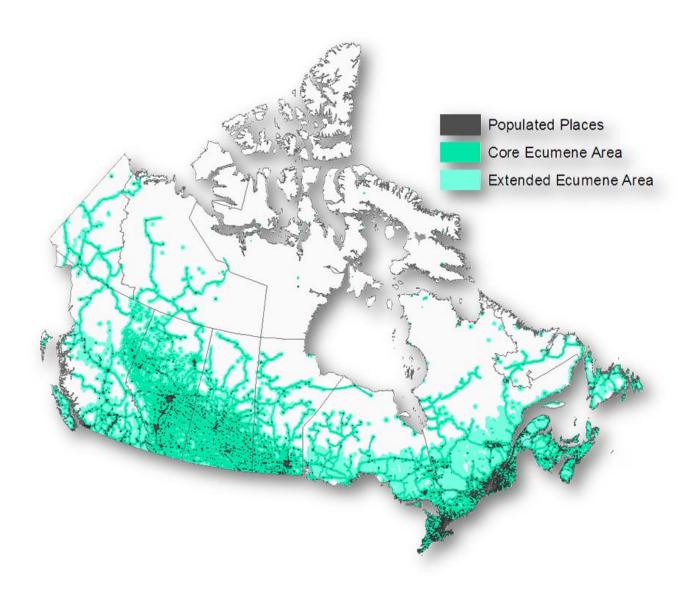
# CanEcumene GIS Database Version 3.0



# **Technical Reference**

October, 2023

© Natural Resources Canada, 2023

## Introduction

By definition, an 'Ecumene' is generally defined as the geographical pattern and extent of human settlement in relation to the biophysical environment (Eddy et. al. 2020). Such a definition implies the natural area of the earth where humans have established permanent or persistent settlement, including infrastructure such as transportation and utilities. Mapping an ecumene may require some flexibility for some users depending on the application.

The purpose of the Canadian Ecumene GIS database is therefore twofold:

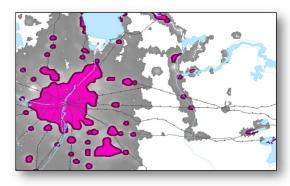
- 1) to provide a set of GIS layers that map the physical extent of human settlement and infrastructure across Canada, and
- 2) to use as an alternative spatial framework for analysis of social, economic, and demographic patterns.

These two aspects serve different users for different purposes. They may be applied separately, or in combination. In the former case, the aim is to provide a set of layers that provide flexibility in how users may define the physical extent of Canada's Ecumene depending on the need. In the latter case, one of the key aims is to provide an alternative to Census Sub-Divisions (CSDs) for mapping and Census data at regional to national scales. This is particularly important for integrating socio-economic data with ecological, natural resources and environmental data. The CanEcumene 3.0 Geodatabase was developed to support both aspects.

This technical reference accompanies the CanEcumene-V3.GDB database provided on the Federal Geospatial Portal (FGP) and OpenMaps Canada, and is supplemental to Eddy et. al. 2020. In following the two main purposes, this document is divided into are two parts: Part A – Populated Places (the Core layers), and Part B – Ecumene Extents (Extended Layers). Figure 1 illustrates the distinction between these two:

Briefly, the Core layers are the populated places (polygons and points) that contain many reference attributes, and demographic, social and economic variables derived from Census data. Part A discusses how the natural boundaries of populated areas were mapped and integrated with official place names and Census data.

The Ecumene Extents are a set of layers used for mapping the spatial extents of settlement areas and infrastructure in various ways. These are supplemental to the Core layers, and can be used in combination with the CanEcumene\_CoreV2.gdb for cartographic or additional analytical purposes. Part B describes these layers in more detail and provides some guidance on their use, either in combination with the Core populated places layers, or on their own.



Part A - Core Layers

- Populated Places
- Boundaries, Names
- Census Data Integration



Part B - Extension Layers

- Population Density
- Transportation / Utilities
- Ecumene Extents

Figure 1. Illustration of the two aspects of the CanEcumene 2.0 database and respective GIS layers: 1) Core Populated Places and 2) Extension Layers.

# Part A - Populated Places

#### Rationale

The most common method used in mapping Census-based socio-economic variables on regional to national scales is the use of the Census Sub-Divisions (CSD) spatial framework. However, there are a number of known limitations with using the CSD framework (see Eddy et. al. 2020). The most significant limitation is that CSDs are defined on the basis of what constitutes a 'municipality' in a given province. However, because municipalities are defined differently in different provinces, the result is that CSDs cannot be directly compared from one province to another on an equivalent basis.

This is problematic from both a cartographic and statistical analysis standpoint. Some provinces (e.g. Quebec – which defines municipalities for very small communities) have a much denser number of CSDs than other provinces (e.g. Ontario – which clusters many communities into regional municipalities). Cartographically, it gives the impression that some provinces have more 'populated places' than other provinces.

This issue can be more serious for statistical analyses. Provinces that have smaller CSD's (e.g. Quebec) generally have smaller population values, whereas provinces that have larger CSDs (e.g. Ontario) tend to have larger population values. Because of this lack of normalization, it is not appropriate to compare statistics among provinces that use ratios or proportions that were calculated based on population values. Other limitations with the CSD framework pertain to the physical location and identification of 'communities' as they are known locally or officially recognized according to the national place names registry (CGDNB, 2018).

The Canadian Ecumene GIS Database was developed to address these issues while retaining the important value of census data. Essentially, it uses a 'natural boundary' approach for identifying populated places as 'Human Habitats' (see Eddy and Dort, 2011, Eddy et. al. 2020) as opposed to administrative or census boundaries. This makes it particularly useful for integration with regional ecological, natural resources and environmental datasets. The following section describes the method used in mapping the locations (points) and boundaries (polygons) of populated areas and integrating Census-based demographic and socio-economic data.

#### Data Integration Method<sup>1</sup>

The method used is based on a triangulation approach using three primary data sources: 1) DMSP Night Lights Imagery (NOAA, 2010), 2) NRCan's Official Populated Place Names database

<sup>&</sup>lt;sup>1</sup> It is important to note that because the CanEcumene database is intended for use at regional to national scales/extents, our aim was to create a nationally consistent geospatial representation of populated places that are not limited by census or administrative boundaries. Whereas many larger populated centres (e.g. large cities) are comprised of multiple CSD's/municipalities, the Ecumene database aggregates most of these areas into larger single populated places (.e.g Greater Montreal, or Greater Toronto Area). The CanEcumene database is therefore not suitable for mapping socio-economic data 'within' larger populated areas.

(GeoGratis, 2018), and 3) Statistics Canada's Census Sub-Division Geographic Boundary File and centroids (StatCan, 2018a,b) (Figures 1). Three steps were taken and described as follows:

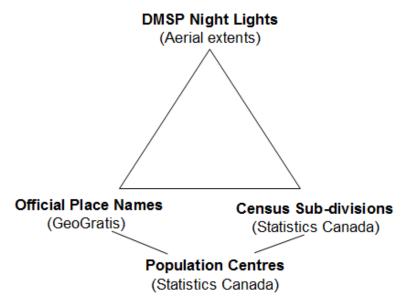


Figure 1. Schema of triangulation method for the development of the Canadian Ecumene GIS Database.

**Step 1. Delineating Populated Areas:** The DMSP Night Lights imagery was extracted for all of Canada, and used to identify approximate boundaries of populated areas. Image radiance values for Canada range from 1-63, and it was determined that values equal to or larger than 30 provided the most suitable delineation of populated areas. Values less than 30 were 'overshadow' or 'halo' effects of populated areas to surrounding areas (Figure 2). The image was reclassified into a binary raster for all values >= 30, and individual cluster areas were vectorized into individual polygons.

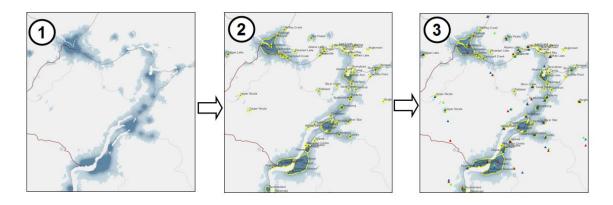


Figure 2. Illustration of the three-step method for aligning and matching populated places from three data sources.

Step 2. Assigning Official Place Names: Official place names from GeoGratis were then overlaid with the DMSP derived polygons and a spatial join was made for all locations where matches were found. Each polygon was assigned the most logical place name based on a detailed inspection and comparison with Atlas of Canada maps, topographic maps, and Google Earth. It is worth noting that not all DMSP polygons correspond directly with official place names, nor did all place names have a corresponding DMSP polygon. For example, some DMSP polygons were identified to be large industrial work areas or remote stations without any official population status (and were therefore excluded). Many official populated places that did not have a corresponding DMSP polygon were identified to be very small communities not large enough to register in the DMSP radiance image. In such cases where communities were found to correspond with official Census CSD locations, they were retained and given a circular buffer with a 3 km diameter.

Step 3. Assigning Census Data Attributes: The third step involved assigning CSD identifiers to the populated places identified in step 2. This was a more intricate task that required detailed inspection at a local scale to determine which CSDs matched the ecumene populated places most appropriately. CSD centroids were extracted from the GeoSuite (StatCan, 2018b) database and were overlaid on ecumene places layer with the aim of performing a 'point-in-polygon' analysis to match the ecumene place ID with corresponding CSD IDs. This process was repeated for each of the 2001, 2006, 2011 and 2016 census periods, as each period can have different CSD ID's and/or centroids locations. While most CSDs aligned well with the ecumene places, there were some cases in which centroids were required to be re-located manually. Figure 3 illustrates how this was done.

The process resulted in either identifying a match (with match) or leaving some CSD centroids unmatched (without match). Three types of matches were made in locations where there was alignment: 1) multiple centroids aligned with a larger multi-contiguous area (MCA), 2) single centroids aligned with a single contiguous area (SCA), or 3) single centroids aligned with a single populated place (SPP).

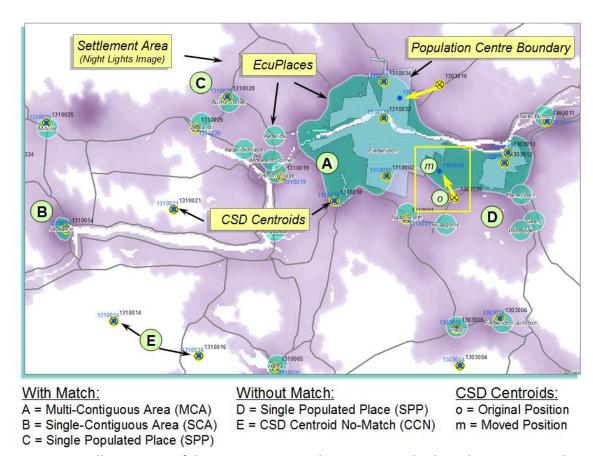


Figure 3. Illustration of how CSD centroids were matched with ecumene place boundaries.

A MCA is regarded as a larger urban centre for which there is a continuous night light signature among multiple, yet physically contiguous, populated areas (e.g. the city of Fredericton and Oromocto in Figure 3) (Case 'A'). MCA's usually have multiple CSDs and in some cases, the positions of nearby centroids were re-located to within the MCA boundary to ensure the point-in-polygon routine worked properly (see 'o' and 'm' in Figure 3). A SCA is a single populated area with a strong enough night light signature to have a boundary derived from the DMSP data (Case 'B'), whereas a SPP are smaller populated places for which there was no night light signature and for which polygons were created by a 3 km circular point buffer (Case 'C'). As Figure 3 also illustrates, not all SPP's have a corresponding CSD match (Case 'D'), nor do all CSD centroids have an ecumene place match (Case 'E'). Such mis-matches have implications for overall statistical representation.

#### Results

The populated places in the ecumene database include all of the original official place names, but not all of these have corresponding CSD matches. Those that do not are most often small reference places with little or no population, or for which census data are aggregated in nearby communities. Although these places were retained in the ecumene database for additional reference in Versions 1.0 and 2.0, they are excluded in Version 3.0. This is primarily due to them not having any associated CSD linkage, and consequently, not linkage to Census data.

CSD centroids for which no matches were identified are considered outliers in that the location of the centroids is usually the centre of a larger CSD that encompasses many small isolated areas not directly associated with any particular populated place. These centroids most often have very small population values, however, collectively they do implicate the overall statistical representation achieved for the integration of the census data and ecumene places.

The representation achieved is summarized in Table 1 by census period. As noted above, each census period has a different total number of CSD's which results in a different number of matches with ecumene places. Differences among the five periods are relatively small as a 99% representation of the total population was achieved for each census period.

In Versions 1.0 and 2.0, attributes were added to the ecumene polygons from a several sources including: 1) ecological zones and regions, 2) ecumene zones, 3) forest zones, and 4) indigenous community indicators. In Version 3.0, StatsCan's Statistical Area Classification (SACType) is added, along with a population category classification (PopCat) for each period. These additional reference variables are included to expand the analytical flexibility of the ecumene data (see end notes for further details).

Table 1. Summary of statistical representation of ecumene places in comparison with Census Sub-Divisions (CSDs). a) Official Census Population Counts (by CSD), b) CanEcumene Population Counts (by Ecumene Places), and c) CanEcumene Population Percent Captured (by Ecumene Places). Note:  $c = (b/a) \times 100$ 

	a. Census P	opulation C	ounts (CSD	)	
ProvTerr	CSDTotPop01	CSDTotPop06	CSDTotPop11	CSDTotPop16	CSDTotPop21
Alberta	2,975,198	3,282,308	3,644,571	4,067,175	4,262,63
British Columbia	3,926,626	4,113,720	4,450,900	4,648,940	5,001,900
Manitoba	1,123,967	1,151,026	1,209,075	1,278,365	1,342,153
New Brunswick	729,442	731,069	752,661	747,101	775,610
Newfoundland and Labrador	528,475	520,515	520,547	522,987	513,729
Northwest Territories	37,238	41,449	41,447	41,786	41,070
Nova Scotia	912,428	914,657	926,044	923,598	969,383
Nunavut	26,660	29,448	31,886	35,944	36,858
Ontario	11,390,433	12,159,005	12,860,440	13,448,499	14,223,957
Prince Edward Island	140,942	138,475	142,645	145,296	155,075
Québec	7,246,175	7,550,038	7,916,793	8,173,994	8,511,448
Saskatchewan	985,500	966,419	1,035,174	1,103,974	1,138,156
Yukon	28,661	30,532	33,999	35,874	40,232
Canada	30,051,745	31,628,661	33,566,182	35,173,533	37,012,206
b. Car	iEcumene P	opulation C	Counts (Ecul	Places)	
DuoisTour	Fourtet DonO1	Four Tot Don Of	ECuTet Den11	Fourtet Don 16	EcuTotPop21
ProvTerr Alberta	EcuTot Pop01	EcuTotPop06	ECuTot Pop11	EcuTot Pop16	
	2,941,105	3,238,845	3,602,600	4,034,582	4,234,528
British Columbia	3,855,466	4,050,671	4,396,508	4,603,083	4,961,325
Manitoba Nava Brancoviale	1,084,140	1,111,221	1,172,102	1,272,799	1,337,464
New Brunswick	700,363	702,557	724,576	732,932	765,946
Newfoundland and Labrador	517,865	510,042	512,342	519,798	510,513
Northwest Territories	37,017	40,794	40,812	41,142	40,456
Nova Scotia	886,193	880,595	900,315	907,025	969,383
Nunavut	26,660	29,448	31,886	35,882	36,858
Ontario	11,315,362	12,078,647	12,775,788	13,419,314	14,191,833
Prince Edward Island	123,997	121,833	125,673	128,697	142,567
Québec	7,172,308	7,513,154	7,875,914	8,153,452	8,478,068
Saskatchewan	940,380	925,542	996,144	1,070,927	1,100,786
Yukon	26,528	28,189	29,771	31,351	35,264
Canada	29,627,384	31,231,538	33,184,431	34,950,984	36,804,991
c. CanEcum	ene Popula	tion Percer	nt Captured	(EcuPlaces)	
ProvTerr	EcuPerPop01	EcuPerPop06	EcuPerPop11	EcuPerPop16	EcuPerPop21
Alberta	98.9%	98.7%	98.8%	99.2%	99.3%
British Columbia	98.2%	98.5%	98.8%	99.0%	99.2%
Manitoba	96.5%	96.5%	96.9%	99.6%	99.7%
New Brunswick	96.0%	96.1%	96.3%	98.1%	98.8%
Newfoundland and Labrador	98.0%	98.0%	98.4%	99.4%	99.4%
Northwest Territories	99.4%	98.4%	98.5%	98.5%	98.5%
Nova Scotia	97.1%	96.3%	97.2%	98.2%	100.0%
Nunavut	100.0%	100.0%	100.0%	99.8%	100.0%
Ontario	99.3%	99.3%	99.3%	99.8%	99.8%
Prince Edward Island	88.0%	88.0%	88.1%	88.6%	91.9%
Québec	99.0%	99.5%	99.5%	99.7%	99.6%
Saskatchewan	95.4%	95.8%	96.2%	97.0%	96.7%
Yukon	92.6%	92.3%	87.6%	87.4%	87.7%
Canada	98.6%	98.7%	98.9%	99.4%	99.4%

#### **Core Layers Data Description**

The Canadian Ecumene GIS Database (Basic) V2 is comprised of a set of layers as follows:

Layer	Туре	Description
CanEcumene_BufferArea10kmV1	Polygon	CanEcumene Extent Layer
CanEcumene_PopPlacesPointV3	Point	Populated Places Point File
CanEcumene_PopPlacesPolyV3	Polygon	Populated Places Polygon File
CanEcumeneMasterTableV3X	Table	Populated Places Master Attribute File

Table 2. Descriptions of layers in the CanEcumene GIS Database Basic Vers. 2.0

- The main GIS files are provided in the CanEcumenePopPlacesPoly/Point shape files. Each of these files contain the basic descriptive attributes as follows:
  - EDUID, Ecumene Place Name, Province/Territory, and Latitude/Longitude (points/centroid of polygons).
  - The CanEcumene polygon file contains these same attributes with additional fields describing polygon area and perimeter values calculated in kilometer units.
  - Records with PlySource=1 (Night Lights) are variable depending on the night lights extent, and PlySource= 2 (CGNDB) are constant due to a circulate buffer of 1.5 km radius applied to the CGNDB lat-long location
  - The PointDense field shows the number of points (replicates) used in the multipoint file used for interpolation purposes (Advanced version).
- The 10 km buffer area layer is provided for cartographic purposes, which was created by combining the ecumene polygons with road and rail transportation networks and buffered at 10 km. This file may also be used as an analytical extent for spatial interpolation of point attribute data.
- The CanEcumeneMasterTable3X contains all of the descriptive identifier attributes for the CanEcumene places. Descriptors for each field are provided in Appendix A, along with key field relations among all CanEcumene GDB tables.
  - EcuType is the descriptor of the aggregation type for each point or polygon as described above.
  - Ecumene Zones (alt. 'ecunomic zones') (EcuZones) are unofficial economic zones indicating a community's location relative to core economic areas.
  - This new version contains StatsCan's Statistical Area Classification (SAC) Codes for each record. This is now included for analytical purposes to compare aggregations by SAC versus EcoZones.
  - EcoZones15 are the standard EcoZones delineated by the Canadian Council on Ecological Areas (CCEA). EcoZones18 contain east and west sub-divisions of some of the CCEA EcoZones, used by the Canadian Forest Service (CFS).
  - EcoRegions are subdivisions of EcoZones.

- Forest Zones are delineated on the basis of intersections between forest and nonforest areas (from MODIS kNN, 250 imagery), and forest dominated and nonforest dominated ecozones.
- o In identifying 'Indigenous Communities' for Ecumene mapping purposes two fields of information are used: IndigCSD indicate indigenous communities according to CSDType categories, IndigPer16 is the % of indigenous population from the 2016 Census. A comparison of values in both IndigCSD and IndigPer16 shows many communities that have high % values are not formally indicated by CSDType. Therefore, a third field, IndigEcu, uses values from both fields to indicate Indigenous community status as follows:
  - Y Yes (indigenous community, > 50% population)
  - X Mixed (partially indigenous, 5-49.9% indigenous population)
  - N No (non-indigenous community, < 5% or 0 indigenous population)
  - U Uncertain (conflicting or not enough information to determine)
- Version 3 now contains both Population and Dwelling counts for each CanEcumene place, plus a Population Category field for each Census Period.

#### Concordance Look Up Tables and Socio-Economic Data Tables (XLSX file)

- The CanEcumeneMasterTablesV3.xlsx file contains a duplicate of the CanEcumene MasterV3 table as described above, plus both the concordance look-up tables (LUTS) and Ecumene Data Tables that contain a selection of socio-economic variables.
- The CSDECU\_LUT{YEAR} tables provide the CSD ID and Ecumene Place ECUID links for each of the four census periods (Table 3). Users who wish to map their own census data will need to use these files to aggregate the census data from CSD to Ecumene Places. This file contains the CSD identifiers associated with each Ecumene place for each Census year. Note there are one or many CSD identifiers for each Ecumene place.

Field	Description
CSDUID{YR}	Census Subdivision Unique Id (Year) (Key Index)
CSDName{YR}	Census Subdivision Name (Year)
ECUID	Ecumene Unique ID
ECUName	Ecumene Place Name
ProvTerr	Province or Territory

Table 3. Field structure of the CSD-ECU look-up tables.

 The ECUData{YEAR} tables contain a selection of socio-economic attributes that describe eight variables used to map communities as 'human habitats'. Additional information on the rationale for the selection of these variables can be found in Eddy and Dort (2011).
The field structure for these tables is straightforward as they match the CanEcumene point and polygon files 1:1 (Table 4).

Field	Description
TotPop{PYR}	Total Population (Previous Census Year)
TotPop{CYR}	Total Population (Current Census Year)
TotYut{YR}	Total Youth (<15 yrs)
TotSen{YR}	Total Seniors (>65 yrs)
TotPost{YR}	Total Post-Secondary Education
TotLForce{YR}	Total Labour Force
TotLFUNM{YR}	Total Unemployed
TotValDwell{YR}	Total Value of Dwellings

Table 4. Field structure of the ECUData{year} tables.

# Part B – Extension Layers

The CanEcumene 3.0 GDB contains a number of extension layers as described in the Introduction. Table 4 provides a description of each layer. As also provided with previous version of the CanEcumene GDB, these layers remain unchanged in Version 3.0.

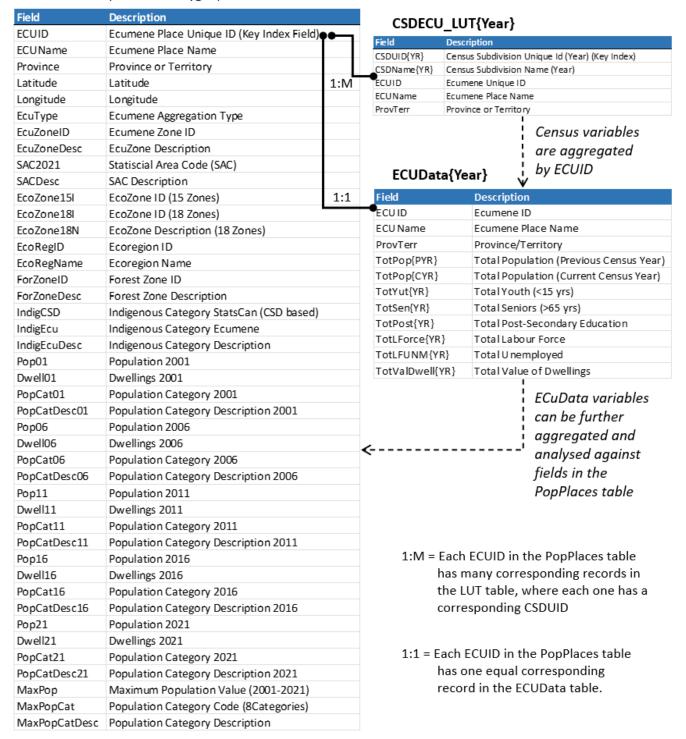
Layer	Description	Туре
CanEcumene_BaseX1_[1,2,5]VSC	Ecumene Extent – Transportation only (1, 2, 5 km)	Area
CanEcumene_BaseX3_[1,2,5]VSC	Ecumene Extent – Transportation and Utilities (1,2,5 km)	Area
CanEcumene_PopPlacesPointMulti_V2	Populated Place Multiple Points (for Interpolation)	Point
CanEcumene_PopPlacesPolyRoadBuff_V2	Populated Places by Road Buffers (100 m)	Area
CanEcumene_PopPlacesRDensityV2	Population Density – Categorical (250 m res)	Raster
CanEcumene_TransAllRoadsV2x	Transportation – Road Buffers – 100 m (100 m res)	Raster
CanEcumene-TransRailwaysV2x	Transportation – Rail Buffers – 100 m (100 m res)	Raster
CanEcumene_UtilPipelinesV2x	Utilities – Pipeline Buffers – 100 m (100 m res)	Raster
CanEcumene_UtilPowerLinesV2x	Utilities – Power Lines Buffers – 50 m (100 m res)	Raster
CanEcumene_XDMSPNightLights2010	DMSP Night Lights Image – 250 m	Raster

Table 4. Layer descriptions for the CanEcumene 3.0 GDB 'extension' layers.

#### APPENDIX A - FIELD DESCRIPTORS AND RELATIONS AMONG CANECUMENE TABLES.

## CanEcumene\_PopPlaces

(Point and Polygon)



### **References and Links**

Eddy, B.G., Muggridge, M., LeBlanc, R., Osmond, J., Kean, C., and Boyd, E., 2020. An Ecological Approach for Mapping Socio-Economic Data in Support of Ecosystems Analysis: Examples in Mapping Canada's Forest Ecumene. One Ecosystem 5: e55881. https://doi.org/10.3897/oneeco.5.e55881

Eddy, B.G. and Dort, A., 2011. Integrating Socio-Economic Data for Integrated Land Management (ILM): Examples from the Humber River Basin, Western Newfoundland. GEOMATICA Vol. 65, No. 3, 2011 pp. 283-291

National Oceanic and Atmospheric Administration (NOAA). 2010. Version 4 DMSP-OLS Nighttime lights time series. Available from:

http://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

Gao, J., 2017. Downscaling Global Spatial Population Projections from 1/8-degree to 1-km Grid Cells, NCAR Technical Note NCAR/TN-537+STR.

https://sedac.ciesin.columbia.edu/data/set/popdynamics-1-km-downscaled-pop-base-year-projection-ssp-2000-2100-rev01

CGNDB, 2018. Link to Canadian Geographic Names Database places: <a href="https://open.canada.ca/data/en/dataset/e27c6eba-3c5d-4051-9db2-082dc6411c2c">https://open.canada.ca/data/en/dataset/e27c6eba-3c5d-4051-9db2-082dc6411c2c</a>

StatCan, 2018a. Link to Statistics Canada's census boundary files: <a href="https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-eng.cfm">https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-eng.cfm</a>

StatCan, 2018b. Link to Statistics Canada's GeoSuite database: https://www12.statcan.gc.ca/census-recensement/2011/geo/ref/geosuite-eng.cfm