

Exploring Open Data

The Open Database of Greenhouses (ODG)

Metadata document: concepts, methodology and data quality

Version 1.0



Data Exploration and Integration Lab (DEIL)
Centre for Special Business Projects (CSBP)
in partnership with
Agriculture Division (AGRI)

Release date: February 2, 2023



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 STATCAN.infostats-infostats.STATCAN@canada.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 |

Depository Services Program

- | | |
|------------------|----------------|
| • Inquiries line | 1-800-635-7943 |
| • Fax line | 1-800-565-7757 |

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. 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

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2018

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

Cette publication est aussi disponible en français.

Table of Contents

1. OVERVIEW.....	3
2. DATA SOURCES.....	3
3. REFERENCE PERIOD	4
4. GEOGRAPHY	4
5. TARGET POPULATION	6
6. COMPILATION METHODOLOGY	6
GEOCODING	7
DATA STANDARDIZATION	7
<i>Address Parsing.....</i>	<i>Error! Bookmark not defined.</i>
<i>Removal of duplicates.....</i>	7
7. DATA DICTIONARY	8
8. DATA ACCURACY	9
9. CONTACT US.....	9

Acknowledgments

A first version of the database was made possible because of the open data availability, data agreements and partnerships across several municipalities and provinces across Canada. More specifically we would like to acknowledge the Ontario Ministry of Natural Resources and Forestry for their support of our project and ongoing data partnerships with Statistics Canada. Further we would like to acknowledge the City of Surrey, for their open data availability of very high resolution orthophotos.

1. Overview

For the purpose of exploring open data for official statistics and to support geospatial research across various domains, the Agriculture Division and Data Science Division undertook a project to use earth observation data for the modernization of conventional surveys conducted at Statistics Canada with a collection of high-resolution earth observation sources released as open data by various levels of government within Canada¹. This data was created in response to Statistics Canada's modernization initiatives that uses leading-edge methods, data integration and advanced technologies to reduce the response burden on farmers, and as a response the ODG is used to facilitate and train with machine learning models aimed at automatization of collecting greenhouse information across Canada. This is in an effort in to reduce response burden of greenhouse operators in Canada.

This document details the process of collecting, processing, and standardizing the earth observation and the imageries derived product of digitized greenhouses with the first version of *Open Database of Greenhouses* (ODG), which 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 and four provinces. The database is expected to be updated periodically as new open datasets become available. The ODG is provided as a geographic shapefile.

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 licence, 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:

<https://www.statcan.gc.ca/eng/lode>

2. Data sources

Multiple data sources were collected to create the ODG. The data providers include multiple levels of government or were provided 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.

Details on the data sources are provided in Table 1 below. There are a total of 10 municipalities covered in 4 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.

¹ This includes municipal, regional, and provincial.

² [Open Government Licence - Canada | Open Government, Government of Canada](https://open.canada.ca/en/open-government-licence-canada) <https://open.canada.ca/en/open-government-licence-canada>

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: <https://www.canada.ca/en/employment-social-development/programs/laws-regulations/labour/interpretations-policies/standing-offers.html>)

Table 1: Data Sources

Data Sources	Credits	Licences Agreement
MDA Geospatial Services	<u>York, Ontario:</u> WorldView3 - 8 Band, TIFF, DRA off, 30cm Pan+1.2m Date : 2020/06/13 <u>Laval, Québec:</u> WorldView2 - 8 Band Ortho Tif FTP NN NAD83 UTM 8-bit, DRA off, 40cm Pan+1.6m Date: 2021/05/24 <u>St-Eustache, Québec:</u> WorldView2 - 8 Band Ortho Tif, DRA off, 50cm Pan+2m Date: 2020/05/20 <u>Medicine Hat, Alberta:</u> WorldView2 - 8 Band Ortho Tif, DRA off, 50cm Pan+2m Date: 2021/05/01;	The NMSO for MDA is under the contract number W1786-180002/001/ST and standing offer #E60SQ-120001/003/SS
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
City of Burnaby	City of Burnaby – GIS. 2020 Orthophotos.	Open Government Licence – British Columbia https://data.burnaby.ca/pages/open-government-licence
City of Surrey	City of Surrey, 2018 orthophotos	Open Government License – City of Surrey. https://data.surrey.ca/pages/open-government-licence-surrey
Ministry of National Resources and Forestry Ontario	Ontario Ministry of Natural Resources and Forestry	Orthophotography under Licence with the Ontario Ministry of Natural Resources and Forestry© King's Printer for Ontario, 2018 and 2020.
City of Chilliwack - GIS	City of Chilliwack – GIS. 2021 Orthophotos	The data is provided as a public service by the City of Chilliwack. https://www.chilliwack.com/main/page.cfm?id=2391

3. Reference period

The data sources in Table 1 list the most recent date which the source was available at time of access or provided through various other partnerships or agreements and used for digitizing greenhouses or used within the machine learning models for automatization of greenhouses. The data is provided for years 2017-2021 across various sources and locations. Data was accessed or downloaded between the years 2019-2022.

4. Geography

The ODG geographic frame is referenced by regions provided by the open data portals of the cities of Burnaby, Surrey, and Chilliwack, and the Township of Langley as outlined in the Data Sources Section. The named regions listed in Table 1 and the ODG reference Statistics Canada's census agricultural regions (CAR) ([Statistics Canada Geographic Boundaries](#)). These sources are not solely limited to the Statistics Canada geographic bounds for any one specific area and may extend into neighboring municipalities. The product may also not be complete for one specific region listed in Table 1 or limited to the bounds based on Statistics Canada's 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. Any plants that start cultivation in a greenhouse but are finished before sales in a nursery should be considered a nursery product.

Additionally, a nursery and nursery products are defined as a diverse range of non-edible, living plant material grown 'in field' or in containers outdoors and sold with their root system intact. Plants range from tree seedlings to full-grown trees. Include annual and perennial plants.

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.

The database was created by digitizing greenhouses in provided earth observation imagery, with reference to labelled greenhouses in Google Earth Pro. Minimal editing and validation is done to the shape of buildings digitized and validation of buildings captured in the database have similar visual characteristics. Greenhouses identified within the dataset do not discriminate greenhouse type, what is growing inside, and are not labelled different based on any features that could help classify them.

The database does not include linkages to business information or refer to Statistics Canada surveys, business registers, or other sources. This is to enable the database to maintain an open database component.

6. Compilation methodology

The creation of the ODG 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. A data dictionary of the variables used is provided in section 7. To compile the data into the final geographic shapefile database:

- Earth observation data was extracted, uncompressed and converted to TIF format if not originally in this format once acquired.
- Satellite imagery sources were pansharpened from 1.5 meters to 50cm pixel resolution using PCI Geomatica, Pansharpening toolbox and the pansharpening band included in the dataset when acquired.
- Imagery was visualized into GIS software, and a new geographic shapefile was created for each earth observation dataset. Greenhouses visually comparable to known greenhouses were identified in the earth observation, and a new record was created within the shapefile.
- Concatenated geographic datasets were created to represent each dataset used within each municipality where data was acquired.

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, centroid 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.

⁴ Annual Greenhouse, Sod and Nursery Survey - 2020
https://www.statcan.gc.ca/en/statistical-programs/instrument/3416_Q1_V21

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

Geocoding

Records in the ODG v1.0 include latitude, longitude, province, or territory and in some cases municipal information, when applicable. Records do not include further locational information such as address or postal code.

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.

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 productions, 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.

Variable – Province Unique Identifier	
Name	PRUID
Format	Long
Source	Converted from province code.
Description	Province unique identifier.

Variable – Longitude	
Name	Longitude
Format	Double
Source	Calculated geometry of centroid-x of each record in decimal degrees
Description	Longitude.

Variable – Latitude	
Name	Latitude
Format	Double
Source	Calculated geometry of 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.

8. Data accuracy

All greenhouses digitized in the ODG were in reference to the imagery within a certain date range, provided by government, or open-source portals on public webpages. 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.

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.