial imagery
- **Rural – all other areas**

Table 1: Building layer and attribution for different Area Categories.

Class / Attribute	Urban	Rural
Building Attributes		
Address count	Y	Y
Building centroid	Y	Y
Building outline (polygon)	Y	Y
Eave height	Y	N
Ground elevation	Y	N
Ground level Z value for vertices and centroid*	Y	Y
Maximum roof height	Y	N
Number vertices for a polygon	Y	Y
Polygon area	Y	Y
Roof colour	Y	N
Primary roof material	Y	N
Roof type	Y	N
Solar panel indicator	Y	N
Swimming pool adjacent indicator	Y	N
Building volume	Y	N
Planning zone	Y	Y
State/Territory	Y	Y
Estimated levels	Y	N
Roof shape	Y	N
Tree overhang	Y	N
Geoscape Data linkages		
Cadastre	Y	Y
G-NAF	Y	Y
Locality	Y	Y
Mesh Block	Y	Y
Property	Y	Y

* Not available in MapInfo TAB format

4.1 Linkages

Buildings uses the following Geoscape products for inputs into the processing:

- G-NAF,
- Cadastre,
- Property,
- Localities and ABS 2016 Mesh Blocks and
- Geoscape Planning

The linkages with buildings are explained in detail below.

Users should note that the listed Geoscape products above are not part of Buildings. G-NAF, ABS 2016 Mesh Blocks and Localities are available through Geoscape's Partner Network or under an open licence from the Commonwealth of Australia at www.data.gov.au.

The Mesh Block data is sourced from the Australian Bureau of Statistics (ABS) and is part of their Australian Statistical Geography Standard (ASGS).

4.1.1 building_cadastral linkage and relationship confidence

A building will be linked to a cadastral parcel where the area of overlap is 7% or greater of the buildings area or 40% or greater of the cadastral parcels area. Cadastral parcels with a parcel type of road are not used in the creation of these relationships. If a building is related to multiple cadastral parcels and one of those parcels fully contains another of these parcels, a building linkage to the container cadastral parcel is not created.

Relationship confidence values are assigned to describe the confidence in these building_cadastral relationships. The relationship confidence values use the area of overlap to broadly describe the likelihood that a building is related to a cadastral parcel, with higher overlap inferring a greater confidence in the relationship.

Spatial Overlap	relationship_confidence
>= 80% building area overlap or >= 40% parcel area overlap	90
60% to <80% building area overlap	80
40% to <60% building area overlap	50
20% to <40% building area overlap	30
7% to <20% building area overlap	20

4.1.2 building_property linkage and relationship confidence

A building will be linked to a property parcel where the area of overlap is 7% or greater of the buildings area or 40% or greater of the property parcels area. If a building is related to multiple property parcels and one of those parcels fully contains another of these parcels, a building linkage to the container property parcel is not created.

Relationship confidence values are assigned to describe the confidence in these building_property relationships. The relationship confidence values use the area of overlap to broadly describe the likelihood that a building is related to a property parcel, with higher overlap inferring a greater confidence in the relationship.

Spatial Overlap	relationship_confidence
>= 80% building area overlap or >= 40% parcel area overlap	90
60% to <80% building area overlap	80
40% to <60% building area overlap	50
20% to <40% building area overlap	30
7% to <20% building area overlap	20

4.1.3 building_address linkage and relationship confidence

A building will be linked to a Geoscape Address where the address either intersects the building geometry or the address intersects a cadastral or property parcel that has been related to the building through the above building_cadastral or building_property linkage rules. Linkage is limited to active principal addresses relating to the building centroid, property centroid, frontage centre setback or property access point setback. Geoscape Addresses is an address dataset containing a standardised and quality assured national set of addresses that are linked to G-NAF and updated as frequently as addresses are provided by each jurisdiction. This link to Geoscape Addresses assigns an address string to the building_address record and where the Geoscape Address is related to a G-NAF record the address_detail_pid is also assigned to the building_address. Not all G-NAF records will be linked, for example at developments with multiple units only the primary address might be linked to a building record. Relationship confidence values are assigned to describe the confidence in these building_address relationships. The relationship confidence value for a building_address relationship will be 95% if the address intersects

the building. If the building_address relationship has been created through cadastral or property relationships, the relationship confidence value for the building_address record will match that used for the building to parcel relationship, described above.

4.1.4 Building to Localities linkage

A building will be assigned a locality_pid value relating to the gazetted locality boundary that the building is primarily within, based on maximum area of intersection. This linkage will not be made if the maximum area of intersection is less than 7% of the building area. If a gazetted locality does not intersect the building, the locality_pid of a district (ACT) may be assigned based on maximum area of intersection if this intersection is 7% or greater based on the building area.

4.1.5 Building to Mesh Block linkage

A building will be assigned a mesh_block_code value relating to the 2016 Australian Bureau of Statistics (ABS) Mesh Block that the building is primarily within, based on maximum area of intersection. This linkage will not be made if the maximum area of intersection is less than 7% of the building area.

4.1.6 Building to Planning Zone linkage

A building will be assigned a generalised planning zone description through linkage with the Geoscape Planning product, using the Cadastre relationship with the highest confidence. Where two or more cadastral relationships exist for a building with the same confidence, the generalised planning description is assigned using the following priority order (with "Residential" as the highest priority):

Residential > Commercial/Business > Industrial/Utilities > Community Use > Mixed Use > Special Use > Transport/Infrastructure > Rural/Primary Production > Conservation/National Park > Recreational/Open Space > Water

4.2 Attributes

4.2.1 Building Source

Informs about the source imagery, whether it is a satellite or aerial capture and the resolution of the imagery in centimetres, used for each building record.

4.2.2 Quality Class

Categorises each building record into either 'Urban', areas with a population greater than 200, or with significant industrial/commercial activity in a visual assessment, or 'Rural', which includes all other areas. The quality class of 'Urban' denotes buildings with high population of attributes and high horizontal positional accuracy. Buildings with a quality class of 'Urban' are either captured from satellite or aerial imagery. The quality class of 'Rural' denotes buildings with generally lower population of attributes and lower horizontal positional accuracy than the 'Urban' records.

4.2.3 Address Count

Provides the total count of addresses at a building record.

4.2.4 Swimming Pools

The `swimming_pool_adjacent` attribute is an indicator of the presence or absence of a swimming pool on the properties associated with the building feature. The `swimming_pool_review_date` provided the date at which the presence of a swimming pool at the property was last reviewed.

4.2.5 Solar Panels

Provided as an attribute (`solar_panel`) on each building feature, this is an indicator of the presence or absence of a photovoltaic solar panel on the roof surface. Other types of solar panels e.g. solar hot water can be captured as false positives and impact the classification correctness of this attribute. The `solar_panel_review_date` provides the date at which the presence of solar panels was last reviewed.

4.2.6 Roof Height

This is the relative height in metres above ground surface of the identified maximum roof height within the building boundary and may include the height of plant rooms and building fixtures. Roof height is measured from the lowest elevation where the constructed walls intersect the ground to the highest point of the constructed building that is not on a spire, antenna or similar.

4.2.7 Eave Height

This is the relative height in metres above the ground surface of the identified eave of a building. All efforts have been taken to identify an 'eave' for each building boundary. Multilevel buildings and irregular facades may impact the accuracy of this value. For records within the category Urban (aerial source) the eave height reflects the height of the highest eave of a multi-level building.

4.2.8 Ground Elevation

This value represents the building ground surface height in meters with respect to the Australian Height Datum (AHD) at the location of the building centroid. The centroid is calculated as the geometric centroid of the building polygon, for irregularly shaped features the centroid is ensured to lie within the bounds of the polygon. This centroid location is intersected with the best available data to assign ground elevation. The source of the data used for assignment of elevation for each building feature is contained within the `ground_elevation_source` attribute.

4.2.9 Primary Roof Material

Provided as an attribute on each building, the roof material is the classification of the primary roof material found across the building surface. The `primary_roof_material` attribute is classified from image pixels that intersect the building. The image signature of the intersecting pixels is matched to a spectral library of known material compositions to determine the best matching material for each pixel.

4.2.10 Roof Type

Provided as an attribute on each building feature is the classification of the type of shape of a roof structure. The roof type is an approximation of the roof shape based off a statistical calculation of the elevation profile of the roof.

4.2.11 Roof Colour

Provided as an attribute on each building, the roof colour is the mean hexadecimal value from all image pixels considered to be comprised of the primary roof material for that building.

4.2.12 Area

For buildings captured from satellite imagery the area attribute value is a derived value calculated during processing. For buildings captured from aerial imagery the area value does not include parts of a building with a roof material of 'Fiberglass/Plastic'. These 'Fiberglass/Plastic' parts of the building are not included as they generally represent unenclosed outdoor coverings. Area values are calculated with features projected in a Transverse Mercator coordinate system.

4.2.13 Volume

Provided as an attribute on each building feature, this is the volume in cubic metres of a building estimated using the roof height and building area for buildings derived from satellite imagery. Where a building has been captured from aerial imagery as a complex polygon and generalised into the Buildings product as a single outline the volume is calculated from the complex building representation and excludes parts of the building with 'Fiberglass/Plastic' roof material.

4.2.14 Geometry Quality

A quality flag has been included to identify buildings that have been modified during data correction processes, or that a quality issue was identified but an automatic correction could not be made.

4.2.15 Is Residential

The linkage table between address and building features a separate attribute that has been included to indicate if the address has been identified as currently or having previously thought to be residential. The source of the residential indicator is currently derived from Commonwealth Government sources.

4.2.16 Number Vertices

A total count of the number of vertices for all rings of the polygon feature representing a building. Polygons have the vertex representing the start and end point counted as one therefore a rectangular polygon has four vertices.

4.2.17 Estimated Levels

The number of levels that has been estimated for an urban building. Rural buildings will be null as height information is required to calculate the attribute.

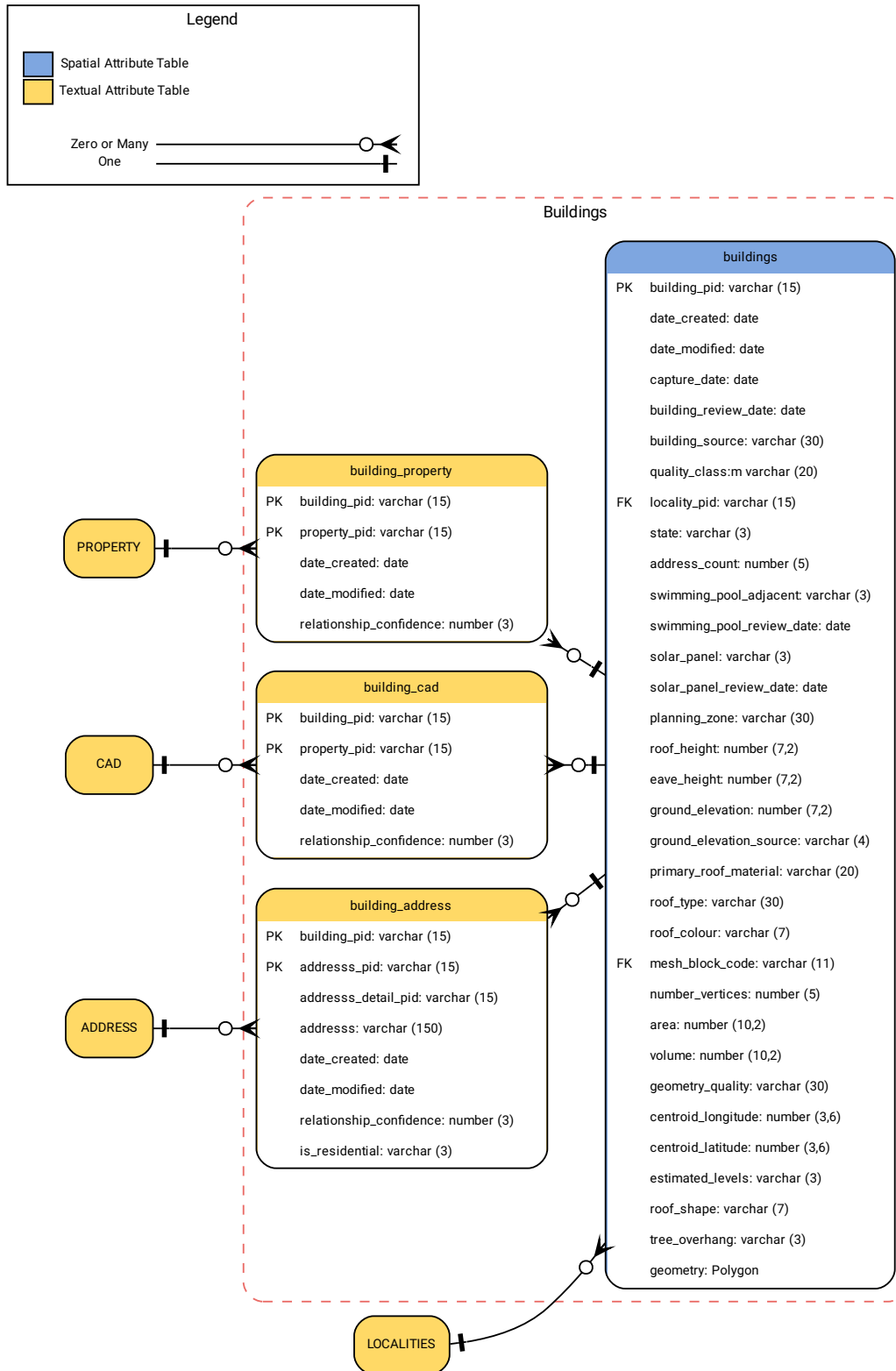
4.2.18 Roof Shape

The shape of the building's roof (e.g. 'Hipped', 'Gabled', 'Flat' etc.) which aligns to domain values in the CityGML standard. The attribute is currently available for aerial derived buildings.

4.2.19 Tree Overhang

Identifies a tree overhang for an aerial building. A 'Yes' value is used where a tree overhangs the building and a 'No' value is used where there is no tree overhanging the building. The attribute is currently available for aerial derived buildings.

4.3 Data Model



4.4 Data Dictionary

Attribute	Data Type	Description
building_pid	character string (15)	Persistent identifier for the building.
date_created	date (dd-mm-yyyy)	The date of record creation for the building.
date_modified	date (dd-mm-yyyy)	The most recent date that an attribute (not including building_review_date, swimming_pool_adjacent) has been modified for the building.
capture_date	date (dd-mm-yyyy)	The capture date of the source imagery used to derive the original geometry of the building.
building_review_date	date (dd-mm-yyyy)	The most recent date of source imagery from which the building was reviewed for update.
building_source	character string (30)	The methodology of capture for the building attributes.
quality_class	character string (20)	Represents whether the building has been classed as either 'Urban' or 'Rural'.
*locality_pid	character string (15)	The locality the building is primarily within.
state	character string (3)	The abbreviated name of the State or Territory that the building is primarily within.
address_count	number (5)	Total number of addresses related to a building.
swimming_pool_adjacent	character string (3)	Indicates whether a swimming pool is on a property related to the building.
swimming_pool_review_date	date (dd-mm-yyyy)	The most recent review date for the swimming_pool_adjacent field.
solar_panel	character string (3)	Describes whether a visible photovoltaic solar panel is identified for the building.
solar_panel_review_date	date (dd-mm-yyyy)	The most recent review date for the solar_panel field.
planning_zone	character string (30)	The planning zone type assigned to a building through linkage with the Geoscape Planning Zone.
roof_height	number (7,2)	The height of the building.
eave_height	number (7,2)	The average height of the part of a buildings roof that meets or overhangs the walls (eave).
ground_elevation	number (7,2)	The elevation of the ground at the centroid of the building.
ground_elevation_source	character string (4)	The grid separated distance (GSD) of the source elevation model used to extract ground elevation.
primary_roof_material	character string (20)	The material that the roof is primarily made from.
roof_type	character string (30)	Indicative shape of the roof of the building.
roof_colour	character string (7)	The measured Red/Green/Blue (RGB) colour value in hexadecimal format (#rrggbb) for the roof.
*mesh_block_code	character string (11)	The 2016 Australian Bureau of Statistics (ABS) Mesh Block that the building is primarily within.
number_vertices	number (5)	The number of vertices for the building geometry.
area	number (10,2)	The area in square metres of the building geometry.
volume	number (10,2)	The volume of the polygon in cubic metres.
geometry_quality	character string (30)	Describes quality issues with the geometry of the building.
centroid_longitude	number (9,6)	The longitude of the building centroid.
centroid_latitude	number (9,6)	The latitude of the building centroid.
estimated_levels	number (3)	The estimated number of levels for the building.
roof_shape	character string (7)	The shape of the aerial building's roof (e.g. 'Hipped', 'Gabled', 'Flat' etc.) in line with the City of Sydney Aerial Building Footprint.
tree_overhang	character string (3)	Provides an attribute that describes whether an identified tree overhangs the aerial building footprint.
geometry	Polygon	The geometry field for the buildings table, which represents the building polygon (including any overhangs).

Table 3: buildings_cad

Attribute	Data Type	Description	Primary Key	Mandatory	10 Character Alias
building_pid	character string (15)	The persistent identifier of the building related to a cadastral record.	Yes	Yes	BLD_PID
*cadastre_pid	character string (15)	The unique persistent identifier for each cadastre record associated with the building record by spatial intersection.	Yes	Yes	CAD_PID
date_created	date (dd-mm-yyyy)	The date of record creation for the building_cadastre relationship.	No	Yes	DT_CREATE
date_modified	date (dd-mm-yyyy)	The most recent date that an attribute has been modified for the building_cadastre relationship.	No	No	DT_MOD
relationship_confidence	number (3)	The percentage confidence that a building has a relationship with a cadastre parcel.	No	Yes	REL_CONF

Table 4: buildings_property

Attribute	Data Type	Description	Primary Key	Mandatory	10 Character Alias
building_pid	character string (15)	The persistent identifier of the building related to a property record.	Yes	Yes	BLD_PID
*property_pid	character string (15)	The unique persistent identifier for each property record associated with the building record by spatial intersection.	Yes	Yes	PR_PID
date_created	date (dd-mm-yyyy)	The date of record creation for the building_property relationship.	No	Yes	DT_CREATE
date_modified	date (dd-mm-yyyy)	The most recent date that an attribute has been modified for the building_property relationship.	No	No	DT_MOD
relationship_confidence	number (3)	The percentage confidence that a building has a relationship with a property parcel.	No	Yes	REL_CONF

Table 5: buildings_address

Attribute	Data Type	Description	Primary Key	Mandatory	10 Character Alias
building_pid	character string (15)	The persistent identifier of the building related to an address record.	Yes	Yes	BLD_PID
address_pid	character string (15)	The address pid that is related to the building feature. The persistent identifier is unique to the address feature this record relates to.	Yes	Yes	ADD_PID
*address_detail_id	character string (15)	The unique persistent identifier for each address_detail record associated with the address record.	No	No	ADD_DT_PID
address	character string (150)	The address string from the related address record.	No	Yes	ADDRESS
date_created	date (dd-mm-yyyy)	The date of record creation for the building_address relationship.	No	Yes	DT_CREATE
date_modified	date (dd-mm-yyyy)	The most recent date that an attribute has been modified for the building_address relationship.	No	No	DT_MOD
relationship_confidence	number (3)	The percentage confidence that a building has a relationship with an address record.	No	Yes	REL_CONF
is_residential	character string (3)	Value of "Yes" allocated where address information has identified the land use as residential.	No	No	IS_RESI

4.5 Domain Values

Table 6: state

Domain Value	Description
ACT	The data is located within the Australian Capital Territory.
NSW	The data is located within the state of New South Wales.
NT	The data is located within the Northern Territory.
OT	The data is located within the Other Territories classification. Other Territories covers the external Australian territories of Cocos (Keeling) Islands, Christmas Island, Jervis Bay and Norfolk Island.
QLD	The data is located within the state of Queensland.
SA	The data is located within the state of South Australia.
TAS	The data is located within the state of Tasmania.
VIC	The data is located within the state of Victoria.
WA	The data is located within the state of Western Australia.

Table 7: primary_roof_material

Domain Value	Description
Flat Concrete	The roof is primarily made from concrete.
Fiberglass/Plastic	The roof is primarily made from fiberglass or plastic.
Metal	The roof is primarily made from metal.
Tile	The roof is primarily made from tile.
<NULL>	The primary roof material is not assigned.

Table 8: roof_type

Domain Value	Description
Flat	The general shape of the roof is flat.
Moderate pitch or complexity	The general shape of the roof has a moderate pitch or complexity.
Steep pitch or high complexity	The general shape of the roof has a steep pitch or high complexity.
<NULL>	The general shape of the roof is not assigned.

Table 9: roof_shape

Do-main Value	Description
Flat	The primary roofing system is constructed as a single horizontal plane.
Gabled	The primary roofing system is constructed as two rectangular angled planes supporting each other along the central ridgeline.
Hipped	The primary roofing system is constructed as multiple angled planes meeting at a central ridgeline oriented along the longest centreline of the building. The larger roof planes are typically trapezoidal in shape while the roof planes of the shorter ends are usually triangular in shape.
Mansar	The primary roof system is constructed as multiple planes with part of the planes on an angle and a flat section at the top of the roof.
Mixed	The roof shape is both hipped and gabled
Other	The roof type is not one of the above types.
Shed	The primary roofing system is constructed as a single plane on a slope.
<NULL>	The building has not been assessed for roof shape.

4.6 Data Maintenance

Attribution, data quality and potential capture timelines will vary across Australia based on three categories. Each category has been developed based on several factors defined by the population distribution (categorised based on population size), industrial/commercial activities, the probability of natural events (e.g. flooding) and the image source.

- Urban (satellite source) - areas with a population greater than 200, or with significant industrial/commercial activity in a visual assessment, digitised from satellite imagery
- Urban (aerial source) - areas with a population greater than 200, or with significant industrial/commercial activity in a visual assessment, digitised from aerial imagery
- Rural – all other areas

The age of source imagery used for Buildings also varies across each of the categories to enable national coverage. 'Urban' areas have been captured with as recent as possible source imagery, where possible within 18 months of the building feature creation date. 'Rural' capture was undertaken across an image archive that spans a large timeframe, with the earliest source imagery being captured in 2002. The original age of imagery for Rural capture is shown below. Some limited update has been undertaken across Rural areas since the original capture.

4.6.1 Update Frequency

Updates to Buildings are applied continuously and released on a quarterly schedule

4.6.2 Update Scope

Buildings updates occurs for all existing objects with changed geometry, attributes and/or metadata, as well as data for new objects supplied prior to the release time period.

Updates to the product include:

1. **Feature level building change management:**
 1. Addition of newly captured buildings,
 2. Retiring of non-identified buildings,
 3. Validation of existing building geometry,
 4. Updates to building geometry to improve real-world representation, and
 5. Updates to building attribution: height, roof material, roof type.
2. The inclusion of any new captures of buildings received from third-party partners for inclusion within Buildings.
3. All Buildings relationships to other Geoscape products (G-NAF, Cadastre, Property, ABS Boundaries and Localities) to account for any changes in either product.
4. Corrections and/or improvements to production processes in generating Buildings.

4.6.3 Update Rules

The update process describes rules that are applied to records to determine persistency. A record can be updated, retired or created.

The following table outlines the required attributes needed to be changed to cause a record to be retired.

Table 10: Attributes used for Persistency.

Table Name	Attributes Used for Persistency
building	building_pid
building_address	building_pid, address_pid
building_property	building_pid, property_pid
building_cadastre	building_pid, cadastre_pid

4.6.4 Building Change Management

When updating an area of Buildings new imagery is captured to review for building changes, new buildings and demolished buildings. Automated processes are applied to extract buildings from the new capture imagery.

Newly captured buildings are compared against existing features in Buildings to determine if a footprint is new, an existing footprint no longer exists, or an existing footprint can be better represented with confidence to increase fidelity and accuracy. The logic for several change management scenarios is outlined in the sections below.

The following applies to all change management scenarios:

- Where a new Buildings record is created, new related aspatial linkage records are also created if relationships to Cadastre, Property or Address exist.

- Where an existing Buildings record is retired, all existing related aspatial linkage records are also retired.

4.6.5 Adds, Retires and Updates.

The following rules describe the building change management process used for adding, retiring and updating records in the buildings table. Buildings from higher quality captures will be used in preference over other buildings where there is overlap.

For all buildings in the previous area of interest related to the new capture area, an overlap is performed between the previous buildings and the new capture of buildings. The rules for this are as follows:

- Any buildings that have no overlap with the new capture of buildings are retired.
- Any buildings that match against multiple buildings in the new capture are retired.
- Any buildings that have an overlap of less than 20% with a new capture building are retired and the new capture building is added.

For all buildings in the new capture, an overlap is performed with the previous buildings related to the new capture area. The rules for this are as follows:

- Any buildings that overlap multiple previous buildings will not result in any new or retired buildings.
- Any buildings that overlap a previous building by less than 20% will be added.
- Any buildings that overlap a previous building by between 90% and 99% will be treated as a minor update for that existing building. The previous building will be retired, and the new building polygon will be added with the same building_pid as the previous representation.
- Any buildings that overlap a previous building by between 20% and 90% will be treated as a major update for that existing building. The previous building will be retired, and the new building polygon will be added with the same building_pid as the previous representation.
- Any building that has 100% overlap with a previous building will be treated as an update. In this scenario the building polygon will remain the same as the previous representation, with the same building_pid.

4.7 Data Quality

The quality of Buildings data is assessed by averaging the results of measures applied to samples from the full dataset. Therefore, any figures set out below are only indicative of the quality of the Buildings data.

4.7.1 Positional Accuracy

Positional accuracy is an assessment of the closeness of the location of the spatial objects in relation to their true positions on the earth's surface. Positional accuracy consists of 2 assessments:

- Horizontal accuracy
- Vertical accuracy

The horizontal and vertical positional accuracy is the assessed accuracy after all transformations have been carried out.

Horizontal Accuracy

The horizontal accuracy of Buildings data reflects the positional accuracy of source sensors utilised in the data collection, and the reliability of feature classification and associated orthogonalisation processes.

The horizontal accuracy of imagery used for the extraction of urban buildings ranges from:

- +/- 3 pixels RMSE for aerial imagery. The building_source attribute refers to the imagery resolution used for the capture of the building.
- +/-2.5m Circular Error 90% (CE90) for satellite imagery.

The horizontal accuracy of imagery used for the extraction of rural buildings ranges from:

- +/- 3 pixels RMSE for aerial imagery. The building_source attribute refers to the imagery resolution used for the capture of the building.
- +/-10.9 Circular Error 90% (CE90) for satellite imagery.

The horizontal accuracy of rural buildings is known to be less than 10.9m CE90 and has been measured across various rural locations with CE90 accuracies between 1.8m and 7m.

Influences on horizontal accuracy

The positional accuracy of aerial buildings is yet to be benchmarked as the sample size isn't large enough to provide a definite result. As the number of aerial derived buildings grows we'll assess the accuracies and provide further information.

The accuracy of the aerial imagery used in the buildings product is measured at ground level. Across the imagery there is occurrences of building lean which can influence a building's position. In these locations elevation models are utilised to adjust the building feature to its base.

The positional accuracy of the vertices of unobstructed building features will reflect the accuracy of the source imagery from which it is extracted. Obstructed vertices will have their position estimated with building shapes orthogonalised using trained algorithms with some operator assistance. Users should note that anomalies from the building extraction algorithm may cause erroneous capture and further reduce the positional accuracy of the vertices of building features.

4.7.2 Vertical Accuracy

Source elevation accuracy is dependent on the reference data used for the assignment of height and elevation attributes. Heights are derived either from satellite derived DSM or aerial derived stereo digitisation.

Source elevation data used for the derivation of building height attributes have absolute spatial accuracies described below:

- Absolute vertical (LE90) accuracy: ranging from 1m (aerial) to 2m (satellite).
- Relative vertical (LE90) accuracy: ranging from 1m (aerial) to 2m (satellite).

Multiple factors can impact the quality of the assigned elevation or height, these include but are not limited to:

- Age of source imagery: Where any imagery used within the production of the DSM is older than the date of construction of a building then the heights attributed to that building are likely to be erroneous.
- Correct classification of the feature: Where a building is not correctly defined (i.e. the highest point is not within the representation) then the height assigned to the feature has an increased likelihood of being erroneous.
- The omission of the feature: Where a building is not captured it cannot be assigned a height.

- Obscured building: Where a building is obscured by a tree or other feature then there is an increased likelihood of erroneous height values being assigned despite processes being run to
- Tree coverage surrounding a building: Where a building is surrounded by trees then the algorithm to calculate the roof height may struggle to obtain a representative ground elevation value. In these circumstances, there is an increased likelihood of an erroneous height assignment.
- The off-nadir angle of source imagery: Where imagery used for the classification of buildings is off-nadir the side of a building may be represented within the boundary of the footprint. Intersecting this part of the building against the DSM will return lower elevation values than those expected for the roof of the building. Where this occurs, there is an increased likelihood of an erroneous value being assigned to the eave height. The likelihood and impact of this issue are increased relative to the height of a building.

4.7.3 Thematic Quality

Thematic accuracy is defined as the accuracy of quantitative attributes, the correctness of non-quantitative attributes, and of the classification of features and their relationships. Data within the category Urban (aerial source) is targeting higher quality levels than data from the Urban (satellite source) category.

Classification Correctness

Classification correctness is an assessment of the reliability of values assigned to features in the dataset in relation to their true 'real world' values.

Building Solar Panel

The rate of classification correctness of the solar_panel attribute has been measured at above 85% where a photovoltaic solar panel is visible within the source imagery. Where a photovoltaic solar panel is not visible in the source imagery it is expected that the rate of classification correctness (in the negative) is greater than 90%. Other types of solar panels (e.g. solar hot water) may be captured and included in error.

Building Roof Type and Roof Material

The primary_roof_material attribute is spectrally classified from imagery with pixel sizes greater than the imagery used to extract the building polygon feature. Pixels that intersect the boundary of a building polygon may return values from ground surfaces or neighbouring buildings impacting the results obtained. The primary material attributed to a building is represented by a minimum number of intersecting pixels which comprise at least 10% of the total roof area of the building polygon.

The rate of classification correctness of the roof_material attribute has been measured in excess of 70% across suburban residential areas. Commercial, industrial, and higher density residential areas are known to have lower classification correctness of the roof material.

Characteristics of a building that can impact the roof_type classification correctness of the roof shape include but are not limited to:

- Roof furniture e.g. plant rooms, steeples, aerials.
- Buildings that have a larger footprint than roof area.
- Multiple buildings represented as a single building due to proximity.
- The height of the building.
- Overhanging trees.

The rate of classification correctness of the `roof_type` has been measured at above 90% on unobstructed buildings.

Swimming Pool

The rate of classification correctness of a swimming pool has been measured at above 85% where a swimming pool is present.

Zoning

The `planning_zone` attribute is classified by mapping State, Territory or Local Government planning zone scheme information against high-level, generalised national zoning codes developed by Geoscape Australia. The mapping of the planning schemes to the national zoning codes is a generalised process and is not based on a detailed examination of each scheme. The classification does not take into consideration standard planning overlays, multi-zoned areas or amendments which may be made to planning schemes from time to time which may not be reflected in the base zoning datasets.

The generalised national zoning codes attempt to reflect the general intention of the individual planning schemes. However, they do not reflect differences in planning legislation or its interpretation by State, Territory and Local Government planning authorities.

The `planning_zone` attribute provides only a general indication of the planning zone for a building and should not be treated as authoritative. It is therefore not suitable for purposes such as planning and development decision making. Users should contact the relevant State, Territory or Local Government planning authority for authoritative planning scheme information.

The source of data for the allocation of zones to buildings has large areas where no zoning information exists. This is predominately in rural areas for the State/Territory of NT, WA and north-western NSW.

4.7.4 Logical Consistency

Logical consistency is a measure of the degree to which data complies to a technical specification. The test procedures are a mixture of software scripts and manual visual analysis. The data structure of Buildings has been tested for conformance to the data model. The following have been tested and confirmed to conform:

- File names
- Attribute names
- Attribute lengths
- Attribute types
- Attribute domains
- Attribute order in the file
- Object type
- Compulsory attributes populated

4.7.5 Topological Consistency

Topological consistency is the measure of how features spatially relate to other features within and across the Buildings theme. Topological inconsistencies are identified using a combination of automated rules. Where topological inconsistencies are identified, they are notified back to the supplier for remediation. Some minor topological inconsistencies are corrected during product processing. The level of topological consistency is dependent on the data supplied to Geoscape. Where topological inconsistencies continue to exist after supplier remediation they are identified within the product through the population of the geometry_quality attribute.

4.7.6 Temporal Accuracy

Temporal accuracy is an assessment of both temporal consistency (how well-ordered lifecycle events are) and temporal validity (validity of data with respect to time). Building polygons are attributed with the capture date of the imagery from which the original outline has been captured for the building pid.

4.7.7 Completeness

Completeness is an assessment of the extent and range of the dataset with regard to completeness of coverage, and completeness of attribution.

Attribute Completeness

The layer within the Buildings has a full population of attributes in accordance with the data model.

Feature Completeness

Building features are considered to have an omission rate below 5% in urban areas, with the majority of omission being smaller buildings such as garden sheds. In rural areas, buildings with an area of less than 100m² are considered to have an omission rate of less than 10% and buildings with an area of greater than 100m² are considered to have an omission rate below 5%.

4.7.8 Geometry Validity

Building geometry is validated to ensure polygons are a valid representation and free of self-intersection. Issues being detected and resolved include spikes, bow ties, duplicate vertices, null geometries, multi-part geometries, and self-contacts.

Overlaps are present in the buildings data and flagged in the geometry_quality attribute. Overlapping polygons can occur due to the differing horizontal accuracy of source imagery as well as small overlaps occurring between features extracted from the same source imagery.

Polygon orientation conforms to the following specifications:

- OGC Simple Feature Access Specification v1.2.1 [Section - 6.1.11.1]
- The GeoJSON Specification RFC7946 [Section 3.1.6 dot point 4]

This means the polygon outer boundary will be counter clockwise and the inner boundary will be clockwise for file formats that support the above standards.

4.7.9 Further Comments

Buildings has been processed to assure all polygons are stored as single part features to improve compatibility with a range of software applications.

4.8 Extent/Geographic Description

The spatial coverage includes Australia's land mass and surrounding offshore islands. Buildings does not currently include data for other territories of Christmas Island, Cocos (Keeling) Islands and Norfolk Island.

The Bounding Box for this data is as follows:

- North bounding latitude: -9°
- South bounding latitude: -44°
- East bounding longitude: 160°
- West bounding longitude: 100°

A detailed description of the coverage for each State and Territory is provided in the table below.

State	Specific Area	Coverage
ACT		Complete coverage
NSW		Complete coverage
NT		Complete coverage
OT	Christmas and Cocos (Keeling) Islands	No coverage
	Jervis Bay	Complete coverage
	Norfolk Island	No coverage
QLD		Complete coverage Additional coverage of coastal sea areas
SA		Complete coverage
TAS		Complete coverage
VIC		Complete coverage.
WA		Complete coverage.

4.9 Spatial Reference System

4.9.1 GDA94

Horizontal Datum: The Geocentric Datum of Australia 1994 (GDA94) is the target horizontal datum.

Coordinate System: Geographic Coordinate System Geocentric Datum of Australia 1994 (GDA94).

4.9.2 GDA2020

Horizontal Datum: The Geocentric Datum of Australia 2020 (GDA2020) is the target horizontal datum.

Coordinate System: Geographic Coordinate System Geocentric Datum of Australia 2020 (GDA2020).

4.10 Delivery Format

Buildings is provided at a National and State/Territory level, depending on the file format selected. The data is made available in the File Geodatabase, GeoJSON, ESRI Shapefile and MapInfo TAB formats described below.

Format	National	State/Territory
File Geodatabase	Yes	Yes
GeoJSON	Yes	Yes
ESRI Shapefile	.	Yes
MapInfo TAB	.	Yes

4.10.1 File Geodatabase

Format name

File Geodatabase – ESRI™

Specification

This format includes files with the following extensions: *.gdb

ESRI File Geodatabase Technical Description. Follow this link: <http://desktop.arcgis.com/en/desktop/latest/manage-data/administer-file-gdbs/file-geodatabases.htm>

<http://desktop.arcgis.com/en/desktop/latest/manage-data/administer-file-gdbs/file-geodatabases.htm>

Language

English

4.10.2 GeoJSON

Format name

GeoJSON

Specification

This format includes files with the following extensions: *.geojson

GeoJSON specification: <https://tools.ietf.org/html/rfc7946>

Language

English

Note: The GeoJSON specification states that the coordinate reference system for all GeoJSON coordinates is:

“a geographic coordinate reference system, using the World Geodetic System 1984 (WGS 84) datum, with longitude and latitude units of decimal degrees”

Buildings will be provided with coordinates using the datum selected for download (GDA94/GDA2020) with longitude and latitude units of decimal degrees.

4.10.3 JSON

Format name

JSON

Specification

This format includes files with the following extensions: *.json
JSON specification: <https://www.json.org/json-en.html>

Language

English

4.10.4 ESRI Shapefile

Format name

Shape – ESRI™

Specification

This format includes files with the following extensions: *.shp, *.shx, *.dbf
ESRI Shapefile Technical Description, an ESRI White Paper, July 1998. Follow this link:
www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Language

English

4.10.5 MapInfo TAB

Format name

TAB – MapInfo Professional™

Specification

This format includes files with the following extensions: *.tab, *.dat, *.id, *.map
The MapInfo TAB format is a popular geospatial vector data format for geographic information systems software. It is developed and regulated by MapInfo as a proprietary format. TAB files support geospatial standards such as Open GIS, the OGC, ISO, W3C and others.

Language

English

4.11 Product Versioning

Versioning is managed through incrementing when there is a change to the product schema or a significant change in data population, these are described further below:

- A schema change can affect a major or minor increment to the versioning. Additive changes (changes that won't break customers' ability to work with the data) will be incremented with a minor version increment, an example is the addition of a new attribute. Removal of attributes or changing the structure of the schema will enact a major change to identify that this requires the attention of all customers and partners.
- Where a significant geography of Australia either has a new population of data for an attribute or is populated from a much higher quality source a minor increment will be applied to the product version.

Therefore, versioning will not increment with every data update. Published releases will have a name e.g. 'May 2021' and will reference a version of the product e.g. '1.0'.