Natural Resources Canada | Ressources naturelles Canada

Automatically Extracted Buildings

Product Specifications

Edition 2.1

2023-01-05

Government of Canada
Natural Resources Canada

Client Services
Telephone: +01-819-564-4857 / 1-800-661-2638 (Canada and USA)
Fax: +01-819-564-5698
E-mail: geoinfo@canada.ca
URL: http://open.canada.ca/en/open-maps

Canada

Copyright notice

© His Majesty the King in Right of Canada, as represented by the Minister of Natural Resources.
All rights reserved.

RELEASES HISTORY

Date	Version	Description
2018-11-30	1.0	Original version.
2021-11-01	2.0	Version 2.0.
2023-01-05	2.1	Minor modifications

TABLE OF CONTENTS

1 OVERVIEW

1.1 TITLE

Automatically Extracted Buildings: Product Specifications

1.2 REFERENCE DATE

2018-11-30

1.3 RESPONSIBLE PARTY

Natural Resources Canada
Strategic Policy and Innovation Sector
Canada Centre for Mapping and Earth Observation
Client Services

Telephone: +01-819-564-4857 / Toll free: 1-800-661-2638 (Canada and USA)
Fax: +01-819-564-5698
E-mail: geoinfo@canada.ca
URL: http://open.canada.ca/en/open-maps

1.4 LANGUAGE

Languages in which the product specifications are available according to ISO 639-2 standard:

eng – English
fra - French

1.5 TERMS AND DEFINITIONS

CSRS	Canadian Spatial Reference System
EPSG	European Petroleum Survey Group
ISO	International Organization for Standardization
NAD	North American Datum
NRCan	Natural Resources Canada
URL	Uniform Resource Locator
WGS	World Geodetic System

1.6 INFORMAL DESCRIPTION OF THE DATA PRODUCT

“Automatically Extracted Buildings” is a raw digital product in vector format created by NRCan. It consists of a single topographical feature class that delineates polygonal building footprints automatically extracted from airborne Lidar data, high-resolution optical imagery or other sources.

The product is available at no cost on the Open Government of Canada website:
http://open.canada.ca/en/open-maps

There are two methods to access the data:

- Prepackaged files (static files): These prepackaged files are published when changes occur. They are available in Shapefile (ESRI™) and File Geodatabase (ESRI™) formats.
- Customized area (dynamic files): The dynamic geospatial data extraction tool allows user to define a customized area. They are available in GeoPackage (OGC), Shapefile (ESRI™) and File Geodatabase (ESRI™) formats.

2 SPECIFICATION SCOPE

2.1 SCOPE IDENTIFICATION

Main

2.2 LEVEL

Dataset

2.3 LEVEL NAME

Main scope of “Automatically Extracted buildings”

2.4 EXTENT

2.4.1 Description

The production will take several years and will provide buildings footprints for part of Canada.

2.4.2 Vertical extent

The data is two-dimensional. The minimum and maximum elevations at the base of buildings when provided in attributes are expressed in meters according to the Canadian Geodetic Vertical Datum of 2013 (CGVD2013).

2.4.3 Horizontal extent

2.4.3.1 West bounding longitude

-141.0

2.4.3.2 East bounding longitude

-52.0

2.4.3.3 South bounding latitude

+41.0

2.4.3.4 North bounding latitude

+84.0

2.4.4 Temporal extent

2.4.4.1 Beginning date

2006

2.4.4.2 Ending date

Today

3 DATA PRODUCT IDENTIFICATION

3.1 TITLE

Automatically Extracted Buildings

3.2 ABSTRACT

“Automatically Extracted Buildings” is a raw digital product in vector format created by NRCan.

It consists of a single topographical feature class that delineates polygonal building footprints automatically extracted from airborne Lidar data, high-resolution optical imagery or other sources.

3.3 OBJECTIVE

This product tries to have an accurate and current representation of building footprints in Canada. NRCan also wants to provide a range of raw products derived automatically from airborne Lidar data, high-resolution optical imagery or other sources.

3.4 TOPIC CATEGORY

Main topics for the product, as defined by the ISO 19115 standard:

Structure

3.5 SPATIAL REPRESENTATION TYPE

Type of spatial representation for the product as defined by the ISO 19115 standard:

Vector

3.6 SPATIAL RESOLUTION

Not applicable.

3.7 GEOGRAPHIC DESCRIPTION

3.7.1 Authority

3.7.1.1 Title

ISO 3166-1:1997 Codes for the representation of names of countries and their subdivisions – Part 1 Country codes

3.7.1.2 Date

1997-10-01

3.7.1.3 Date type

Type of date according to ISO 19115 standard:

Publication

3.7.2 Code

Code of the geographical region covered by the product according to the ISO 3166-1 standard:

CA - Canada

3.7.3 Code Type

Type of code of the delimitation polygon of the extent according to the ISO 19115 standard:

Inclusion (polygon delineation is inclusive)

3.8 REFERENCE TO SPECIFICATION SCOPE

Main

4 DATA CONTENT AND STRUCTURE

4.1 DESCRIPTION

This product is composed of a single topographic feature class defined by a name, a definition, a list of attributes and a geometric representation.

4.2 DATA MODELLING SCHEMA

4.2.1 Application schema

The application schema of the product has not been modeled considering the few quantity of features.

The feature catalogue contains the pertinent information about the data structure and the content of the product. The geometric representation of the spatial components is in conformance with the ISO 19107 standard document: Geographic Information – Spatial schema that defines the geometric primitives.

4.2.2 Feature catalogue

The complete description of the Feature Catalogue can be found on the Open Government of Canada website (http://open.canada.ca/en/open-maps).

4.3 REFERENCE TO SPECIFICATION SCOPE

Main

5 REFERENCE SYSTEMS

5.1 SPATIAL REFERENCE SYSTEM

Spatial data are expressed in geographic coordinates of latitude (φ) and longitude (λ) according to the North American Datum of 1983 in Canadian Spatial Reference System (NAD83CSRS - EPSG:4617). The longitude is expressed with a negative number to represent a position to the west of the central meridian (0°).

Note: Data is also available according to the WGS84/Pseudo-Mercator (EPSG:3857) and NAD83/Canada Atlas Lambert (EPSG:3979) projections while issued from the dynamic geospatial data extraction tool.

5.1.1 Authority

5.1.1.1 Title

EPSG Geodetic Parameter Registry

URL: http://www.epsg-registry.org

5.1.1.2 Date

2018-10-06

5.1.1.3 Date type code

Publication

5.1.1.4 Responsible party

OGP - International Association of Oil and Gas Producers

URL: http://www.epsg.org

5.1.2 Code

4617

Note: Data is also available according to the WGS84/Pseudo-Mercator (EPSG:3857) and NAD83/Canada Atlas Lambert (EPSG:3979) projections while issued from the dynamic geospatial data extraction tool.

5.1.3 Code space

EPSG - European Petroleum Survey Group

5.1.4 Version

9.5.4

5.2 REFERENCE TO SPECIFICATION SCOPE

Main

6 DATA QUALITY

6.1 COMPLETENESS

The datasets are visually inspected to verify content against the data source used for production or an independent source. No detected error is corrected. No data quality assessment (including completeness and thematic accuracy) is done during data production. Data comes from airborne Lidar, high-resolution optical imaging or other sources.

Airborne Lidar

Airborne Lidar-derived datasets are produced from an NRCan automatic process. This process creates vector building footprints from the building class (6) of the lidar files. Optionally, the process can automatically classify the buildings (6) and ground (2) classes of lidar files, whether they originally have these classes or not.

The quality of the building footprints varies significantly from one acquisition project to another depending on several influencing factors:

- The point cloud density of lidar files.
- The availability of a rigorously corrected and validated building class (6).
- The presence of vegetation obstructing the buildings.
- The proximity of the buildings to each other. A very dense urban scene can lead to the merging of some buildings during the creation of the product.
- The complexity of the buildings. A complex building may be misrepresented as a result of the squaring step of the footprint.
- The type of roof covering. Some roof covering absorb the lidar signal resulting in no return to the sensor.
- The presence of artifacts in the lidar files.

Proulx-Bourque et al. (2021) demonstrated that the NRCan extraction process detected over 96% of the buildings with fewer than 5% commission error in an airborne Lidar dataset that had a prior classification of buildings. However, the variability of the influencing factors can lead to significant differences for other acquisition projects.

Proulx-Bourque JS., McGrath H., Bergeron D., Fortin C. (2021) Extraction of Building Footprints from LiDAR: An Assessment of Classification and Point Density Requirements. In: Singhroy V. (eds) Advances in Remote Sensing for Infrastructure Monitoring. Springer Remote Sensing/Photogrammetry. Springer, Cham. https://doi.org/10.1007/978-3-030-59109-0_11

High-resolution optical imaging

Not applicable

Other sources

Not applicable

6.2 LOGICAL CONSISTENCY

6.2.1 Conceptual Consistency

Not applicable

6.2.2 Domain Consistency

The domain of values included in the feature catalogue is validated in the source database containing the product. This approach ensures the domain consistency between the Feature catalogue and the product.

6.2.3 Format Consistency

The use of well-established commercial software to generate distribution formats ensures format consistency for product distribution.

6.2.4 Topological Consistency

Topological relationships of features are validated in the source database for each dataset. This approach ensures topological consistency between the feature catalogue and the product.

6.3 POSITIONAL ACCURACY

6.3.1 Absolute or external accuracy

Each dataset feature occurrence has a horizontal positional accuracy attribute that indicates the accuracy of the occurrence expressed as the Circular Map Accuracy Standard (CMAS) if possible.

Each dataset feature occurrence has also a vertical positional accuracy attribute that indicates the accuracy of the minimum and maximum elevation attributes expressed in metres as a linear map accuracy standard (LMAS) if possible.

Otherwise, these accuracies are estimated according to the data source and the operating mode.

Generally, products generated from airborne LiDAR have an accuracy better than 1 metre. The horizontal positional accuracy is estimated at 2 metres to also consider the difference caused by geometry processing.

For a dataset coming from the dynamic geospatial data extraction tool, the overall dataset horizontal positional accuracy is shown in its associated metadata file. This accuracy is expressed by an interval made of the minimum and maximum accuracy values. This interval is determined from the accuracy of each source file used to produce the dataset and not from each feature occurrence accuracy value contained in the dataset. Since data sources generally cover a larger area than the resulting dataset, it is highly possible that the overall accuracy of a dataset does not correspond to the smallest and/or the greatest accuracy value attached to each instance of the dataset.

The same method also applies for validity dates written in the metadata file.

6.3.2 Relative or internal accuracy

Unknown

6.4 TEMPORAL ACCURACY

6.4.1 Accuracy of a Time Measurement

Not applicable

6.4.2 Temporal Consistency

Not applicable

6.4.3 Temporal Validity

Not applicable

6.5 THEMATIC ACCURACY

6.5.1 Classification Correctness

Not applicable

6.5.2 Non Quantitative Attribute Correctness

Not applicable

6.5.3 Quantitative Attribute Accuracy

The accuracy of minimum and maximum elevations at the base of buildings is expressed by the vertical positional accuracy described in "6.3.1 Absolute or External Precision".

The accuracy of the minimum and maximum heights and the area of the buildings are unknown.

6.6 REFERENCE TO SPECIFICATION SCOPE

Main

7 DATA CAPTURE

7.1 DESCRIPTION

The "Automatically Extracted Buildings" dataset is derived from airborne Lidar data, high resolution optical imagery or other sources. Entities in the dataset are building footprints that correspond to the roofs. A building footprint can be as much a detached house as a semi-detached or row house. Building footprints are extracted using a method adapted to the source, rectified and automatically attributed. A minimum area of 10 square metres is used to eliminate buildings considered too small. An additional attribute allows the identification of buildings footprints with a potential incoherence. After some basic validations of the format and not of the content, the dataset is translated into different vector formats.

Datasets generated from Lidar data use the building and ground classification points. We first extract the geometric boundary that contains all building points for each building according to a grouping tolerance. Then we add the following attributes: estimated quality of the source data, the minimum and maximum heights and elevations. The courtyards are not delimited in the current version.

Additional information on the calculation method of the attributes:

- The minimum and maximum elevation values are respectively the lowest and highest lidar point elevation among all lidar points classified as ground found in a 2.5 meter buffer around the building footprint. There is no averaging or other statistical method applied. See section 2.4.2 Vertical extent for more information on the height reference system of the attributes.

- The minimum and maximum height values are the height above ground level of the lowest and highest lidar point classified as Building within the building footprint, respectively. Note that the minimum and maximum height do not necessarily represent the base of the walls and the top of the roof. Depending on the context, these points can be a gallery, a chimney, a tree branch, or any other element included in the building footprint.

- The estimated quality of the building footprints is set using a multi-criteria grid that contains 3 properties of the lidar point cloud used for the footprints extraction, namely (1) the presence of class 6 (building) points from a semi-automatic classification, (2) the absence of leaf during most of the data acquisition period and (3) the overall nominal density of lidar pulses (pts/m²). Depending on the combination of these criteria, the attribute value corresponds to one of the 5 estimated building footprints quality levels: Excellent, Good, Fair, Poor, Very poor. These quality levels indicate the ability of the building footprint extraction process to extract a good geometric representation while minimizing commission errors.

		Point cloud properties		
		Presence of class 6 (building) from a semi-automatic classification	Mostly leaf-off	Density (pts/m²)
Estimated footprints quality levels according to the source data	**Excellent**	**x**	**x**	**>= 12**
	Good	**x**	**x**	**< 12**
	Fair		**x**	**> 2**
	Fair	**x**		**> 2**
	Poor			**> = 4**
	Very poor			**< 4**

7.2 TO THE SPECIFICATION SCOPE

Main

8 DATA MAINTENANCE

8.1 DESCRIPTION

Data maintenance activities are aligned around a number of government issues, including emergency management and floodplain mapping. These issues may lead to partnerships with the provinces and territories to release existing Lidar data or to participate in new acquisitions.

Partial coverage of Canada's territory is phased in gradually. Datasets are processed and made available as they are acquired. The source data of the product is acquired by multiple projects of different partners.

A product update is done when changes are made to the data following the reception of new data sources. Only data sources with a good potential for building footprint extraction are produced.

The updates frequency is varied and depends on the availability of new data sources.

8.2 REFERENCE TO SPECIFICATION SCOPE

Main

9 DATA PRODUCT DELIVERY

9.1 DELIVERY FORMAT INFORMATION: GEOPACKAGE

9.1.1 Format Name

OGC® GeoPackage

9.1.2 Version

1.2

9.1.3 Specification

GeoPackage - 1.2, OGC® GeoPackage Encoding Standard, *2017-08-25*, http://www.geopackage.org/spec120/index.html

Note: GeoPackage format is only available from the dynamic geospatial data extraction tool (see Section 1.6)

9.1.4 Language

eng – English

fra - French

9.1.5 Character Set

UTF8

9.2 DELIVERY FORMAT INFORMATION: SHAPE

9.2.1 Format Name

Shapefile - ESRI™

9.2.2 Version

July 1998

9.2.3 Specification

ESRI Shapefile Technical Description, an ESRI White Paper, July 1998,
(http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf)

9.2.4 Language

eng - English

fra – French

9.3 DELIVERY FORMAT INFORMATION: FGDB

9.3.1 Format Name

File Geodatabase - ESRI™

9.3.2 Version

Unknown (Outside the public domain)

9.3.3 Specification

Not available. This format was launched with the ArcGIS (ESRI™) software, version 9.2.

9.3.4 Language

eng - English

fra – French

9.4 DELIVERY MEDIUM INFORMATION FOR STATIC FILES

9.4.1 Units of delivery

The files are compressed and delivered by province and project:

Example for the FGDB format: Autobuilding_MB_Cooks_Creek_fgdb.zip

Example for SHAPEFILE format: Autobuilding_MB_Cooks_Creek_shp.zip

9.4.2 Transfer Size

The file size changes according to the selected projects.

9.4.3 Medium Name

Open Government of Canada website (http://open.canada.ca/en/open-maps)

9.4.4 Other Delivery Information

Information regarding the use of the data is defined in the Open Government Licence - Canada (http://open.canada.ca/en/open-government-licence-canada).

9.5 DELIVERY MEDIUM INFORMATION FOR DYNAMIC FILES

9.5.1 Units of delivery

Depend on the user preference.

9.5.2 Transfer size

Variable

9.5.3 Medium Name

Open Government of Canada website (http://open.canada.ca/en/open-maps)

9.6 REFERENCE TO SPECIFICATION SCOPE

Main

10 METADATA

Not applicable