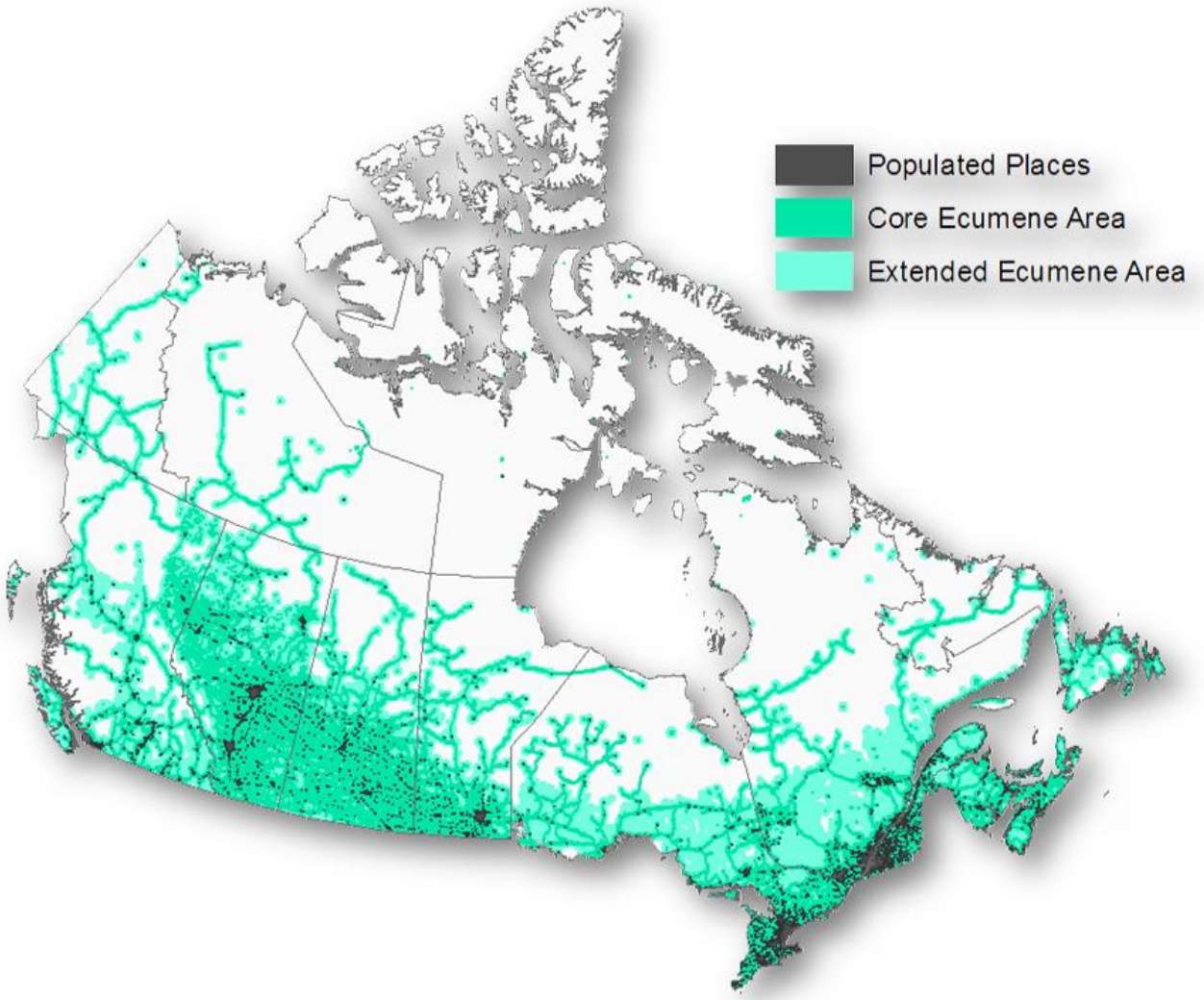
CanEcumene GIS Database
Version 2.0



Technical Reference

November 2020

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. However, 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:

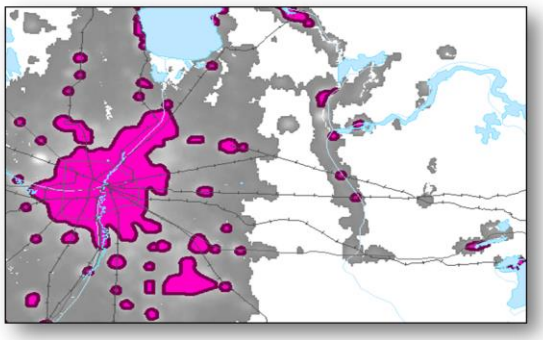
1) to use as an alternative spatial framework for analysis of social, economic and demographic patterns, and
2) to provide a set of GIS layers that map the physical extent of human settlement and infrastructure across Canada.

These two aspects may serve different users for different purposes. They may be applied separately, or in combination. 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 and 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 2.0 Geodatabase was developed to support both aspects. This technical reference accompanies the CanEcumene-V2.GDB database provided on the Federal Geospatial Portal (FGP) and OpenMaps Canada, and is supplemental to Eddy et. al. 2020.

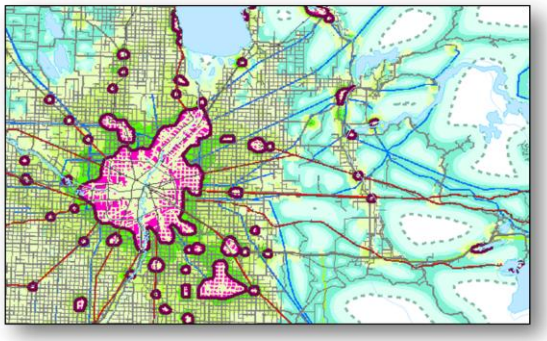
This document is divided into are two parts: Part A – Populated Places (the Core layers), and Part B – Ecumene Extents (Extended Layers). Figure 1 illustrates the distinction between these two:

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 portions of the CanEcumene 2.0 database and respective GIS layers: 1) Core Populated Places and 2) Extension Layers.

Part A – Populated Places

Rationale

The CanEcumene 2.0 database provides an alternative means for mapping demographic, social and economic variables. Until now, 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. 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 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[1]

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

[1] 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:

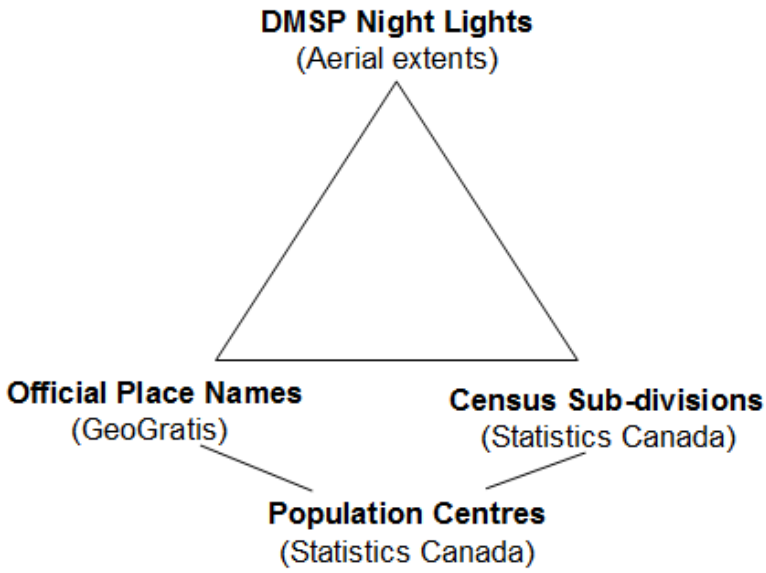


Figure 1. Schema of triangulation method for the development of the Canadian Ecumene GIS Database.

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.

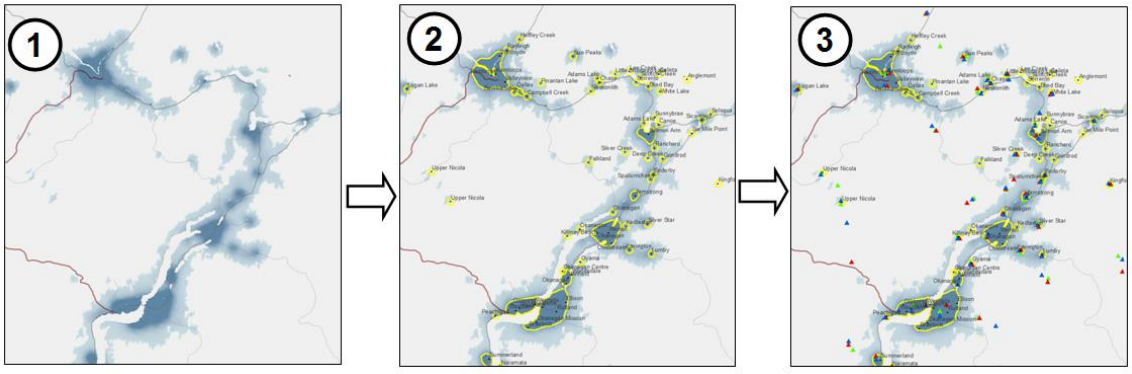


Figure 2. Illustration of the three-step method for aligning and matching populated places from three data sources.

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 and why this was done.

The process resulted in either identifying a match (with match) or leaving some CSD centroids un-matched (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).

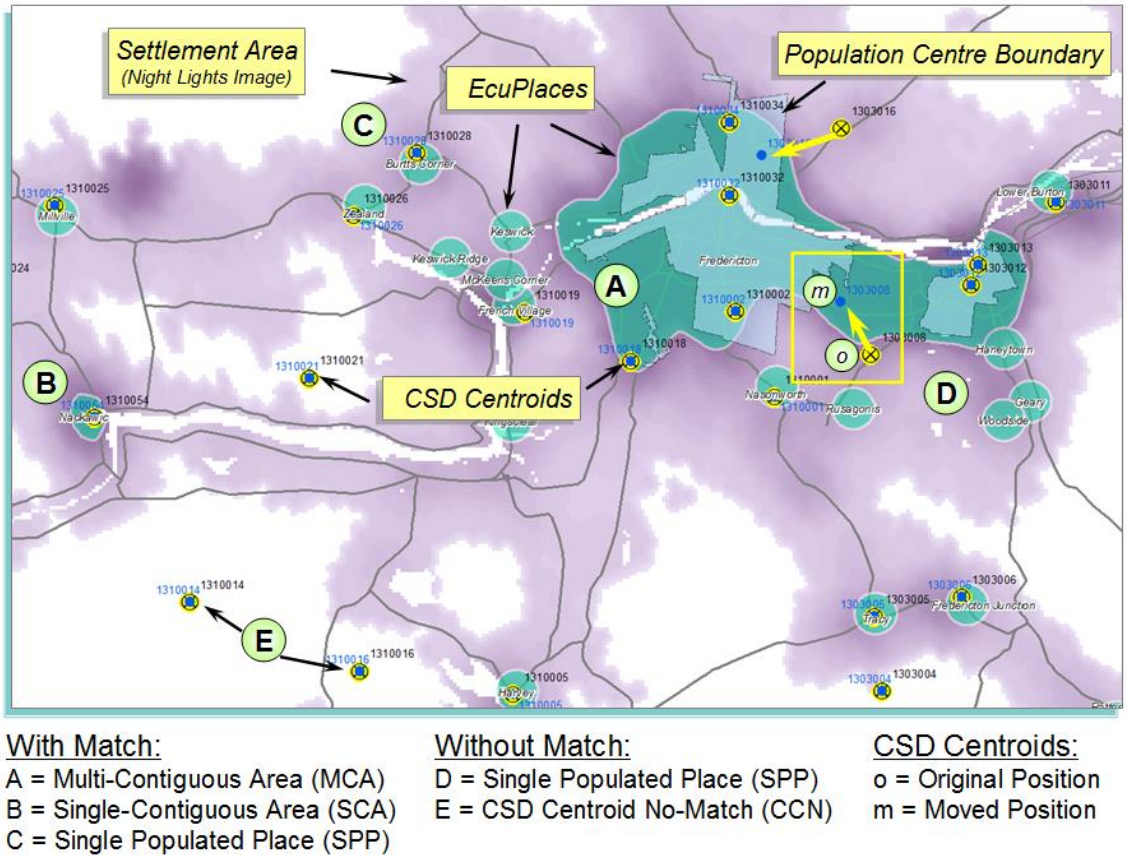


Figure 3. Illustration of how CSD centroids were matched with ecumene place boundaries.

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. They are retained in the ecumene database for additional reference.

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 three periods are relatively small as a 98% representation of the total population was achieved for each census period. The remaining 2% of the population are dispersed among 1300 places, in rural and remote areas. Although CSD data variables could not be linked for these communities, supplemental population estimates were acquired from overlaying the ecumene polygons with the pixel values of NASA's Socio-Economic Data Analysis Center (SEDAC) population grids (Gao 2017). This data provided the needed complement to the CSD data to achieve a near 100% representation by population. Additional attributes were added to the ecumene polygons from a several sources including: 1) ecological zones and regions, 2) unofficial ecunomic zones, 3) forest zones, and 4) indigenous community indicators. These additional reference variables were 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).

	2001		2006		2011		2016	
	Total Pop.	Count	Total Pop.	Count	Total Pop.	Count	Total Pop.	Count
CSDs	29,978,397	4,808	31,563,035	4,550	33,429,076	4,573	35,151,728	5,162
Ecumene Places	29,313,759	2,966	30,910,864	2,897	32,761,384	2,908	34,509,624	2,934
CSD Pop.: % Rep	97.8%	n/a	97.9%	n/a	98.0%	n/a	98.2%	n/a
SEDAC Pop	665,544	1,350	674,031	1,350	684,676	1,350	697,041.2	1,350
SEDAC Pop.: % Rep.	2.2%	n/a	2.1%	n/a	2.0%	n/a	2.0%	n/a
Total	29,979,303	4,316	31,584,895	4,247	33,446,060	4,258	35,206,665	4,284
Total % Rep.	100.0%	89.8%	100.1%	93.3%	100.1%	93.1%	100.2%	83.0%

Core Layers Data Description

The core layers of the CanEcumene 2.0 GIS Database are comprised as follows:

Layer	Description	Type
CanEcumene_PopPlacesPoint_V2	Populated Places - Point Reference	Point
CanEcumene_PopPlacesPoly_V2	Populated Places - Polygon Reference	Area
LFD_EMP_[year]_ESF	Labour Force Data - 2001,2006,2011,2016	Table
LFD_INC_[year]_ESF	Labour Force Income Data - 2001,2006,2011,2017	Table
HHI_AGG_[year]	Human Habitat Indicators - 2001,2006,2011,2016	Table
X_CSDECUConcord[Year]	CSD-ECU Concordance Tables - 2001,2006,2011,2016	Table

Table 2. Descriptions of 'Core' layers in the CanEcumene 2.0 geodatabase.

- The core GIS files are provided in the CanEcumene_PopPlaces Polygon and Point shape files.

- A selection of aggregated labour force data are provided in the LFD tables for mapping and analysis. The data are primary counts aggregated according to a custom aggregation of economic sectors which are described in Figure 5 below.

- The HHI_AGG tables contain a limited set of socio-economic attributes that describe eight variables used to map communities as 'human habitats'. Additional information on these variables can be found in Eddy and Dort (2011).

- The X_CSDECUConcord tables provide the CSDUID and Ecumene Place ECUID links for each of the four census periods. Users who wish to map their own selection of census variables will need to use these files to aggregate the census data from CSD to Ecumene Places.

- The relations among these tables are described in Figure 4 along with descriptions of the fields in each table. The Key Index Field in each table is the ECUID field. This table has a 1:1 relation with the LFD tables which are also keyed on the ECUID field. The concordance files have both CSDUID and ECUID as key fields, and may have 1 to many relations to the point and polygon files.

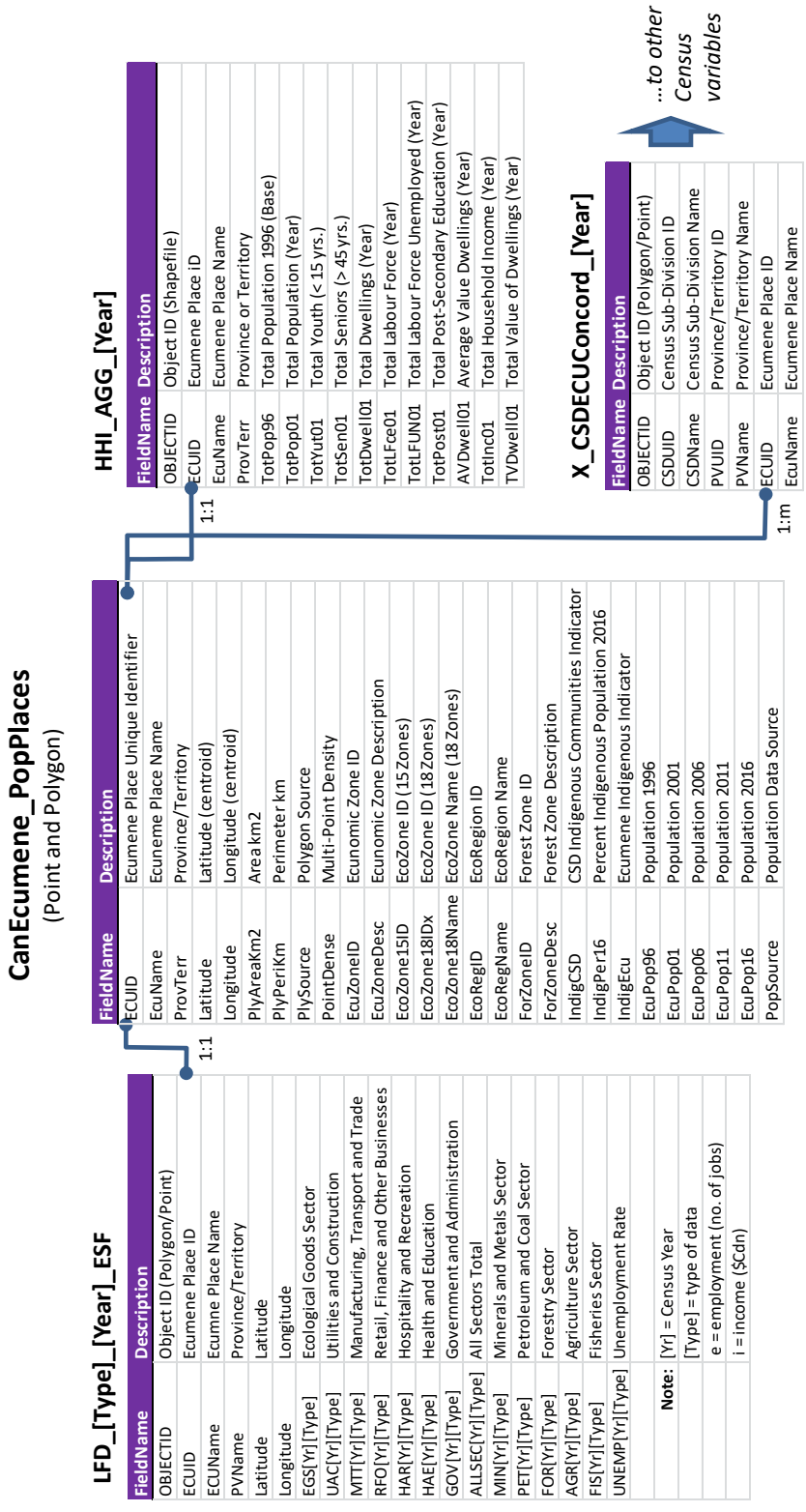


Figure 4. Field Descriptions and Element relations for the CanEcumene GIS Database V2 (Basic).

- The core PopPlaces table contains the same set of fields for both the point and polygon versions, and also contains number of additional reference fields to support various types of analyses. For example, population values can be aggregated according to province/territory, Ecunomic zones (unofficial), ecological zones or regions, forest zones or by indigenous communities. By extension, primary count data from the LFD and HHI tables can also be aggregated in different ways using these fields. Other variables extracted from Census data will need to be aggregated twice; first from CSDUID to ECUID, then by choice of reference field.

- Additional notes on PopPlaces fields:

    - Polygon Area and Perimeter values were calculated from polygon file 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 multi-point file used for interpolation purposes (Advanced version).
    - Ecunomic Zones are unofficial economic zones indicating a community's location relative to core economic areas.
    - 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 non-forest areas (from MODIS kNN, 250 imagery), and forest dominated and non-forest dominated ecozones.
    - 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)

- Data provided in the Labour Force tables (LFD) are aggregated according to the major economic sectors model presented in Figure 5. The specific categories of the North American Industrial Classification System (NAICS) are provided for each major sector.

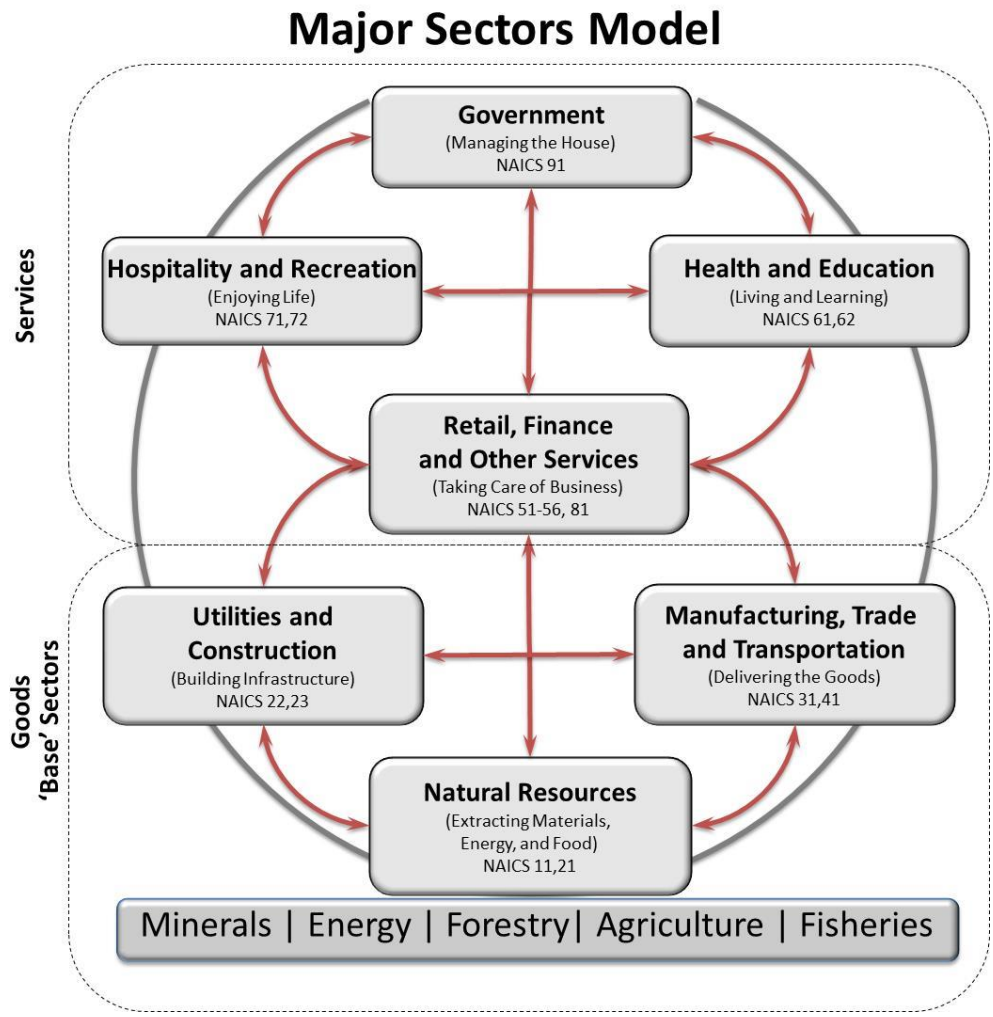


Figure 5. Major Sectors Model for the CanEcumene V2 Labour Force data. Note: Primary data for income and employment were aggregated from the indicated NAICS (North American Industrial Classification Systems) categories. Google 'NAICS' for more details.

Part B – Extension Layers

The extension layers can be viewed in three categories, each of which provide different options for spatial representation of 1) populated areas, 2) transportation and utility networks, and 3) the combination of 1 and 2 to map the overall extent of Canada's Ecumene. Table 3 provides a summary list of these layers, which are graphically illustrated in Figure 6.

Layer	Description	Type
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 3. Descriptions of 'Extension' layers in the CanEcumene 2.0 geodatabase.

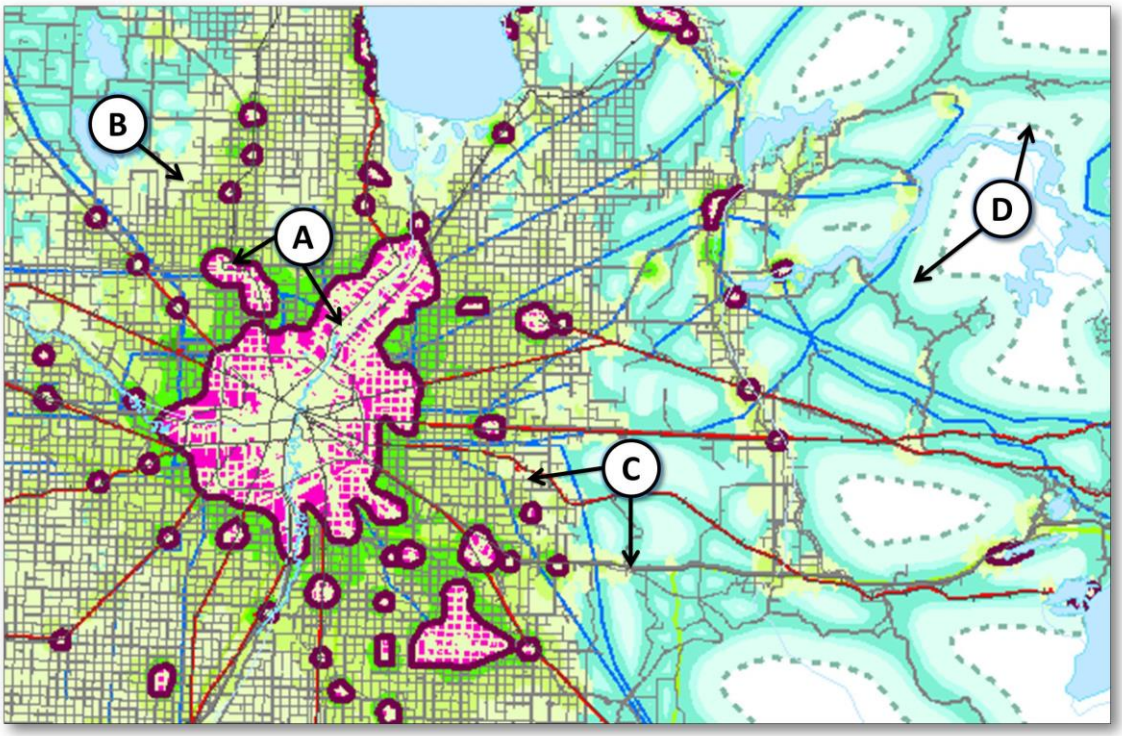


Figure 6. Map illustration of CanEcumene 2.0 Extension layers. See text for description of reference areas.

Populated Places and Areas (A and B)

- As discussed in Part A, populated places are primarily represented primarily by polygons derived from DMSP night lights imagery (larger areas) and buffered points (smaller areas) indicated as 'A' in Figure 6.

- Also in 'A' are more detailed polygons derived by buffering road networks that intersect with the core polygons (PopulatedPlacesPolyRoadBuffv2). These polygons are also coded with the same ECUID as the more general polygons, and may be used for applications that require more detailed boundaries for populated areas (e.g. fire risk analysis, natural hazards, etc.). They contain the same set of reference attributes as the core PopPlacesPoly/Point files, including population values from 1996-2016.

- A third representation of populated areas is a population density raster (PopPlacesRDensityV2), derived from classifying the radiance values in the DMSP imagery (XDMSPNightLights2010). The categories were derived according to the following values:

Radiance Values	Inferred Population Density
1-8	Sparse
9-16	Low
17-32	Moderate
33-63	High

Note: The population density categories are inferred based on visual comparison with the PopPlaces polygon files and other sources such as Google Earth. They are approximate and should only be considered a 'proxy' for population density, and not an actual quantitative estimate.

Transportation and Utilities (C)

- Transportation and Utilities are represented by the TransAllRoads, TransRailways, UtilPipelines and UtilPowerlines raster layers. The source data for these layers is primarily CanVec (NRCan, 2017), supplemented by older GeoBase data (GeoBase, 2006/2016)[2].

- Vector data from these sources were combined into national coverages and were rasterized to a 100 meter resolution. They were then buffered (using pixel inflation) at 100 meters.

- The result is national coverages in raster format of transportation and utility footprints. The 100 meter buffer was selected as an approximate area of influence of human influence on the natural landscape surrounding these features.

[2] Occasional gaps in the CanVec map series required filling in with coarser scaled data from GeoBase layers.

Ecumene Extents (D)

- Six additional layers are provided that map the overall extents of the Ecumene by combining all of the populated places and transportation and utility layers. Taken in combination, these layers provide a reasonable, national level approximation of the overall Ecumene of Canada

- As discussed above, users may need flexibility in defining what to include in mapping these extents. The BaseX1 layers use transportation networks only, whereas the BaseX3 layers include powerlines and pipelines.

- Each set of layers are provided at three buffered intervals: 1 km, 2 km, and 5 km. This allows them to be used in combination or individually depending on the application.

- The 5 km extent is more suitable for national (small scale) mapping, whereas the 1 and 2 km intervals are more likely to be useful at regional scales.

- Vectors from each of these layers were smoothed to provide a more appealing visual effect, and to also include areas in sharp corners of transportation and utility corridors. Therefore, the 1, 2, and 5 km intervals are approximate in some areas.

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

CanVec, 2017. Link to Topographic Data of Canada – CanVec Series: https://open.canada.ca/data/en/dataset/8ba2aa2a-7bb9-4448-b4d7-f164409fe056

GeoBase, 2006. Link to National Roads Network: https://open.canada.ca/data/en/dataset/3d282116-e556-400c-9306-ca1a3cada77f

GeoBase, 2016: Link to National Railway Network (FGP only): https://gcgeo.gc.ca/geonetwork/metadata/eng/ac26807e-a1e8-49fa-87bf-451175a859b8

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

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

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