

Catalogue no. 32-26-0005
ISBN 978-0-660-78062-7

The Open Database of Greenhouses

The Open Database of Greenhouses metadata document: Concepts, methodology and data quality, 2025

Version 2.0



Release date: August 18, 2025



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “Contact us” > “[Standards of service to the public](#).”

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2025

Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

Cette publication est aussi disponible en français.

Table of contents

1 Overview	4
2 Data sources	5
3 Reference period	7
4 Geography	7
5 Target Population	7
6 Compilation methodology	8
7 Data dictionary	10
8 Data accuracy	12
9 Contact Us	13
References	14

The Open Database of Greenhouses metadata document: Concepts, methodology and data quality, 2025

1 Overview

To explore open data for official statistics and support geospatial research across various domains, the Agriculture and Data Science and Innovation Division launched a project utilizing earth observation data to modernize conventional surveys conducted at Statistics Canada. This initiative leverages high-resolution earth observation sources made available as open data by multiple levels of government across Canada.¹ Developed in response to Statistics Canada’s modernization efforts and the [AgZero](#) project, this project integrates advanced methods—including geographic information systems, data integration, and cutting-edge technologies. In alignment with these efforts, the ODG is used to support and train machine learning models aimed at automating the collection of greenhouse information nationwide. This document outlines the process of collecting, processing, and standardizing earth observation data and imagery-derived products for digitized greenhouses, as established in the second version of the ODG. The ODG is made available under the Open Government Licence – Canada.²

In the first version (version 1.0), the ODG contains 2,476 individual records across ten municipalities across four provinces. The current version of the ODG (version 2.0) contains 3,905 manually digitized greenhouses and individual records across 38 municipalities and four provinces, an increase of 58%. The database will be updated periodically as new open datasets become available. The ODG is provided as a geographic shapefile.

In addition to manually digitized greenhouses that provide updated information on greenhouses from the first version, Statistics Canada also included 1,006 greenhouses detected automatically through innovative data science methods within open source satellite imagery.

The machine learning component is comprised of two mechanisms: (i) super resolved Sentinel imagery, and (ii) greenhouse detection. The super resolution process is a model developed to use multiple 10-meter resolution Sentinel-1 and Sentinel-2 images of the same site over time to generate one 1.25 meter “super resolved” image. This super-resolved image of the site is then used by the greenhouse detection model for training, and ultimately greenhouse detection on new super-resolved images. The super-resolved imagery removed the cost barrier to obtaining high-resolution imagery. The greenhouse detection model is a U-net model developed to segment pixels of an image into “not greenhouse” or “greenhouse” categories. The model is trained on imagery and greenhouse labels and learns features that distinguish greenhouses from other objects or background.

The machine learning results are experimental and thus may not represent a fully accurate and complete data of greenhouses on the ground. As with all modelling techniques, limitations exist: (i) The machine learning model was trained on Southern Ontario data and its ability to scale to other regions has not been tested, (ii) The machine learning model performance is affected by the resolution of the imagery and the size of greenhouses, specifically (like with human digitisation) small greenhouses may be too small to detect.

This dataset is released as part of the Linkable Open Data Environment (LODE). The LODE is an initiative that aims at enhancing the use and harmonization of open data from authoritative sources by providing a collection of datasets released under a single license, as well as open-source code to link these datasets together. Access to the LODE datasets and code are available through the Statistics Canada website and can be found at:

[The Linkable Open Data Environment](#)

1. This includes municipal, regional, and provincial.
2. [Open Government Licence - Canada](#)

2 Data sources

Multiple data sources were collected to create the ODG version 2. The data providers of ODG version 1 included multiple levels of government sources whom have made their high-resolution imagery available free, or were provided to Statistics Canada with existing companies who hold a National Standing Offer³ with the Federal Government, including attribution to each of these sources as per the license requirements.

However, Statistics Canada has shifted to different satellite imagery sources for the second version, which relies on free and frequent satellite imagery sources. For the second version of the ODG, manually digitized greenhouses were created using Sentinel-2 for the 2023-2024 time period. Higher resolution imagery within open-source platforms was often used to support accurate identification and digitization of the greenhouse structures, where possible. Sentinel-2 was used to digitize new greenhouses where more updated imagery for our target date was not available within open-source platforms. Sentinel-2 has a Multispectral Imager (MSI) with a spatial resolution of 10 meters, which was leveraged to validate the presence or absence of large greenhouses but is limited in accuracy when attempting to digitize smaller structures.

In addition to the use of medium resolution (10m) open-source imagery for digitization, Statistics Canada has also adapted an innovative approach of generating high resolution imagery (1.25m) through medium resolution imagery sources by using a method called super resolution (Wang, 2022).

The data from Sentinel-1 and Sentinel-2 imagery was downloaded from European Space Agency's (ESA) Copernicus Hub and processed through the super resolution method, and was then integrated into a model for training as well as testing the greenhouse detection model. The imagery was collected for two sites in southern Ontario, Canada: Niagara and Essex.

Details on the data sources are provided in Table 1 below. There are a total of 38 municipalities covered in four provinces.

For further information on the individual licences, users should consult directly with the information provided on the open data portals of the various data providers.

3. A standing offer is an offer from a supplier to Canada that allows Canada to repeatedly purchase goods and/or services, or a combination of goods and services at pre-arranged prices, under set terms and conditions, when required. A standing offer is not a contract and Canada is under no obligation to purchase. A standing offer only becomes a contract once Canada issues a "call-up" against the standing offer. (See: [Federal Contractors Program – Standing Offers and Supply Arrangements – IPG-085](#))

Table 1
Data Sources

ODG Version	Data Sources	Credits	Image(s) Details	License Agreement
1	MDA Geospatial Services	York, Ontario	WorldView3 - 8 Band, TIFF, DRA off, 30cm Pan+1.2m Date: 2020/06/13	The NMSO for MDA is under the contract number W1786-180002/001/ST and standing offer #E60SQ-120001/003/SS
		Laval, Québec	WorldView2 - 8 Band Ortho Tif FTP NN NAD83 UTM 8-bit, DRA off, 40cm Pan+1.6m Date: 2021/05/24	
		St-Eustache, Québec	WorldView2 - 8 Band Ortho Tif, DRA off, 50cm Pan+2m Date: 2020/05/20	
		Medicine Hat, Alberta	WorldView2 - 8 Band Ortho Tif, DRA off, 50cm Pan+2m Date: 2021/05/01	
1	Township of Langley	Township of Langley – GIS.	2017 Orthophotos	Contains information licensed under the Open Government License – Township of Langley. https://www.tol.ca/en/connect/open-data-licence.aspx
1	City of Burnaby	City of Burnaby – GIS.	2020 Orthophotos	Open Government Licence – British Columbia Open Government Licence - City of Burnaby https://data.burnaby.ca/pages/open-government-licence
1	City of Surrey	City of Surrey	2018 Orthophotos	Open Government License – City of Surrey. https://data.surrey.ca/pages/open-government-licence-surrey
1	Ministry of National Resources and Forestry Ontario	Ontario Ministry of Natural Resources and Forestry	Orthophotos	Orthophotography under Licence with the Ontario Ministry of Natural Resources and Forestry© King's Printer for Ontario, 2018 and 2020.
1	City of Chilliwack - GIS	City of Chilliwack – GIS.	2021 Orthophotos	The data is provided as a public service by the City of Chilliwack. Terms of Use - City of Chilliwack https://www.chilliwack.com/main/page.cfm?id=2391
2	European Space Agency	Copernicus Sentinel-2	Multispectral 10 meters resolution. Coverage: At least one imagery is available and cloud-free covering the agricultural areas across Canada. Date: Images available between April and September 2020, 2023, 2024.	ESA Open Access License https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/TermsConditions
2	European Space Agency	Copernicus Sentinel-1 GRD.	Coverage: At least one imagery is available and cloud-free covering the agricultural areas across Canada. Date: Images available between April and September 2020, 2021	ESA Open Access License https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/TermsConditions

3 Reference period

The data is covering years 2023-2024 for the manually digitized greenhouses across Canada. Data sources were accessed, consulted and processed in Spring 2023 and 2024. The satellite imagery used for the super resolution imagery for the machine learned digitized greenhouses is covering 2020-2021 timeframe for Essex and Niagara regions in Ontario.

4 Geography

The ODG geographic frame is referenced by five key regions, notably (1) the Fraser Valley and Greater Vancouver regions in BC; (2) Essex and (3) Niagara, in Southern Ontario ;(4) North of Montreal, Quebec; and (5) Medicine Hat and Edmonton regions in Alberta. These regions and the ODG reference Statistics Canada's census divisions (CD) (Statistics Canada Geographic Boundaries: [2021 Census – Boundary files](#)).

The sites for the machine learning generated greenhouses were Essex and Niagara, Ontario. Sentinel-1 and Sentinel-2 imagery for these sites were downloaded and then used to generate a super-resolved image for each site.

These sources are not solely limited to the Statistics Canada geographic bounds for any one specific area and may extend into neighboring municipalities. The ODG may also not be complete for one specific region listed in Table 1 or limited to the bounds based on Statistics Canadas geographic regions.

Geographic Representation

The Open Database of Greenhouses is available on the Statistics Canada website in the following geographic representation:

- Projection: Lambert conformal conic
- False easting: 6200000.000000
- False northing: 3000000.000000
- Central meridian: -91.866667
- Standard parallel 1: 49.000000
- Standard parallel 2: 77.000000
- Latitude of origin: 63.390675
- Linear unit: metre (1.000000)
- Datum: North American 1983 (NAD83)
- Prime meridian: Greenwich
- Angular unit: degree
- Spheroid: GRS 1980

The North American Datum of 1983 (NAD83) is an adjustment of the 1927 datum (NAD27) that reflects the higher accuracy of geodetic surveying.

5 Target Population

Statistics Canada defines a greenhouse and greenhouse products⁴ as a space for growing seedlings, potted plants, bedding plants, cuttings and other propagating material, vegetables and fruit grown for sale in a permanent, artificially heated enclosed structure made of plastic, plexiglass, poly-film or glass.

4. [Annual Greenhouse, Sod and Nursery Survey \(GSNA\)](#)

As a result of this definition, additional buildings which do not fit into the greenhouse definition, as outlined above, can possibly be included in the dataset based on their common visual characteristics. The database does not include linkages to business information, which would differentiate agricultural versus non-agricultural facilities.

Minimal editing and validation are done to the shape of buildings through methods of digitization and automated machine-learning in instances where validation of buildings captured in the database have similar visual characteristics. The dataset identifies greenhouses without distinguishing their type, the crops grown inside, or any features that could aid in their classification. The database does not include linkages to business information or refer to Statistics Canada surveys, business registers, taxation data or other sources. This ensures the database to maintain an open database component.

6 Compilation methodology

The creation of the ODG version 2.0 comprised of two main processing steps: first, the processing of earth observation data and, second, the creation and formatting of the dataset overlaying the earth observation data and the mapping of the original dataset attributes to standard variable (column) names. The ODG version 2.0 has two separate methods of construction: manual digitization and automated machine learning.

A data dictionary of the variables used is provided in section 7. To compile the data into the final geographic shapefile database, both compilation methods are outlined below:

Digitized

- Earth observation data were extracted, uncompressed and converted to TIF format if not originally in this format once acquired.
- Sentinel-2 imagery was visualized into GIS software using web services, removing the steps of downloading and processing the imagery, and a new geographic feature classes were created for each region of interest.
- Greenhouses visually comparable to known greenhouses were identified in the earth observation, and a new record was digitized within the feature class.
- In the case where a greenhouse building was identified in ODG v 1.0 but not visible in ODG v 2.0 timeframe, the greenhouse polygon(s) were removed.
- Once a region was completed, attribute tables were completed with the correct and up to date information.
- Once all the regions were covered; the feature classes were merged to form the final ODG file.

Super-resolved imagery

- The super resolution model is a neural network with the following architecture
 - ▶ 1x3x3 CGR layer followed by 1x3x3 DenseNet block
 - ▶ 1x1x1 CGR layer followed by multi-head attention layer
 - ▶ 1x1 CG layer followed by 3x3 DenseNet block
 - ▶ 1x1 CG layer followed by upsampling, followed by 1x1 CG layer, followed by 3x3 DenseNet block
 - ▶ 1x1 CGR layer followed by 1x1 convolutional layer and a sigmoid
- The architecture was based on a literature review (Wang et al., 2022, Dong et al., 2016, Kawulok et al., 2020, Fuoli et al., 2021).
- The first step when running the model is to download Sentinel-1 and Sentinel-2 imagery. Ensure common projection, then provide images to super-resolution model to obtain super-resolved images of desired sites.
- Super-resolved images are reprojected to the same projection as manually digitized greenhouse shapefile.

Machine learning Detection

- The detection model is a U-Net architecture neural network for image segmentation, based on the paper by Ronneberger, O. et al., 2015.
- The architecture was modified to use ResNet34 as the encoder portion to improve training stability.
- The model was trained using super-resolved imagery and manually digitized greenhouse labels.
- For the detection, the first step to run the model is to feed super-resolved imagery to the trained greenhouse machine-learning based detection model.
- Obtain shapefile output of greenhouses from the detections.
- For this release false positives were removed by taking the intersection of detected greenhouses and known greenhouse shapefiles (manually digitized).
- In the future, post processing steps will remove these false positives without relying on known greenhouse polygons. For example, with an NDVI layer.
- Lastly, the manually digitized greenhouse shapefile is merged with machine learning detections.

The original data fields were the unique ID and Shape identified automatically by the software. New fields were created to provide information on the imagery data source, method of collection, centroid X and Y location and province. While effort was made to ensure all greenhouses were identified and other building types were not included, some buildings may be misidentified, or greenhouses could have been missed from the source image. Should any such errors be reported, they will be corrected in future versions of the ODG.

The data included in the ODG is due to visual inspection only and is not linked to official databases, surveys, or private sources.

Geocoding and metadata

Records in the ODG v2.0 include year, data source of information, collection method, building centroid-x longitude, building centroid-y latitude, province unique identifier and name. Records do not include further locational information such as address or postal code. Records do not include additional information regarding the greenhouse covering material, crop being grown inside or ownership.

The images in the greenhouse detection model training process are reprojected to the CRS (EPSG:3347) of the labels to ensure proper alignment with the digitized greenhouses.

Data standardization

Due to the different standards adopted in the original data, steps taken to standardize the data may include some errors. The key principles of the methodology used were the avoidance of false positives and of significant alterations to the data. The methodology and limitations of each technique are described below. Simple cleaning techniques, such as removal of whitespace characters and punctuation removal, are omitted from discussion.

For the machine learning methodology, the standardization process involved processing the labels and images to the same projection and ensuring alignment for training the machine learning model.

Comparisons with Greenhouse, Sod and Nursery Survey

The Statistics Canada's Annual Greenhouse Sod and Nursery Survey (GSNA) is a collection of information of greenhouse production, nursery stocks and sod produced in Canada and is frequently used to perform market trend analysis. Since the GSNA does not use information from this data source, nor does the ODG use data from the GSNA, it is unlikely that the information and total area for a province or region are comparable. The data are kept separate from each other to allow the ODG to be published and used by the public through the open data licence.

Removal of duplicates

For the ODG only entries that seemed to be clear duplicates, overlapping greenhouse shapes, were chosen for removal.

7 Data dictionary

This data dictionary below describes the variables of the ODG.

Variable – Record ID

Name

FID

Format

String

Source

Internally generated during data processing

Description

Unique record ID automatically generated during data processing

Variable - Shape

Name

Shape

Format

Geometry

Source

Internally generated during data processing

Description

Geometry automatically generated during data processing.

Variable - Image Date

Name

ImageDate

Format

Long

Source

Provided in imagery source

Description

Year of imagery acquisition

Variable - Province or Territory

Name

PROV_TERR

Format

String

Source

Province or Territory of record

Description

Province or Territory name

Variable - Province Unique Identifier

Name

PRUID

Format

Long

Source

Converted from province code.

Description

Province unique identifier as per the Statistics Canada official identifier.

Variable - Longitude

Name

Longitude

Format

Double

Source

Calculated geometry of building centroid-x of each record in decimal degrees

Description

Longitude.

Variable - Latitude

Name

Latitude

Format

Double

Source

Calculated geometry of building centroid-y of each record in decimal degrees

Description

Latitude.

Variable - Data Source

Name

DataSource

Format

String

Source

Created based on origins of earth observation data

Description

Name of the entity that provided the earth observation data.

Variable - Method

Name

Method

Format

String

Source

Created based on the collection method used to record the greenhouses.

Description

Collection method used to record the greenhouses. Only two collection methods are possible; Digitized or Machine Learned. Digitized applies when manual digitalization has been used to outline the greenhouse building. Machine Learned applied when the machine learning algorithm has been used to automatically detect and create the greenhouse building within the source imagery.

8 Data accuracy

All greenhouses digitized in the ODG version 2 were in reference to the imagery within a certain date range, provided by ESA's satellite imagery. In general, other than processing and digitization of the features in the dataset, the imagery was used as is and can therefore create errors in the final database where features could not be identified correctly in some cases. Given the nature of the data acquisition and creation of the database, there is the possibility of some errors to be found in the final geographic product.

Super Resolution Model

Quality assessment of the super-resolved predictions was done in the following manner:

- Looking at the mean absolute error (MAE) and root mean squared error (RMSE) for each band.
 - ▶ A sliding tolerance threshold was taken for each window for MAE and RMSE, in which for each 100x100 window the minimum MAE and RMSE for a shift by up to 8 pixels was considered.
 - ▶ This sliding window accepts images that are slightly shifted (for example, a 5 pixel to the left constant shift) compared to the ground images.
 - ▶ Allowing such slight shifts greatly favors more detailed images. The model cannot be certain for the shift between the 10 m Sentinel images and the ground truth images is. Without allowing the shift a blurry image without much enhancement is favored over a detailed image that is shifted by a few pixels.

- Comparing the ratio of MAE of the super-resolved images against a baseline consisting of cubic upsampling of one Sentinel-2 image most closely resembling the target image (similar ratio for RMSE). If this ratio is less than one, super-resolution is outperforming the baseline for the corresponding band. The same sliding window threshold was used for both the baseline and super-resolution.
- Visual inspection of the super-resolved results with input images and the high-resolution ground truth images.

Greenhouse Detection Model

There were 1209 greenhouses in the dataset of which 749 were used for training and 460 used for testing. The greenhouse detection model achieved a pixel-wise F1 score of 80% in testing. Recall was 87%, meaning the model consistently identified true greenhouses, and precision was 75%. The object-wise F1⁵ score was higher, at 91%. Most of the error in performance came from false positives. Upon closer inspection, the false positives occurred where errors in the super-resolved imagery appeared to be shapes that were distinct from surrounding land and had similar characteristics to greenhouses. There were also instances of buildings that became washed out in the super-resolved process and appeared closer to the roof of a greenhouse, and vice versa, where greenhouses in the original imagery became washed out in the super-resolved process. These errors in the super-resolved images contributed to errors in the greenhouse detection results.

9 Contact Us

The LODE open databases are modelled on ongoing improvement. To provide information on additions, updates, corrections or omissions, or for more information, please contact us at statcan.lode-ecdo.statcan@statcan.gc.ca. Please include the title of the open database in the subject line of the email.

How to cite this dataset

Statistics Canada. (2025). Open Database of Greenhouses Version 2. <https://www.statcan.gc.ca/en/lode/databases/odg>.

5. The F1 score is a metric used to evaluate the accuracy of a test or model in classification tasks. It is the harmonic mean of precision and recall, providing a single measure that balances both aspects. Precision refers to the proportion of true positive results among all positive predictions, while recall indicates the proportion of true positive results among all actual positives. The F1 score is particularly useful when the distribution of classes is imbalanced, as it ensures that both precision and recall are given equal importance. A higher F1 score signifies better performance in correctly identifying relevant instances without including irrelevant ones.

References

Peijuan Wang, Bulent Bayrem, Elif Sertel (2022). A comprehensive review on deep learning based remote sensing image super-resolution methods. *Earth-Science Reviews*. 232. 104110. 10.1016/j.earscirev.2022.104110.

Chao Dong, Chen Change Loy & XiaoouTang (2016). Accelerating the Super-Resolution Convolutional Neural Network.

Michal Kawulok, Pawel Benecki, Szymon Piechaczek, Krzysztof Hrynczenko, Daniel Kostrzewa & Jakub Nalepa (2020). Deep Learning for Multiple-Image Super-Resolution. *IEEE Geoscience and Remote Sensing Letters*.

Dario Fuoli, Luc Van Gool & Radu Timofte (2021). Fourier Space Losses for Efficient Perceptual Image Super-Resolution.

Yunhe Li, Yi Wang, Bo Li & ShaohuaWu (2022). Super-Resolution of Remote Sensing Images for 4 Resolution without Reference Images. *Electronics* 2022, 11, 3474.

Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351.