

Data Inventory & Quality Report → **Analytical Report**

Diagnostic Analysis: Canada Structures field completeness

A census of 13,760,754 structure polygons in the 13 held files

Diagnostic

Descriptive

Prepared for Arbutus Visual Analytics

Prepared by Arbutus Visual Analytics

Date September 16, 2026

Confidentiality Public portfolio piece. Open Government of Canada product. No client.

EXECUTIVE SUMMARY

WHAT WE SET OUT TO ANSWER

Which fields in the Canada Structures files can be used in every province, and which patterns sit in the file itself?

WHAT WE USED

We counted every polygon in the 13 held GeoPackages and the product specification. No rows were dropped. Height zero and Height missing were kept as separate classes.

WHAT WE FOUND

- Every Nunavut polygon stores Height as 100% zero (all 11,068 polygons).
- Height is missing on 1,887,979 polygons. Ontario and Quebec hold almost all of them.
- OpenStreetMap type is blank on 11,344,842 polygons. Land cover name is blank on 13,215,202.
- 20,180 overlapping pairs remain. Stored Area disagrees with geometry by more than 1 percent on 79.4% of polygons.

WHAT FOLLOWS FROM IT

Treat Height, type, and land cover as provincial fields until the product publishes remaining zeros, missing values, and blanks.

NUNAVUT HEIGHT RECORDED AS ZERO

100%

Share of Nunavut polygons whose stored Height is zero, not missing.

[↓ jump to finding](#)

POLYGONS WITH HEIGHT MISSING

1,887,979

Almost all sit in Ontario and Quebec.

[↓ jump to finding](#)

POLYGONS WITH OSM TYPE BLANK

11,344,842

Type is blank on most polygons in every province.

[↓ jump to finding](#)

POLYGONS WITH LAND-COVER NAME BLANK

13,215,202

Land-cover name is blank on almost every polygon.

[↓ jump to finding](#)

INTERIOR-OVERLAPPING PAIRS REMAINING

20,180

Quebec and Ontario hold the largest pair counts.

[↓ jump to finding](#)

STORED AREA OFF GEOMETRY BY MORE THAN 1 PERCENT

79.4%

Typical gap is 2.9 percent. None exceed 10 percent.

[↓ jump to finding](#)

Every Nunavut polygon stores Height as zero. Yukon and Saskatchewan are next.



Share of each file whose stored Height is recorded as zero, not missing. The highlighted bar is Nunavut.

ENGAGEMENT OVERVIEW

Diagnostic Analysis: Canada Structures field completeness

13 province and territory files, counted as a census of the receipt

1 PROBLEM

The question

- Q Which stored fields mean the same thing in every province?
- ⚠ Where does Height fail: recorded zero, or missing entirely?
- 🔍 What patterns sit in area, shape, and source flags without leaving the file?



2 WORKFLOW

What we did

- 📂 Read all 13 GeoPackages. National total 13,760,754 polygons. No rows dropped.
- ☑ Separated Height zero from Height missing. Checked identifiers, source flags, and type labels.
- 📊 Measured how area, perimeter, and Height associate on polygons with a measured Height.
- 🎯 Counted remaining interior overlaps and the gap between stored Area and geometry Area.



3 SOLUTION

What we found

- ⚠ Height cannot be used nationally. Nunavut is all zeros. Missing Height sits in Ontario and Quebec.
- 🔍 OpenStreetMap type is blank on most polygons. Land cover names are blank too.
- On polygons with measured Height, footprint area and Height barely associate.

Source: Canada Structures GeoPackages, Public Safety Canada, held assembly (specification reference date 2024-04-30).

Conclusions

Written for a non-technical audience.

WHAT WE SET OUT TO ANSWER

Which stored fields can a national user quote from these 13 files, and which fields only hold in some provinces?

WHAT WE FOUND

- Height cannot be used nationally. Every Nunavut polygon stores Height as zero. Height is missing on 1,887,979 polygons, almost all in Ontario and Quebec.
- OpenStreetMap type is blank on 11,344,842 polygons. Land cover name is blank on 13,215,202. The Open Database of Buildings flag is missing on Manitoba, Newfoundland and Labrador, Nunavut, Prince Edward Island, and Yukon.
- The identifier CS_ID is unique on every row. Province matches the filename. Yukon is stored as YK. Those identifier checks pass.
- On the 9,156,321 polygons with a measured Height, Area and Height barely move together (rank association 0.26). Compactness across all 13,760,754 polygons is typically 0.73.
- Geometry is valid. 20,180 overlapping pairs remain. For a typical polygon among all 13,760,754 polygons, stored Area exceeds geometry Area by 2.9 percent of geometry Area. 10,929,105 of those polygons (79.4 percent of all polygons) sit above a one percent gap.

WHAT WE RECOMMEND

1. Quote the field trust table before you publish a national Height, type, or land cover number from this product. The table shows which fields fail in some provinces.
2. Keep Height zero and Height missing as separate classes in any later copy of these files. Ontario's low zero share is a missing-value problem.
3. Treat stored Height as mean lidar height inside the footprint. Architectural tip is a different quantity. The polygon named CN Tower that holds the national maximum stores 270.435 m.

1. Files, population, and how Height is stored

The object of study is the 13 held Canada Structures GeoPackages, not the stock of real buildings in Canada.

One row is one structure polygon. The population is every polygon in the 13 receipted GeoPackages: 13,760,754 polygons. Inclusion is all rows. No polygon was dropped for missing Height, blank type, tiny Area, or invalid geometry. The time window is this held assembly (specification reference date 2024-04-30). Province-file tables are labelled aggregations of that population, not a second population. Yukon is stored as YK.

The files are an open Government of Canada product (Public Safety Canada, Canada Structures, specification version 1.0.0, 2025-07-31). Area and Perimeter are the producer's calculations in EPSG 102002, in square metres and metres. Height is metres on the Canadian Geodetic Vertical Datum of 2013. The specification states that Height is the mean of Digital Height Model values inside the polygon, attributed only inside a population centre with High Resolution Digital Elevation Model coverage. That is why a named tower can store a Height far below its architectural tip.

Two polygons are named CN Tower. The taller named polygon stores Height 270.435 m on 1791 square metres (CS_ID 11549690). The other

stores Height 54.0 m. The national Height maximum is that taller polygon, 270.435 m, not a clipped antenna.

CS_ID is unique on all 13,760,754 rows. The Province code matches the filename on every row (0 mismatches). Geometry types are Polygon on 13,758,911 rows and MultiPolygon on 1,843. Invalid geometries: 0.

Roles used in this audit: CS_ID is a row key, not a measurement. Province and source_file are grouping fields. OSM, ODB, and MSB are source flags stored as 1 or empty. OSM_Type and LC_Name are stored labels, with blank as a recorded gap. Area, Perimeter, and Height stay continuous. Height null stays null. Height zero stays a recorded zero class, not a missing-value recode. There is no fitted model, so categorical fields were not one-hot encoded.

Typical footprint Area is 129.1 square metres (median). Typical measured Height, on polygons with Height greater than 0, is 5.2 metres (median). Typical compactness ($4 \pi \text{ Area} / \text{Perimeter squared}$) is 0.73.

HEIGHT IS A MEAN, NOT A TIP

The CN Tower polygon that holds the national maximum stores 270.435 m. That is the mean lidar height inside a 1791 square-metre footprint, not the architectural height of the tower.

2. Height zero and Height missing are different problems

A national Height statistic mixes three classes: missing, recorded zero, and a measured value above zero.

Height is recorded as zero on 2,716,454 polygons (19.7% of all polygons). Height is missing on 1,887,979 polygons (13.7% of all polygons). A measured Height above zero is present on 9,156,321 polygons.

Every Nunavut polygon stores Height as zero: 11,068 of 11,068. Yukon is 94.3% zero. Saskatchewan is 76.4% zero. Ontario's zero share is only 1.2%, because Ontario's problem is missing Height, not zero.

Of the 1,887,979 polygons with Height missing, Ontario holds 1,239,128 and Quebec holds 648,848. New Brunswick holds 3. The other ten files hold none.

The rank order of Height-zero share by province is not the rank order of Height-missing share. A user who only looks at zeros in Ontario will miss more than a million missing values.

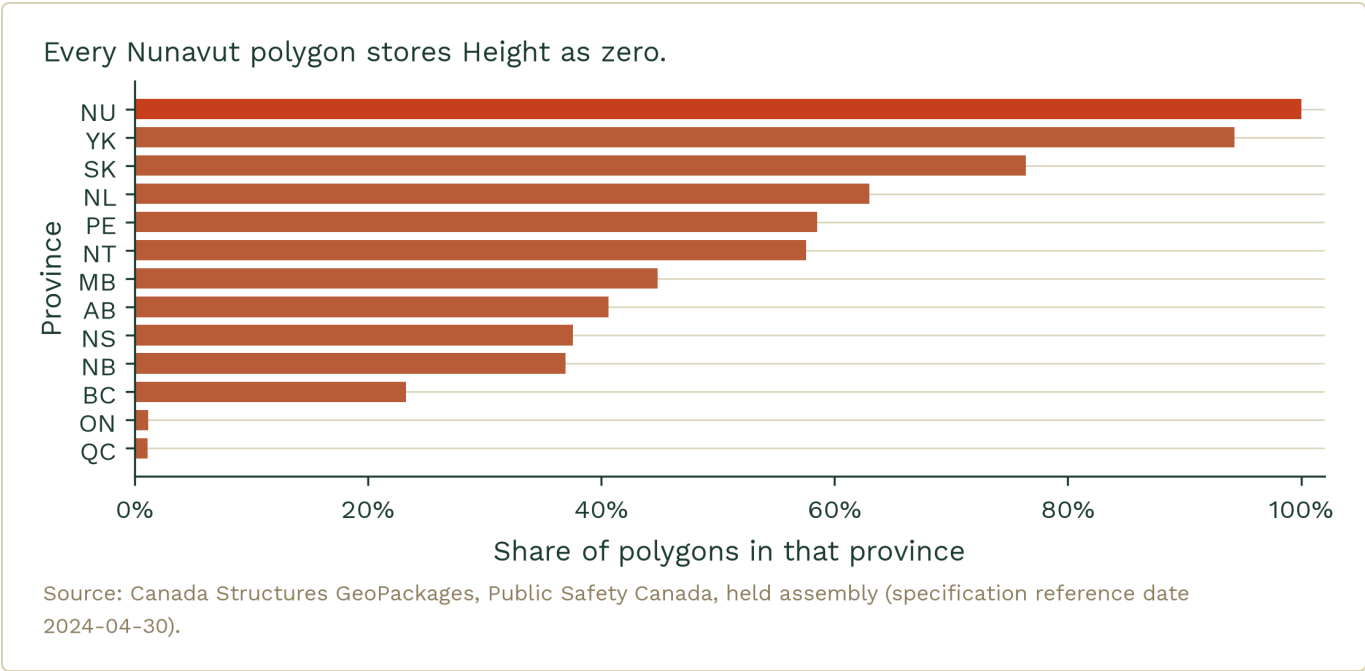
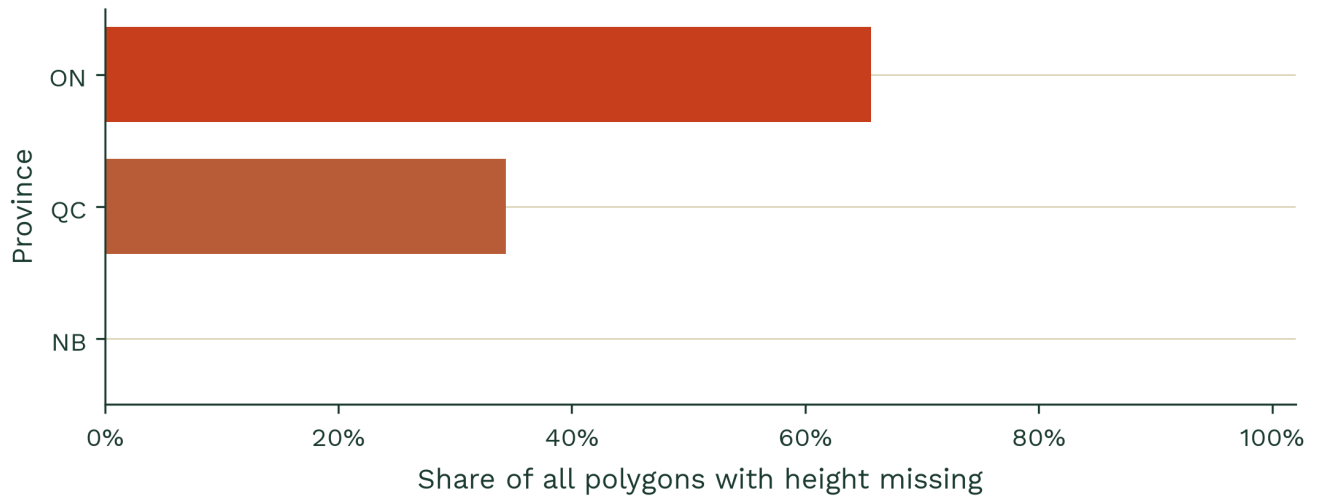


Figure 1. Each bar is that province's polygons whose stored Height is zero. Nunavut is the full bar.

Ontario and Quebec hold almost every missing Height.



Source: Canada Structures GeoPackages, Public Safety Canada, held assembly (specification reference date 2024-04-30).

Figure 2. Ontario and Quebec hold almost every missing Height value. The bars are that national missing mass.

Province	Polygons	Height zero	Zero share	Height missing	Missing share	Total Area (m2)
AB	1,910,493	775,944	40.6%	0	0.0%	383,469,092
BC	1,535,416	357,294	23.3%	0	0.0%	349,561,198
MB	721,810	323,760	44.9%	0	0.0%	126,615,935
NB	403,613	149,160	37.0%	3	0.0%	67,362,169
NL	288,267	181,599	63.0%	0	0.0%	39,392,358
NS	507,673	190,693	37.6%	0	0.0%	76,899,027
NT	19,616	11,294	57.6%	0	0.0%	3,723,146
NU	11,068	11,068	100.0%	0	0.0%	2,401,724
ON	4,732,798	56,927	1.2%	1,239,128	26.2%	1,033,243,002
PE	81,686	47,786	58.5%	0	0.0%	15,910,251
QC	2,794,337	31,977	1.1%	648,848	23.2%	538,177,131
SK	737,953	563,841	76.4%	0	0.0%	126,173,811
YK	16,024	15,111	94.3%	0	0.0%	3,116,750

Table 1. Polygon counts and Height classes by province.

3. Source flags can be used; type and land-cover names cannot

OSM, ODB, and MSB flags are populated. The type and land-cover text fields are not.

OpenStreetMap contributed a flag on 5,931,825 polygons, the Open Database of Buildings on 4,439,708, and Microsoft Building Footprints on 6,785,721. A polygon may carry more than one flag. 3,396,500 polygons carry two or three sources.

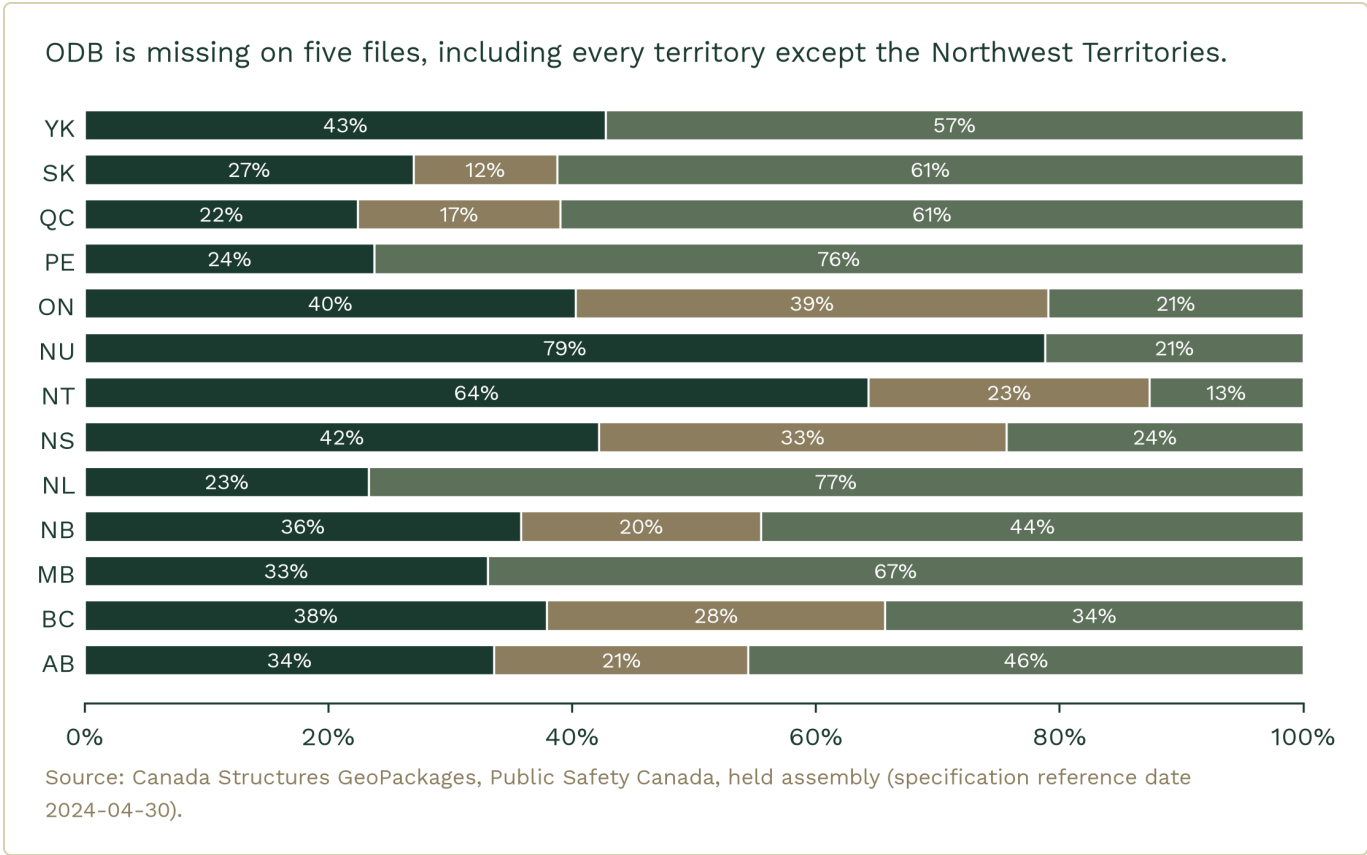


Figure 3. OpenStreetMap, Open Database of Buildings, and Microsoft flags by province. Five files have no Open Database of Buildings flag.

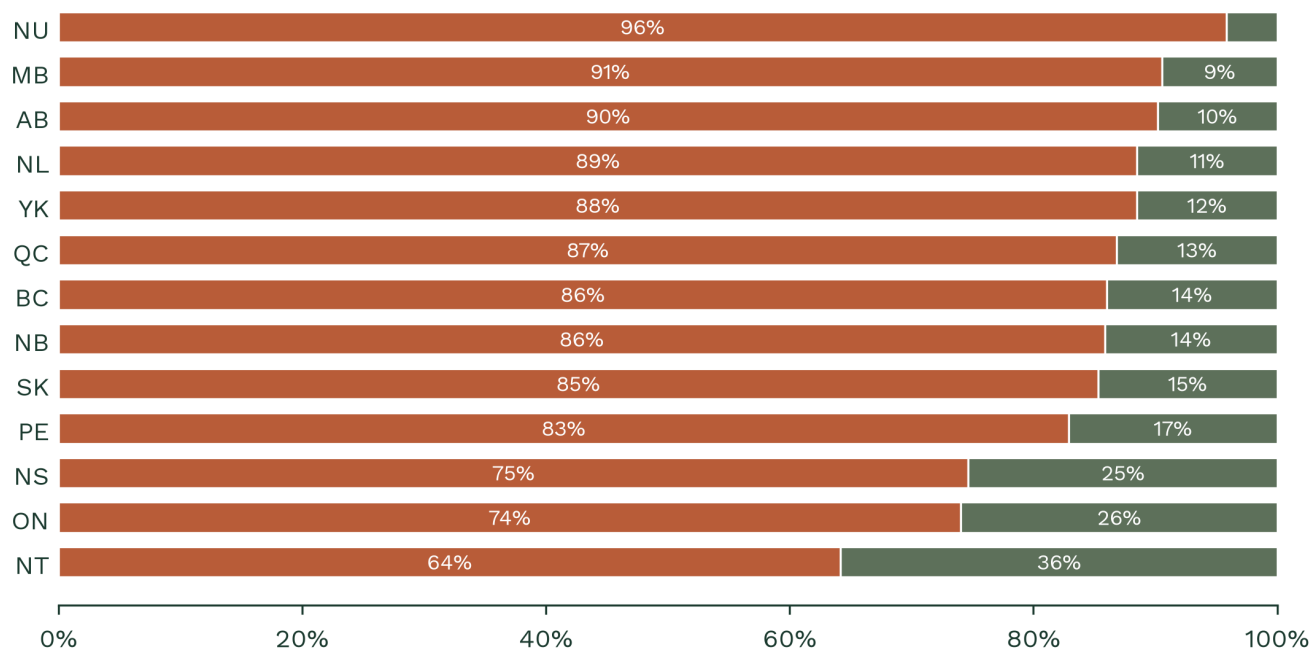
The Open Database of Buildings flag is absent on five files: MB, NL, NU, PE, YK. A national pie of source mix would hide that hole.

OSM type is blank on 11,344,842 polygons. Land-cover name is blank on 13,215,202 polygons. 8,579 remaining land-cover values are placeholder question marks, concentrated in Nunavut.

Where a type is present, 2,028,558 are dwelling-like labels, 205,344 ancillary, and 182,010 other. Nunavut ancillary types number 9, below a reporting floor of 30, so that cell is named rather than charted.

Land-cover name without an OSM land-cover flag occurs on 545,552 polygons. The two fields do not line up as a single land-cover system.

OpenStreetMap type is blank on most polygons in every province.



Source: Canada Structures GeoPackages, Public Safety Canada, held assembly (specification reference date 2024-04-30).

Figure 4. OpenStreetMap type is blank on most polygons in every province.

Province	OSM	ODB	MSB	Type blank	Land-cover blank
AB	41.8%	25.9%	56.7%	90.2%	91.0%
BC	49.6%	36.3%	44.8%	86.0%	95.9%
MB	43.8%	0.0%	88.5%	90.5%	98.1%
NB	43.2%	23.7%	53.6%	85.9%	98.3%
NL	26.9%	0.0%	88.6%	88.5%	96.0%
NS	57.4%	45.5%	33.1%	74.7%	96.7%
NT	83.0%	29.7%	16.3%	64.2%	94.7%
NU	97.0%	0.0%	26.0%	95.9%	60.1%
ON	54.2%	52.1%	28.2%	74.1%	96.0%
PE	29.5%	0.0%	94.4%	82.9%	99.4%
QC	23.7%	17.6%	64.4%	86.9%	98.1%
SK	30.5%	13.3%	69.1%	85.3%	98.6%
YK	54.0%	0.0%	72.2%	88.5%	79.6%

Table 2. Source-flag rates and blank type and land-cover rates by province.

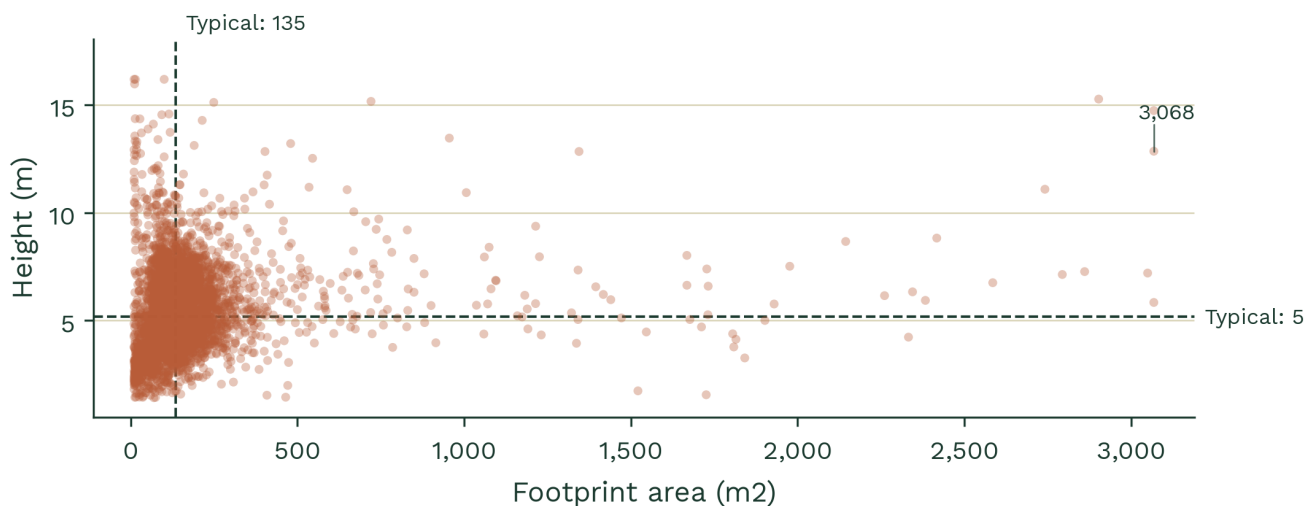
4. Area and Height barely associate once Height is measured

The association questions use only polygons with Height greater than 0, because a zero or a missing value is not a measured height.

On 9,156,321 polygons with a measured Height, the association is 0.07 (Pearson). the rank association between footprint Area and Height is 0.26 (Spearman). The linear

On polygons with measured Height, area and Height barely associate.

Showing 4,000 of 9,156,321 pairs. Medians from all. 178,262 pairs sit outside this view (largest 2,788,966, 270).



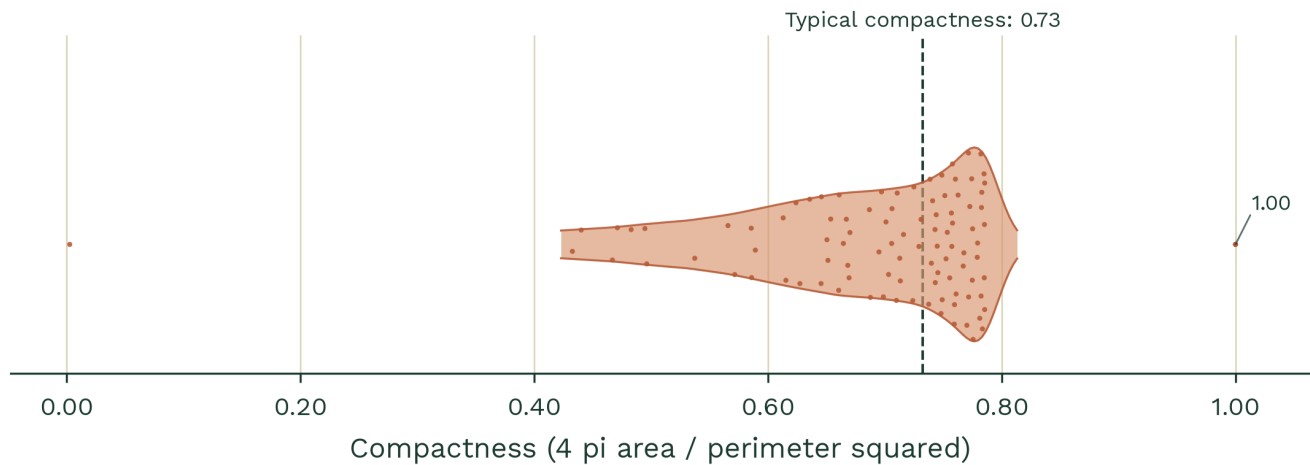
Source: Canada Structures GeoPackages, Public Safety Canada, held assembly (specification reference date 2024-04-30).

Figure 5. Footprint Area against Height, on polygons with a measured Height. The view shows the bulk of the cloud.

Ontario alone is 0.15. Alberta alone is 0.42. Dropping the far tail of Area, using the same three-standard-deviation cut on the logarithm of Area that the robustness check uses, leaves 0.24.

Compactness has median 0.73. 1,137 polygons sit below 0.1. None exceed 1. Multi-source polygons have a slightly higher typical measured Height (5.61 m) than single-source polygons (4.97 m). That is a shift, not a cause.

Typical compactness is 0.73; 1,137 polygons sit below 0.1 and none exceed 1.
13,760,754 values; 100 plotted as points. Median and extremes from all.



Source: Canada Structures GeoPackages, Public Safety Canada, held assembly (specification reference date 2024-04-30).

Figure 6. Compactness of every polygon. The typical value is 0.73. The low tail is a review list of elongated or jagged shapes.

Area and Perimeter move together. Height does not move with Area once Height is measured.

Area	1.00	0.98	0.26	-0.45
Perimeter	0.98	1.00	0.26	-0.58
Height	0.26	0.26	1.00	-0.17
Compactness	-0.45	-0.58	-0.17	1.00
	Area	Perimeter	Height	Compactness

Source: Canada Structures GeoPackages, Public Safety Canada, held assembly (specification reference date 2024-04-30).

Figure 7. Rank associations among Area, Perimeter, Height, and compactness, on polygons with measured Height.

No pair of size and Height measures is strongly rank-associated. Area and Perimeter move together because both describe the same outline. Height does not ride that outline.

5. Geometry holds; stored Area is systematically a little high

Invalid geometries are absent. Remaining interior overlaps and a stored-versus-computed Area gap are the geometry findings.

Invalid geometries: 0. Interior-overlapping footprint pairs remaining: 20,180. Quebec and Ontario hold the largest pair counts.

Height-zero share is highest in the territories and Saskatchewan.



Source: Canada Structures GeoPackages, Public Safety Canada, held assembly (specification reference date 2024-04-30).

Figure 8. Polygons with Height recorded as zero, mapped on province envelopes. Nunavut, Yukon, and Saskatchewan read darkest.

Stored Area sits above planar geometry Area. Counting each polygon equally, the typical relative gap is 2.9%. The unweighted median of the 13 provincial medians is 2.6% (Nunavut and Ontario count equally). 10,929,105 polygons (79.4% of all polygons) disagree by more than 1 percent. 0 disagree by more than 10 percent. The 10 percent cut is a review flag, not a drop rule. Compactness used the stored Area and Perimeter pair, which is the producer pair.

Review flags from low compactness or Area below 10 square metres, plus the 10 percent Area gap (which is zero rows), total 75,685. An in-file proxy for the specification's tile-edge Height warning, Height relative to the square root of Area among positive Heights, flags 104,926 polygons. There is no tile identifier in the receipt, so that proxy cannot name a tile.

Province	Overlap pairs	Median relative Area gap	Gap above 1 percent	Invalid
AB	2,985	2.6%	1,746,262	0
BC	2,570	0.2%	303,344	0
MB	858	0.7%	163,139	0
NB	171	2.5%	373,669	0
NL	9	1.1%	164,432	0
NS	1,329	3.6%	507,673	0
NT	11	5.9%	19,592	0
NU	17	5.7%	10,963	0
ON	5,464	4.7%	4,509,838	0
PE	23	2.3%	81,686	0
QC	5,753	2.7%	2,408,236	0
SK	900	1.8%	624,247	0
YK	90	5.7%	16,024	0

Table 3. Interior-overlap pairs and stored-versus-geometry Area gaps by province.

Field	Nationally usable	Note
CS_ID	yes	Unique on all 13,760,754 rows.
Province	yes	One code on each file; Yukon stored as YK.
OSM / ODB / MSB flags	yes, with holes	ODB is absent on MB, NL, NU, PE, YK.
OSM_Type	no	11,344,842 blank (82.4% of polygons).

Field	Nationally usable	Note
LC_Name	no	13,215,202 blank; remaining Nunavut values include literal question marks.
Area	with limits	Typical stored Area 129.1 m2. 10,929,105 polygons (79.4% of all polygons) have stored Area more than 1 percent away from geometry Area. Typical per-polygon relative gap 2.9%. Unweighted median of the 13 provincial medians 2.6%.
Perimeter	with limits	Producer EPSG 102002 metres; compactness uses this pair.
Height	no	Nunavut is 100% zero. Ontario has 1,239,128 nulls. Values are mean lidar height inside the footprint, not architectural tip.

Table 4. Which fields can be used nationally.

The field-trust table is the verdict for a national user of the held files. Quebec and Ontario contribute the most remaining overlap pairs; those two files also hold most of the polygons.



Figure 9. Downtown Toronto footprints around the tower named CN Tower, highlighted in red. Stored Height is mean lidar height inside the footprint.

The downtown sample shows the same mean-height field at street scale.

Limitations and Caveats

These limits apply to the work.

Census of the held files. Completeness against Overture, Microsoft Canadian footprints, or the Open Database of Buildings as an external register was not measured.

No producer merge record. The report cannot say that the producer's merge rule, or a quality control pass on Microsoft footprints, caused the Height or Area patterns.

One vintage. There is no date field and no second release, so polygon counts cannot be compared across assemblies.

No census dissemination geography. Density per census dissemination area is outside the receipt.

No height tile identifier. Height errors cannot be attached to named High Resolution Digital Elevation Model tiles. A proxy inside the file compares stored height against stored area.

Height is a mean. The column stores mean lidar height inside the footprint. Architectural tip, antenna, and spire height are different quantities.

No review by a subject expert of the findings was obtained. This is a public case study with no live client review audience.

No live client handoff pack, invoice, or operating plan. Those belong with a paying client. A data inventory still ships because the brief is an assessment of the data.

No stakeholder map, business case fee framing, or sponsor written approval. There is no external sponsor on this piece.

Association is a description. Polygons flagged by more than one source having a slightly higher typical Height is a description. It is not a merge-effect finding.

Abbreviations

AB	Alberta, as stored in the Province field.
BC	British Columbia, as stored in the Province field.
CGVD	Canadian Geodetic Vertical Datum of 2013, the vertical datum named for Height.
CN	The letters CN in the stored name CN Tower.
CS_ID	Canada Structures identifier. Unique on every polygon in the held files.
DHM	Digital Height Model: a lidar surface minus ground. Height is the mean of those values inside each footprint.
EPSG	A public numeric code for a coordinate system.
HRDEM	High Resolution Digital Elevation Model, the height-tile product the specification names as a source of some Height errors.
LC	Land cover. LC_Name is the land-cover label stored on a polygon.
MB	Manitoba, as stored in the Province field.
MSB	Microsoft Building Footprints, one of the three sources named on each structure polygon.
NB	New Brunswick, as stored in the Province field.
NL	Newfoundland and Labrador, as stored in the Province field.
NS	Nova Scotia, as stored in the Province field.
NT	Northwest Territories, as stored in the Province field.
NU	Nunavut, as stored in the Province field.
ODB	Open Database of Buildings, one of the three sources named on each structure polygon.
ON	Ontario, as stored in the Province field.
OSM	OpenStreetMap, one of the three sources named on each structure polygon.
PE	Prince Edward Island, as stored in the Province field.
QC	Quebec, as stored in the Province field.
SK	Saskatchewan, as stored in the Province field.
YK	The Province code stored for Yukon in these files, not YT.

Basis of Preparation

This report was prepared as a public portfolio case study for Arbutus Visual Analytics. It asks which fields in the Canada Structures files can be used nationally, and what patterns sit in the held files.

It is not to be relied on by any third party, or for any purpose other than that stated here, without written agreement from Arbutus Visual Analytics.

The analysis uses the 13 Canada Structures GeoPackages and the product specification as received. Row counts, unique CS_ID, Height classes, source flags, and geometry checks were computed from those files. The files were not independently audited as an operating dataset beyond those checks.

Data as of Held assembly, specification reference date 2024-04-30; specification publication 2025-07-31.

AI Usage Disclosure

This report's production and formatting were completed with AI-assisted tools. All methodology, architecture, and analytical review were completed by the author.

1. This report was prepared by Arbutus Visual Analytics on its own initiative, for the purpose set out in Basis of Preparation. No client commissioned it and no party has been engaged to rely on it.

2. Nothing here is legal, tax, accounting, financial, or investment advice. Before acting on anything with legal or financial consequences, consult a lawyer, a Chartered Professional Accountant, or a licensed investment or financial adviser.

3. The findings rest on the data described in Basis of Preparation, relied on as published and not independently audited.
4. Polygon counts and class totals in this report are a census of the held files, not estimates. Association measures and sampled charts are summaries, not guarantees of a future result.

5. There is no client named on the cover. Decisions taken on the basis of this report remain with the reader.

6. Any map in this report is for information only. It is not a legal survey and determines no boundary.