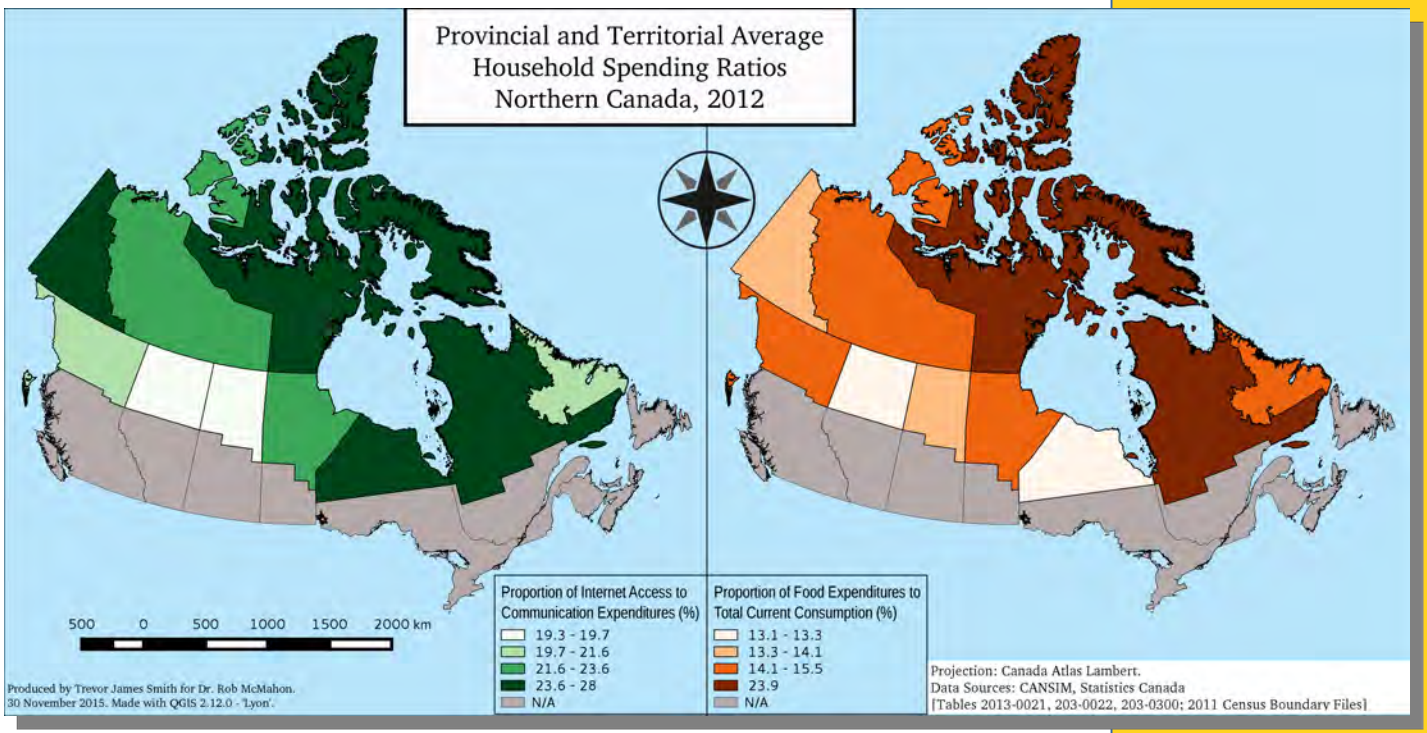
An Open Source GIS and Mapping Methodology for Internet Access in Remote and Rural Indigenous Communities



An Open Source GIS and Mapping Methodology for Internet Access in Remote and Rural Indigenous Communities

The views expressed in this report are those of the authors. This project was funded through a .CA Community Investment Program grant from the Canadian Internet Registration Authority (CIRA). This report is available for download on the First Mile website: http://firstmile.ca

Suggested reference for this report:

Smith, T.J., McMahon, R., Whiteduck, T. (2017). *An Open Source GIS and Mapping Methodology for Internet Access in Remote and Rural Indigenous Communities*. First Mile Connectivity Consortium. February. 43 pages.

FMCC RESEARCH TEAM AND AUTHORS:

Trevor James Smith, M.Sc., Environmental Researcher and Spatial Analyst, QC
Rob McMahon, PhD, Assistant Professor, University of Alberta, AB
Tim Whiteduck, Director of Technology, First Nations Education Council, QC

CONTACT INFORMATION

First Mile Connectivity Consortium (FMCC)
PO Box 104, Fredericton, NB E3B 4Y2
Web: www.firstmile.ca
Email: info@firstmile.ca
Tel: 877-737-5638 ext. 4522
Listserv: features news, updates and discussion about community broadband systems in Canada's rural and northern regions with a focus on remote Indigenous communities. The listserv is hosted by the University of New Brunswick. To join, send a note to Susan O'Donnell at susanodo@unb.ca

The First Mile Connectivity Consortium (FMCC) is a registered national non-profit organization in Canada. The FMCC membership and board of directors consists of First Nations technology organizations serving remote and rural areas. FMCC associates include researchers focused on broadband infrastructure and services in remote and rural communities. The FMCC is engaged in research and public outreach initiatives focused on contributing to telecommunications policy and regulation in Canada. In 2016 FMCC was contracted by Innovation, Science and Economic Development Canada to conduct research on digital technology adoption in northern and remote Indigenous communities. This work includes a comprehensive literature review on this topic, and exploring and testing different methodologies for remotely conducting research on digital technology adoption in remote communities. **For more information, and to access this report, visit: http://www.firstmile.ca**

Acknowledgments

We want to acknowledge and thank the members of the First Mile Connectivity Consortium (FMCC), and residents of the northern and remote Indigenous communities, whose work in digital innovation is an inspiration for this project. We thank all our partners and funders for their support, in particular the First Nation regional community intermediary organizations who compose the membership of the First Mile Connectivity Consortium:

- Keewaytinook Okimakanak K-Net Services (Ontario)
- First Nations Education Council (Quebec)
- Atlantic Canada's First Nations Help Desk (Atlantic region)
- First Nations Health and Social Secretariat of Manitoba
- First Nations Technical Services Advisory Group Inc. (Alberta)
- First Nations Technology Council (B.C.)
- Western James Bay Telecom Network (Northeastern Ontario)

We also acknowledge and thank the university-based researchers involved in this work through the First Nations Innovation project. These include researchers based at:

- University of New Brunswick
- University of Alberta
- Université Laval

The authors also thank Concordia University department of Geography, Planning and Environment Associate Professor Sebastien Caquard, PhD, for his helpful comments on an earlier draft of this paper, and Lecturer Donny Seto, MPPPA, for his helpful advice in the work-flow development and guidance on GIS styling and best practices. We also thank Fenwick McKelvey, PhD, an Assistant Professor in the department of Journalism and Communication at Concordia University, for his support on Internet performance data and monitoring, and Cassandra Lamontagne, M.Sc., for her helpful comments on and review of our early drafts. We also acknowledge and thank the First Nations Innovation Project, based at the University of New Brunswick, for its leadership on this topic over the years. We thank FMCC and FNI researchers for providing comments and feedback on report drafts, including Dr. Susan O'Donnell, Brian Beaton, and Dr. Heather Hudson. Finally, we offer our thanks to everyone who is working to develop and deliver digital infrastructure and services to the unique communities discussed in this report.

Summary

"One test of any new technological advance is whether it makes the problem worse, because it is likely more complex than its predecessors, or whether its design represented a possible improvement."
(Goodchild, 1993, p. 446)

In this report we discuss the efforts of the First Mile Connectivity Consortium (FMCC) to shape a Geographic Information System (GIS) platform into a tool for data-driven policy advocacy. This work took place in the context of a lack of robust, accurate data concerning broadband access in Canada's northern and remote regions. Given this challenge, we sought to develop a transparent methodology to (re)present the limited existing statistical data on broadband access and affordability in maps of remote and Northern Indigenous communities in Canada. This was done to outline a GIS design process that we can adopt and adapt as more accurate data from these regions becomes available, as well as highlight and reflect on the design choices we made throughout this project.

Our methodology involves first defining a geographic community of interest, and then collecting, formatting and spatially encoding statistical data in visualizations that we used to contribute to an intervention in a regulatory proceeding on digital infrastructure and services in Canada's northern, rural and remote regions. We discuss the digital literacy challenges that we encountered in this project, considering how they may impact further adoption, adaption and sustainability of our methodology by community groups.

In the future, we hope these steps can be taken up and used by community-based organizations to generate their own GIS-supported data visualizations. In our opinion, this methodology might be useful for such groups to: 1) collect and display in a useful way all of the statistical information available about a community or region; and 2) show relationships between or among various statistical indicators.

Table of Contents

Digital Divide Policy, Broadband Data and the First Nations Connectivity Consortium

In Canada, the federal government is taking steps to address significant and persistent digital divides experienced by residents of rural, remote and Northern communities. In this context, in 2015 the national telecommunications regulator, the Canadian Radio-Television and Telecommunications Commission (CRTC) launched national consultations to update the definition of the 'basic service objective' for telecommunications, as well as define the regulatory environment to support the delivery of those services to all Canadians (Canadian Radio-Television and Telecommunications Commission (CRTC, 2015; Measurement Lab, 2014). To address these questions, the CRTC's public hearings process allowed interested parties to file a series of written and oral interventions that present arguments and evidence to inform the Commission's regulatory decision on this topic.

However, robust data on existing broadband access and affordability in the northern and remote regions of the country is lacking. This is a major challenge for policy development in this area: simply put, there is little to no robust, accurate data currently available about connectivity in these regions. As acknowledged by agencies including Statistics Canada and the CRTC, much of what we know about current broadband access is made available on a limited basis through private parties (such as telecommunications companies) or is determined by educated guesses based on available information.

Despite these limitations, in the course of the CRTC hearings, several parties presented available geospatial data sets as evidence of broadband access and affordability in northern and remote regions. For example, telecommunications providers provide coverage maps - but these are often redacted for competitive reasons, lack granularity, or fail to capture 'on the ground' reports of affordability and/or quality of service from local communities and households. Third-party researchers, such as M-Lab and the Canadian Internet Registration Authority (CIRA), are now generating data through open source web-based platforms. For example, CIRA assesses Internet performance using a test called the Network Diagnostic Test (NDT) provided by M-Lab (Measurement Lab, 2015).[1]

The CRTC recently launched its own access mapping initiative, the Broadband Measurement Project - conducted by a third-party provider, SamKnows. Started in 2015, this project collected data from approximately 5,000 Canadians located across the country, with the goal of informing the CRTC's future broadband policy. While this initiative provided much-needed data independent from that was issued by telecommunications providers, its early-stage results have been critiqued for failing to include enough granularity to reflect the realities in northern and remote regions of the country.[2] Furthermore, these efforts are still very new, and so have not yet developed a robust database of information on this issue, particularly in remote and Northern regions.

During the CRTC 'basic service objective' hearings in 2015-2016, these various data sources came in tension with one another, as Commissioners were confronted with a multiplicity of representations of

1 For more information on this methodology, and to run the CIRA test yourself, visit: https://cira.ca/cira-internet-performance-test-0

2 For more information on the CRTC's work in this area, visit: http://www.crtc.gc.ca/eng/internet/proj.htm

geospatial data claiming that access and affordability either was or was not a problem in northern and remote regions of Canada. In this context, the First Mile Connectivity Consortium (FMCC) intervened in these regulatory proceedings to contribute to the debate.[3] FMCC is a national non-profit association of community-based First Nations technology organizations and university-based researchers (see www.firstmile.ca). (Disclosure: The study authors are members of this organization). Our work focuses on innovative solutions to digital infrastructure and services with and in rural and remote Indigenous communities. Through advocating for policy and regulatory support for Indigenous-led technology projects, the group aims to enable members of First Nations communities to utilize digital networks and technologies to meet their development goals (Beaton et al., 2016; McMahon et al., 2011).

In our intervention during the CRTC regulatory proceedings, the FMCC aimed to generate a representation of existing geospatial data on broadband access and affordability that portrayed the conditions of people living in northern and remote regions of Canada. Specifically, the team wanted to illustrate the geographic boundaries established by FMCC member organizations, rather than by third-party groups such as government agencies or commercial providers. By illustrating how geospatial data mapped on to regions where the existing networks built and operated by Indigenous technology organizations are located, we sought to show a way to present data in a way that more accurately represented the realities of the FMCC member organizations.[4] Importantly, this effort was still based on the very limited data available from remote and Northern regions; rather than presenting an accurate snapshot of connectivity in these areas, we instead present a methodology that FMCC (and others) can use to graphically present such information once accurate data becomes available. We do not present robust data about northern connectivity, since it is not yet available.

For these reasons, we decided to generate an operational work-flow that FMCC researchers and First Nations IT professionals can use to construct and utilize GIS platforms to collect and reformat geospatial data available through government agencies such as StatsCan. By transparently outlining how we re-appropriated various data sets and re-shaped them to illustrate the broadband access and affordability challenges of a specific community of interest, we provide one example of how other groups can adapt or modify geospatial data to meet their needs. In our opinion, this methodology might be useful for such groups to: 1) collect and display in a useful way all of the statistical information available about a community or region; and 2) show relationships between or among various statistical indicators such as educational attainment and rate of unemployment; poverty and number of people per household; sanitation (piped water and sewer) and frequency of certain illnesses, etc.

With this focus in mind, in the next section we illustrate the design process we developed during the course of the CRTC hearings to aggregate and re-present already-existing geographic and statistical data about broadband access and affordability using freely available, open source software platforms.

3 The FMCC has been involved in past proceedings, and documented those efforts to support similar initiatives. For example, see McMahon, Hudson and Fabian (2014), available at: http://ci-journal.net/index.php/ciej/article/view/1106/1092

4 This work has been supported through funding from several parties, including the Canadian Internet Registration Authority's .CA Community Investment Program: https://cira.ca/community-investment-program

The Right Tools for the Job

Performing spatial analysis has traditionally been a difficult research activity. The various data formats, technologies and techniques typically require a level of specialized training and academic background that allows for researchers to understand what data they have access to, what geographic processes they can perform to interpret that data, how to initiate those analytical processes and, most importantly, how to properly interpret and effectively present their results to a range of audiences. Statistical analyses are essential for quantitative researchers to examine and determine trends from their observations; when the data contains a spatial component, existing in both space and time, the need for heavy computational analysis becomes essential to that activity. In this regard, the field of digital cartography has innovated alongside the rise of personal and globally decentralized computers (Dobson, 1983; Goodchild, 1993), and of the Free and Open Source Software (FOSS) movement. Due to these trends, the digital tools required to examine this spatial information have become almost universally accessible for almost all users.

The modern approach to this type of data analysis was envisioned originally by Canadian geographer Roger Tomlinson, who is credited as the founder of one all-encompassing tool for spatial analysis in the mid-to-late 20th century, Geographic Information Systems (GIS) (Tomlinson, 1962). This tool allows users to create, modify, analyze, and present spatial information. Contemporary, Internet-enabled applications of GIS technologies have contributed to the rise of user-friendly "Web 2.0" software applications such as location-based services (e.g. Google Maps), the widespread sharing and representation of geographic data (e.g. GeoNode software, Natural Resource Canada's *GeoBase* project) and crowd-sourced interactive slippy-maps (e.g the OpenStreetMap project). All of these GIS tools, supported by the explosion of networked computation, aid us in understanding our world, its physical characteristics, its cultural diversities, and more.

In recent decades GIS researchers and professionals identified the capacity for quick computation of spatial data to help inform policy and track socioeconomic trends (Keenan, 2008). Personal computers, the lowering cost of technology devices and software, and the network connectivity of the Internet have all radically changed the potential of GIS for these purposes. Personal and enterprise-level geospatial databases and commercial applications have expanded its role to a broad tool used for understanding environmental conditions, economic flows, urban/rural divides and other information for peoples across the globe. Regardless of discipline, quantitative and qualitative applications of GIS can help answer a myriad of geographically-oriented research questions (Keenan, 2008).

GIS analysis can be a useful tool to simplify, curate, query, manage and visualize geo-data. However, in an age of information abundance, it is difficult to sort through and find data sets that are both accurate and reflect the requirements of user communities. At the same time, the availability of statistical data sets is shaped through government decisions, which can shift over time. Many challenges arose in our attempts to collect and standardize socioeconomic data from census and intercensal years, due to major changes in the methods and variables utilized by statistical agencies. For example, the decision of the former federal government of Canada to retire the long-form census in lieu of a National Household Survey for 2011, as well as the discontinuance of many pertinent tables of

data on socioeconomic variables and digital technologies, posed challenges in locating and interpreting data from certain years (Kingston, 2015). This situation clearly limits the data and knowledge resources available to groups and although the federal administration has reinstated the long-form census (Harris, 2015), the gap in information for the 2011 period will present researchers with significant assessment challenges in the coming years (CPHA, 2010).

This report was developed by FMCC researchers to present a methodology we developed for performing spatial representations of publicly available Canadian socioeconomic data associated with broadband access and affordability using a combination of desktop-based Open Source Software and web-based utilities. In our opinion, Free and Open Source Software (FOSS) is the most viable option for data processing and management. The Open Source movement - realized by software developers who wish to release source code with relatively few restrictions for anyone to inspect, modify, and integrate into larger projects - has provided many free and reliable options for data management and GIS technologies. Concerns with software purchasing and subscriptions concerns are mitigated with the use of FOSS.[5]

The accessibility of geographic analyses and visualizations in recent years, both from a computational and a financial perspective, has broadened the capacity for typical non-GIS researchers and even members of the public with basic computer skills. That said, there still exists several barriers to performing this analysis. In this context, this guide is meant to explain a methodological approach and expose some of the potential limitations or challenges of performing spatial analysis using Open Canada geospatial data.

In our opinion, and building upon some of the factors presented by Sieber (2007), some notable barriers to performing this type of analysis include the following:

- Lack of robust data that accurately represents the realities of Internet access in remote and rural Indigenous communities in Canada.
- Familiarization with the Open Source software ecology takes time and requires some preexisting knowledge of statistical programs and web data extraction techniques.
- Data repositories often use different formats and collection procedures, and vary in terms of relative quality and/or commensurability.
- The technical knowledge required for GIS (as well as the licenses needed for proprietary versions of GIS) have traditionally impeded researchers in many fields from adopting these tools.
- Particularly for web applications, a reliable broadband Internet connection is needed for data collection and interactive visualizations.

5 This methodology can still nonetheless be replicated using proprietary solutions such as ESRI's ArcGIS solutions provided that researchers have licenses access to these products.

- Ethical reservations exist concerning the creation, control, and applications of sensitive information.

With these challenges in mind, we present here an approachable guide for performing statistical/socioeconomic data collection, parsing, and geographical presentation using both a fully-featured Open Source GIS, QGIS, as well as a web-based spatial analysis platform, CARTO (previously CartoDB). Both of these approaches are aimed at different types of users and can help provide different data products depending on the familiarity of the user with GIS software. The guide is organized by platform, with conceptual work-flow suggestions for more experienced users and step-by-step instructions for less experienced users. Considerations when using each tool are briefly explored and both options are presented with reference to their strengths and limitations. Both data analysis options demand different steps but essentially perform the same data handling and presentation processes.

Methodology Summary

Spatial analyses are a variety of research approaches that involve examining data that has a spatial or geospatial component. This can include readings or point observations (points), directional features like transmission lines, rivers, or highways (lines), areal features (polygons), and continuous regional imagery (grids). By relating patterns of data across geographic coordinates and years, patterns can emerge to better understand distribution across space and time.

Using these approaches we present here a methodology for using spatial analysis for examining socioeconomic information geospatially by joining Census and other similar data tables to geographic region shapefiles. The methodology outlines a step-by-step process for performing the following steps:

1. Gathering, Creating, and Preparing Canadian Regional Spatial Data
2. Identifying and Parsing Statistics Canada Socioeconomic Tables for Spatial Analysis
3. Gathering and Formatting Information and Communication Technology Indicators
4. Joining Tables, Combining Data Sources, and Styling and Presenting Analysis Results

The work necessary for Step 1 to manipulate and generate spatial data requires the use of a GIS, while Step 2 can be performed using an Internet browser (e.g. Internet Explorer, Firefox) and standard spreadsheet software (e.g. Microsoft Office, LibreOffice). The approach for Step 3 is dependent on the software platform being used (QGIS or CARTO); the steps used in both platforms are explained in sequence.

Building a Geographic Definition for the 'First Mile' Community of Interest

Step 1: Collecting and Creating Spatial Data

Geo-data Collection and Selection

In order to perform these steps to collect, manage, and generate geospatial data we used a cross-platform, Free and Open Source GIS application called QGIS (QGIS Development Team, 2016). QGIS is a user-friendly, fully-featured GIS application that can be expanded and customized by users to suit specific data processing and visualization needs.

We began by collecting the spatial data sets that would form baseline research units. Throughout the Canadian provinces and territories, many different levels of spatial aggregations of socioeconomic interest are available through the Statistics Canada Census as Boundary Files. These are available as ESRI Shapefile, GML, or older MapInfo format. Shapefiles and GMLs are handled natively by most available GIS platforms, while MapInfo requires some conversion to be used with modern systems. Most socioeconomic results tables available through Canadian statistical agencies provide information for relatively aggregated regional levels, the spatial units being: Provincial/Territorial, Municipal, and Health Regional. These files are available under the Open Canada License through the StatsCan Census website.

Some considerations when collecting this data are that:

- Cartographic Boundary Files (where land boundaries are map accurate) are much larger in size than Digital Boundary Files (where land boundaries are simplified and extended across water and islands) and serve different purposes with regards to presentation.
- Geographic boundaries are updated and changed between Censuses to best represent changes to the socioeconomic realities across Canada, while some boundaries such as Health Region extents are updated almost annually.
- The finer the resolution of the data, the larger the files will be (e.g. ~50 MB for Health Regions vs. ~500 MB for Census Dissemination Blocks).
- Geographic IDs for regions between Censuses often change or are redefined between periods, making it difficult to compare regional change across time.

As such, when gathering this data, it is best practice to ensure that adequate storage space is available and that the years of interest are specified. When creating maps from scratch to be presented in static/paper format, Cartographic Boundary files will provide a much more accurate image of the areas of interest. Digital Boundary files are small in size, making them a better fit for platform services where file sizes are a consideration and are also better for interactive visualizations as they will load faster. Digital Boundaries are best suited as overlays for a basemap where land and water boundaries are already specified.[6]

6 For Northern regions of Nunavut, the Digital Boundary files do not properly overlay all land boundaries due to the simplifications taken in their production. Modifying these extents using a Vertex Editor can be done with relative ease.

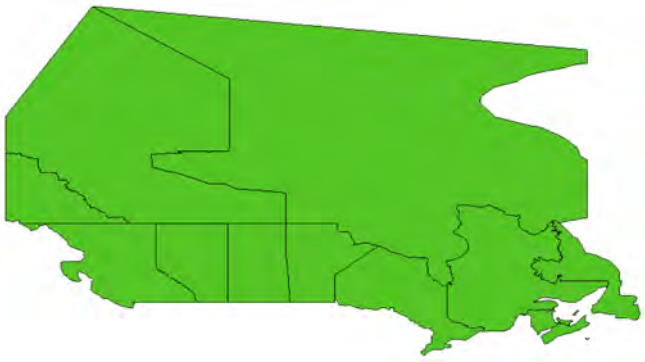


Illustration 2: Provincial and Territorial Digital Boundary File, NAD83 projection

Illustration 1: Provincial and Territorial Cartographic Boundary File, NAD83 projection

Once these files have been gathered, in order to remove areas that are not of concern, it is necessary to perform selection and save the results as a separate file. For most cases, this step requires the QGIS tool "Select by Attribute" or, with the aid of an existing shape of such a boundary, "Select by Location". As this process is similar to a database query, this procedure is nearly identical in format to a command one might find composed for a SQL-like database. For example, a command that asks the GIS to return all shapes that are part of the territory of Nunavik might look similar to as follows:

"FROM shapefile_name SELECT shape AS "Nunavik"

WHERE region_id IS "Nunavik_Region"

When following the case prompt through the QGIS "Select by Attribute" tool, this query translates to:

Input Layer: **shapefile_name**

Selection Attribute: **region_id**

Operator: **'='**

Value: **'Nunavik_Region'**

Output Layer: **'Nunavik'**

Regions can also be selected using the "Select Features by area or single click" tool if the area of interest is a small continuous area or a large region delimited by a square bounding box using the cursor. In cases where the region of interest is irregularly shaped or requires units in an existing shapefile to be split, creating a new shapefile with these coordinates is necessary in order to create an accurate new selection. For our research, this step was necessary to specify regions falling within Canada Revenue Agency's Northern and Intermediate tax zones.

Identifying and Delimiting the Areas of Interest

In Canada, northern Indigenous communities are located in three territories (Yukon, Northwest Territories and Nunavut) and in the northern regions of the provinces. While these small, dispersed communities share challenges, including a lack of digital telecommunications infrastructure, they do not have a common defined geographic community of interest and in fact are jurisdictionally separated by various provincial, territorial and Indigenous borders. This diversity can make it difficult for people living in these areas to present a collective voice in telecommunications advocacy activities.

In this context we created a geographically-defined community of interest to represent the FMCC constituency of communities served by regional First Nations technology organizations. We present a short methodology to define the geographic boundaries of the Northern and remote Indigenous regions that encompass the association's constituent members.

We created this geometry through an intersection of existing data sets developed by government agencies and geographic boundaries that were translated from the *Income Tax Act of Canada* (Justice Laws of the Government of Canada, 2015; Statistics Canada, 2015a). The physical process of developing this boundary was through the digitization of geographic boundaries of the Northern and Intermediate Tax Regions of Canada. This involved collecting Canadian census cadastral data sets for several different levels of government (Statistics Canada, 2015a) and building a vector (shapefile) image to represent the geographic region under consideration.

The figure below illustrates the outcome of this process, displaying the geographic regions that encompass the community of interest of FMCC's member organizations (in blue and teal). While this map was relatively simple to make using GIS techniques, the choices involved in creating it hold political and social implications for people living in these regions.

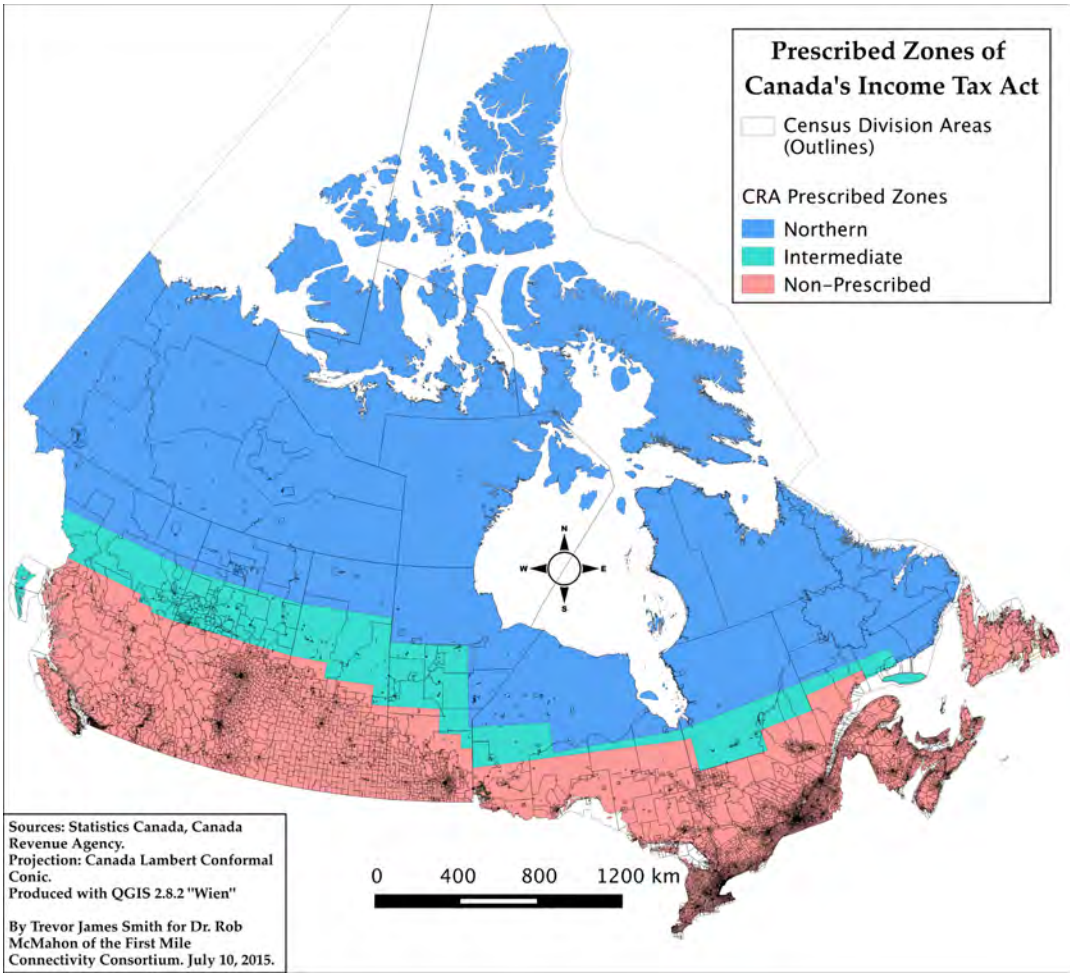


Illustration 3: Canadian Census Division Areas overlaying FMCC Community of Interest (Northern and Intermediate Regions)

For this reason we stress that our work in this area must be reviewed and verified by the Indigenous people living in these regions. Our process discussion reveals the choices we made in constructing our maps, and illustrates some of the ways that we attempted to use multiple sources to address our mapping requirements. We note the need for close partnerships between technical designers and participating communities of interest to help ensure that GIS maps used in policy advocacy accurately reflect interpretations of space held by involved groups.

Digitizing and Intersecting Canada Revenue Agency's Tax Zones

Working with the spatial boundaries defined earlier, we created an initial shapefile with accurate coordinates for the Northern or Intermediate zone which spanned the geographic communities of interest. We began by first digitizing the specific geographic limits of the prescribed Northern and Intermediate zones as polygonal vertices (Justice Laws of the Government of Canada, 2015).[7] For instance, where the Income Tax Regulations (C.R.C., c.945, s.7303.1(2)(c)) states that the intermediate zone comprises the following area:

> *"that part of Saskatchewan that lies*
>
> *i. north of 55°00'N latitude,*
>
> *ii. north of 54°15'N latitude and east of 107°00'W longitude, or*
>
> *iii. north of 53°20'N latitude and east of 103°00'W longitude;"*

These coordinates are converted from Degree-Minutes-Seconds to Decimal-Degrees and used to mark the bounding area, closed by the border of the Northern Zone, as specified in the Act. Using a vertex tool, these coordinates are treated as vertices of a new polygon and added by hand then fine-tuned to precise geographic limits. The *Numerical Vertex Edit* plug-in available through the QGIS plugin repository provides one method of specifying these coordinates (see below).[8]

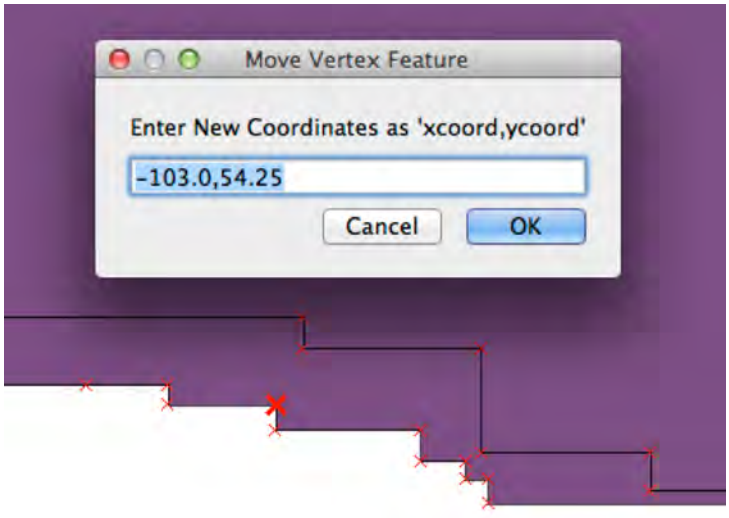


Illustration 4: QGIS Plugin – "Numerical Vertex Editor"

7 Digitizing is relatively difficult to perform without prior technical knowledge. The authors would caution against people with limited to no background in GIS to perform this. For those with GIS knowledge, a comprehensive guide to digitizing can be found at http://www.qgistutorials.com/en/docs/digitizing_basics.html (Gandhi, 2015b)

8 This is not the only possible approach. Another method could include spatially visualizing a spreadsheet with coordinate data and then connecting these points to create new polygons.

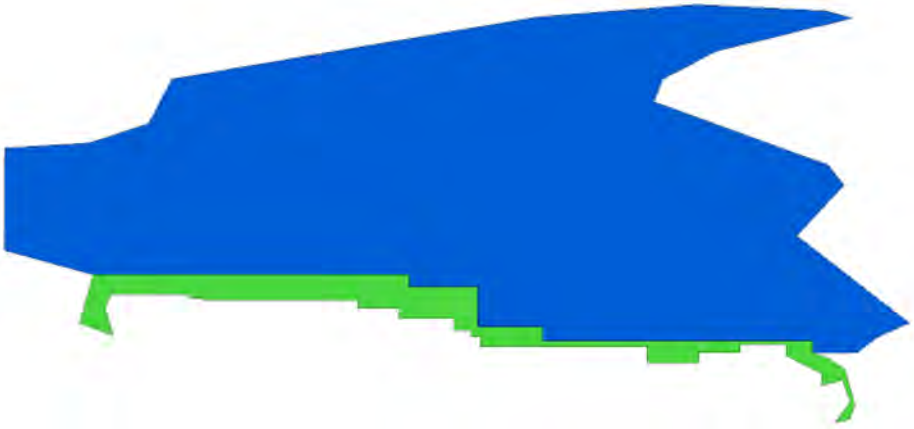


Illustration 5: Completed Income Tax Act Zones Shapefile Output

Replicating this process for both the "Intermediate" zone (seen above in green) and "Northern" zone (seen above in blue), we then saved the objects as a Shapefile. The northern regions of the "Northern" Zone were extended to cover the Canadian landmass and the majority of the portion of Canada delimited in the Digital Boundary file.[9] The justification for this is to facilitate easier intersection analysis (as shown in Illustration 3).

Intersecting Geometries

In addition to this process of creating the boundaries of the area of interest, we also needed to perform spatial intersections between our northern boundaries and the census boundary files to preserve only the geographic areas that we were interested in and assign the zone value to the resulting areas. Intersection analysis (or "clipping") is a geographic process whereby only data (points, lines, or polygons) that is overlapping a defined area (the "mask") is preserved and any data present in both the initial file and the "mask" is transferred to this new hybrid shape.

Using the boundary extents for the "Northern" and "Intermediate" zones as a mask, we performed intersections using Shapefiles for Provinces, Territories, Health Regions, and other geographic boundaries. The resulting products were useful in that they preserved the geographic boundaries, the geographic identifiers (names, region codes, region types, etc.) and the Canada Revenue Agency Tax Zone identifier ("Northern", Intermediate") while not preserving geographic information for areas beyond the "mask". These modified data sets were then used for data joining specified in Step 3.

9 To simplify the work-flow, we included Haida Gwaii, Anticosti, Les Îles-de-la-Madeleine, and Sable Island in the Intermediate Zone by extending segments of the polygon over water and parts of Alaska, as non-land boundary areas would be removed during the spatial intersection process.

These processes require some experience in Geo-data management, presenting another potential barrier to GIS analyses for non-specialized users. Further complicating these matters, for some government-regulated geographic boundaries, periodical (and in some cases, annual) revisions to physical extents, the creation and deletion of units, and changes to names all presented challenges in comparing data across time.

Illustration 6: QGIS – Outcome of Intersection Process between Income Tax Act Zones and Health Regions

(Lambert Conformal Conic Projection)

Step 2: Collecting and Formatting Statistical Data Sets

Second, we collected and formatted quantitative data from a variety of publicly available statistical sources to illustrate broadband access and affordability in our defined region. As discussed earlier, this process was limited by the lack of valid, accurate data from remote and Northern regions. To address these challenge, more data must be made available (and we noted above some efforts to do so currently underway by organizations including the CRTC and CIRA). Working under these limitations we examined data sets from federal, provincial, territorial, Indigenous and non-governmental organizations. The data sources that we consulted to collect socioeconomic indicators are summarized as follows:

Statistical Bureau	Geographic Scope/Area of Interest
Statistics Canada (CANSIM, Census, National Household Survey)	Canada, Provinces/Territories, Census Divisions/Municipalities, Census Dissemination Areas/Blocks, Health Regions, Economic Regions
Indigenous and Northern Affairs Canada [INAC] (Community Well-Being Index)	Census Divisions/Municipalities, Aboriginal Communities
Government of Newfoundland and Labrador	NFLD, Municipalities, Nunatsiavut Communities
Government of Nunavut	Nunavut, Municipalities, Inuit Communities
Government of the Northwest Territories	NWT, Municipalities, Inuvialuit Communities
Nunivaat	Nunavik Communities (Northern Quebec)

Table 1: Statistical Bureaus and Relevant Socioeconomic Data/Information

Examining this data more closely, we learned that much of the available socioeconomic information is presented for large regional geographic areas such as provinces, territories and at the national level.[10] Depending on the area or subject of interest for the specific analysis, the geographic scale of interest (national, regional, municipal, or local) will affect the data availability, appropriate methods and reliability/certainty of the results. For example, the CANSIM statistical data extraction tool lists data according to census table, with some tables providing finer scale geographic output than others. This technical limitation impacted the way we could illustrate data specific to the geographic community of interest that we generated for this project.

Detailed tables from CANSIM — Displaying 1 to 11 of 11 matches

Title	Description	Table no.
Aboriginal population as a proportion of total population, Canada, provinces, territories and health regions, 1996, every 5 years	Description	109-0012
Distribution of the population aged 5 to 24 with Aboriginal identity, by age group and living arrangement, Canada, occasional	Description	477-0091
Distribution of the population aged 5 to 24 with Aboriginal identity, by age group and work activity of parents, Canada, occasional	Description	477-0092
Labour force survey estimates (LFS), by Aboriginal group, sex and age group, Canada, selected provinces and regions, annually	Description	282-0226

Illustration 7: CANSIM – Table Displaying different Categories and Geographic Levels of Socioeconomic Information

Along with the challenges that stem from the on-line functionality and general availability of statistical data sets, another consideration is that the data collected needs to be formatted and presented in ways that met our explanatory goals. This work involved navigating digital literacy challenges that stem from both data abundance and data scarcity. Finding and formatting data through CANSIM is possible through its data extraction tool, but no automated on-line methods (i.e. API calls) are available at this time to preformat this statistical data. As GIS objects can represent any data that can be mapped, a lot of data processing work was required to ensure that each geographic objects and socioeconomic variable was relatable using SQL and common data field.

To illustrate this, one approach we used for extracting, parsing and ensuring relatability of socioeconomic data at the Health Region level extracted from CANSIM was as follows:

1. Determine the geographic region (Provinces/Territories, Health Regions, etc.) and date (2011, 2006, 2001, etc.) that we are interested in and download the corresponding Census Boundary Shapefile.
2. Identify a Census Table or data set organized by an appropriate geographic region.

10 Statistical agencies that cater to specific regions or provinces/territories tend to present national data alongside regional estimates for comparison purposes.

Illustration 8: CANSIM – Manipulate Data Tab Showing Available Geographic Organization Units

3. Using the "Manipulate Data" tab, select the level of geographic detail (Health Regions) that we intend to illustrate.
4. Organize the output data with the setting "for database loading" and save data as a Comma-Separated Values (CSV) file.

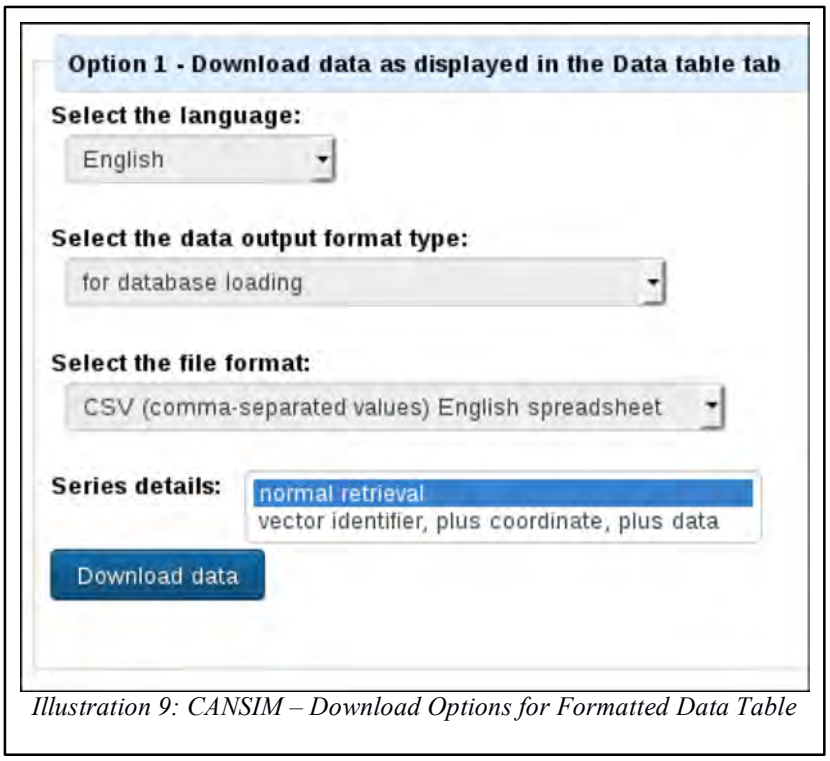


Illustration 9: CANSIM – Download Options for Formatted Data Table

5. To ensure that socioeconomic information corresponds with the geographic boundaries, create a new field with a corresponding geographic identification code (GEO ID) for each entry.[11]

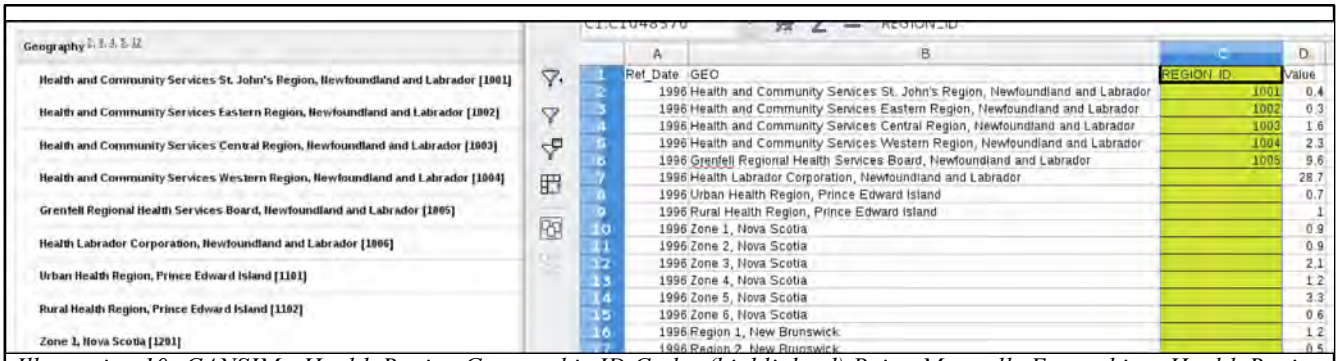


Illustration 10: CANSIM - Health Region Geographic ID Codes (highlighted) Being Manually Entered into Health Region level Socioeconomic Data

6. Optional: For multiple tables, copy and paste the relevant values from each column and create a new table with the "NAME", "GEO ID", "VALUE_1", "VALUE_2", …, "VALUE_n" columns and save the new file as a CSV.

The sourcing of multiple datasets reflects another lesson for GIS and GIS education initiatives: as with the use of maps, the use of statistical data should be problematized with reference to its limitations. As socially-constructed artifacts, statistical datasets reflect choices made in the collection, formatting and presentation of data. These decisions (and the workarounds that we use to address them) must be made as transparent as possible. This allows audiences to critically assess the data sets presented in cases where a community of interest is utilizing them for a defined purpose, such as mapping them onto a geographic region to illustrate a public policy argument.

11 In the example here, the Health Region Unique ID needs to be manually entered to ensure that the CANSIM table and the Census Boundary File share a matching field. This step might not always be necessary.

Step 3: Collection and Geo-visualization of Information and Communication Technology (ICT) Quality Indicators

Alongside socioeconomic figures, we also assembled the availability, throughput, and cost of consumer-level broadband services. As noted above, data in this area is currently limited. At the provincial levels, many of the most recent figures relating to Internet service providers, access cost and user demographics for Northern and Aboriginal communities were derived from year 2013 data held by the Conference Board of Canada (Centre for the North, 2013). Additionally, in this example, we point to the Statistics Canada "Survey of Household Spending on Household Operation, Internet Service" (Discontinued in 2009) as the most extensive and accurate data determining Internet access and service costs by region (recognizing that it is limited).

The primary means of determining the more technical dynamics of the availability of consumer Internet access is through the use of more recently acquired Internet performance test data. Internet performance tests examine numerous quantitative and qualitative indicators derived from the speed and accuracy of data packets sent and received between a user's location and a test server. Throughput measurements, which take into account not only maximum bandwidth capacity but also system hardware performance and, most importantly, latency from physical distance between users and servers are therefore a reliable indicator of Internet quality experienced at the user level, especially so for rural and remote users. For these purposes, our analysis relies on the Network Diagnostic Test [NDT] data for Internet performance for download and upload throughput maintained by M-Lab (Measurement Lab, 2015) and CIRA (Canadian Internet Registration Authority, 2015).

The NDT test data maintained by these agencies is openly accessible to researchers with very few restrictions (CC0 License; Creative Commons, 2015). It is updated daily, and currently provides available data for years 2009 to present. Among numerous different measurements, each individual record also preserves the IP address and geographic location (latitude and longitude coordinates) of the test to allow for geospatial visualization and analysis. However, we stress that these initiatives are currently in the early stages of use, and so do not yet provide a robust dataset from which to draw inferences. We note this qualification to point out the limitations of the visualizations that we generated through this process and present in this report.

As of the time of this writing, the method for collecting this data involves using Google Cloud Services platform and a Google-developed specialized SQL-like language called "BigQuery" (Google, 2016). This cloud infrastructure service allows one to perform specific and resource-intensive queries on very large Google-hosted datasets through a web platform.[12] The method for gaining access to M-Lab data through BigQuery involves registration, opting-in to the Google Cloud Services platform and

12 Other methods of utilizing BigQuery are through interoperability with platforms such as JAVA, Python, R, and Unix-based command-line shells.

performing a BigQuery SQL operation to return the data desired.[13] In the event that direct data access is desired, the raw test record data can be downloaded directly, organized by date, server and test type.

In order to visualize M-Lab data at the national scale, a number of steps were necessary which involved formatting the data using QGIS. During the data gathering phase, we ensured that the NDT test data records were exclusively for areas within the region of interest (Canada). We then performed the following steps:

1. Load the table of M-Lab NDT data points into QGIS[14], specifying that the data contained geographic coordinates, and then save this as a shapefile to allow for processing and manipulation in QGIS.

Illustration 11: Loading a CSV Table of NDT Test Data in QGIS

(Note that X column is set as "Longitude" and Y column is set as "Latitude")

2. We then load the Canada Revenue Agency Income Tax Zone shapefile and perform a "Select by Location" analysis, creating a new shapefile of the data points that fall within the area of interest.

13 While the process of gaining access to the M-Lab data set is not explored here, data and SQL queries used to parse geolocated throughput measurements for year 2014 are openly available through Concordia University's Spectrum Research Repository available at: http://spectrum.library.concordia.ca/980168/ (McKelvey, 2015).

14 As of version 2.14, QGIS handles Comma-Separated Values tables (.csv) as well as Open Document Format (.ods) and Excel/Office Open XML spreadsheet tables (.xls and .xlsx) natively.

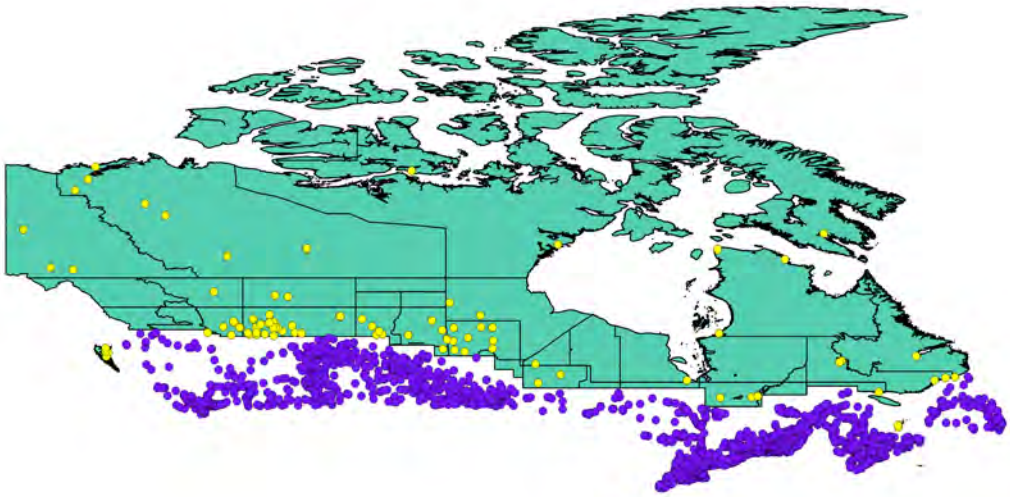


Illustration 12: All M-Lab NDT Test Points Across Canada Overlaying Shape of Canada Revenue Agency Income Tax Zones

(Purple = Points Outside Area of Interest; Yellow = Points Within Area)

3. With this data prepared, we can perform a point-in-polygon aggregate analysis with shapefiles of Provinces/Territories, Health Regions, Census Dissemination Areas, or other boundary. By performing a "Join Attributes by Location" with the boundary file as the target, specifying that intersecting attributes be summarized, the resulting file preserves spatial boundaries while also containing max/min/mean values and aggregated counts for all intersecting M-Lab NDT tests.[15] A processing model to visualize this sequence is shown below.[16]

15 An additional process performed in this model was to remove hexagons with fewer than 100 tests performed within them. In the case of Southern regions this is necessary to ensure statistically accurate representation, however for Northern and remote regions, this results in a much smaller data set with fewer hexagons due to the lower frequency of tests performed.

16 Processing models allow for visual representation and execution of a chain of tasks and a degree of automation for performing the same tasks on multiple sets of data. For more information on processing models see: http://www.qgistutorials.com/en/docs/processing_graphical_modeler.html (Gandhi, 2016)

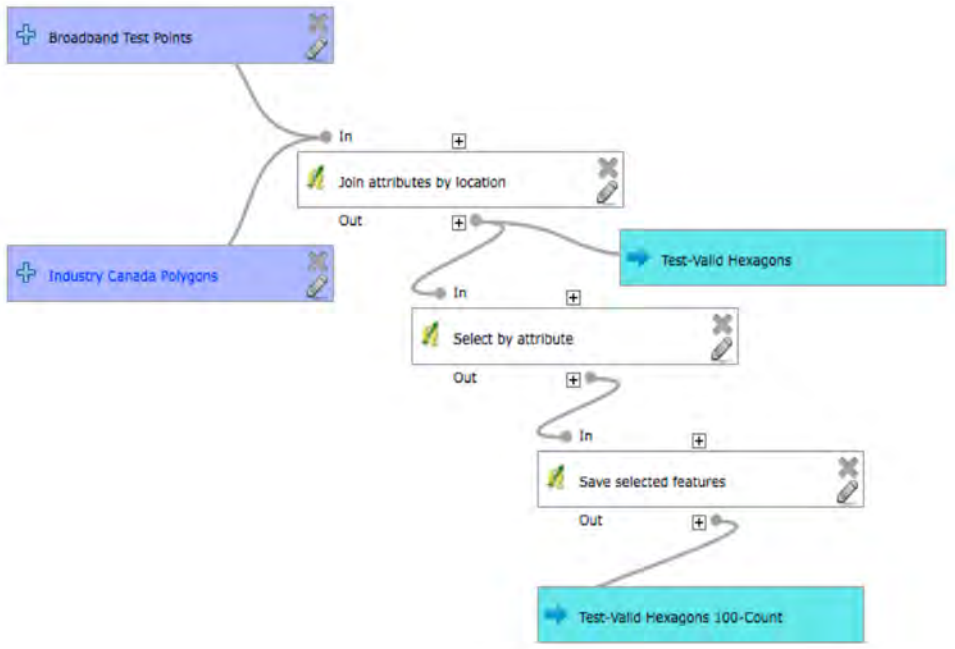


Illustration 13: Graphical Processing Model for Joining M-Lab NDT Data to Shapefile and Performing Additional Functions

One option for visualization at the municipal scale is to use service area hexagons made available from Industry Canada through the *Digital Canada 150* program (Industry Canada, 2010). After performing a "Join by Location" process, the new geometry can then be stylized with graduated categories to display the various aggregated values for ICT indicators according to values now joined to the Shapefile.[17] An example of this output can be seen below.

17 An informative resource for learning how to create, work with, and style vector and raster data in QGIS can be found at: http://www.qgistutorials.com/en/docs/basic_vector_styling.html (Gandhi, 2015a). A brief guide for styling vector data according to chloropleth style is explained in section 4.

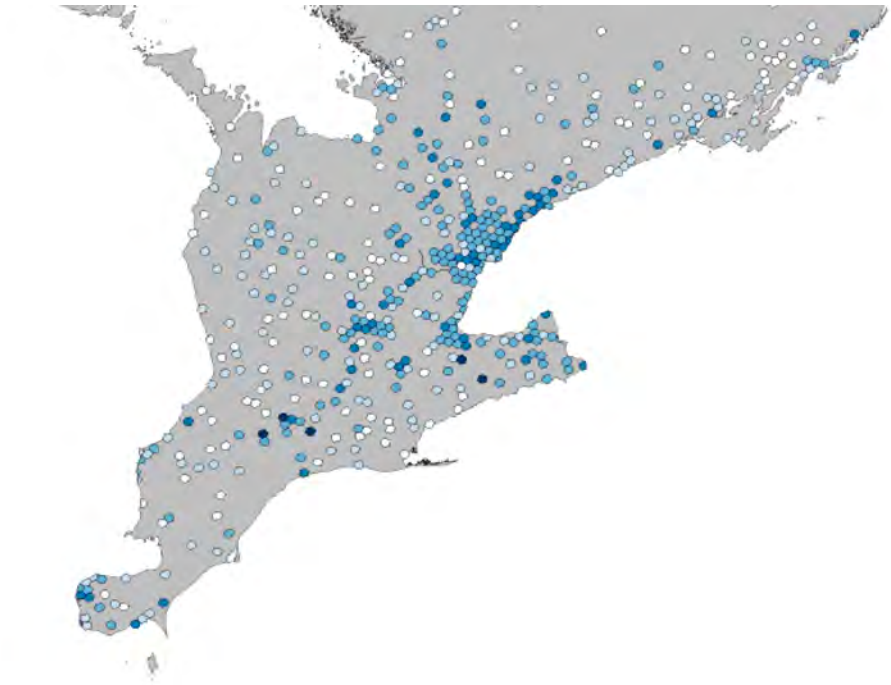


Illustration 14: Example of Municipal-level Industry Canada Hexagons with Joined and Stylized M-Lab NDT Data for Southern Ontario

(Darker Blues = Faster Throughput; Lighter Blues = Slower Throughput)

4. ***Optional:*** A final step to aid in presenting these results at a coarser scale involves converting the polygon Shapefile with joined M-Lab NDT test data into a point Shapefile. The function "Polygon centroids" can be used on the Shapefile to calculate the center of each polygon and assign points at those locations, preserving the joined M-Lab NDT test data. This data can then be combined with the socioeconomic data presented in the next step to help visualize Internet accessibility and quality alongside socioeconomic information.

Step 4: Joining and Representing Spatial and Statistical Data

The fourth step in our work-flow process involved joining our statistical data sets with the spatial data representing our geographic community of interest. This allows researchers to measure and track inequities in broadband access and affordability by generating visuals that layer aggregated data from statistical agencies onto maps of our northern communities of interest. Working through this process involved many points of reflection that illustrate some digital literacy considerations that we encountered in this work.

For the following steps, we present two methods of performing data joins, relating tabular data with geographic data using a common element, in this case a Geographic ID. The first approach relies solely on QGIS and follows through the general process of creating a static map representation while the second approach considers the usage of CARTO, an on-line web-GIS platform designed for interactive representation of geographic data. For both examples, we present a chloropleth (shaded polygon) map of socioeconomic data for Health Regions mixed with graduated categories for point values of Internet throughput data organized by Health Region to show the variety of available symbols for data representation

The QGIS Approach

The QGIS approach involves joining both statistical data and spatial data, transforming the data to an appropriate map projection, and adding cartographic elements to produce a map to do this. We began by loading our geographic data into QGIS in a new project file. For representation purposes, we present geographic points of Internet throughput tests alongside our geographic boundary files. Throughput tests are measures of the upload and download speeds between users and servers. For the analysis of these tests, we followed the steps outlined in Part 3 and performed a point-in-polygon analysis using a Health Regions boundary file then calculated the midpoint (centroid) of each regional boundary (points in yellow). This data was then overlaid with the "Northern" and "Intermediate" zone Health Region boundary (areas in purple), with a Shapefile for Provinces layered beneath it (areas in red).

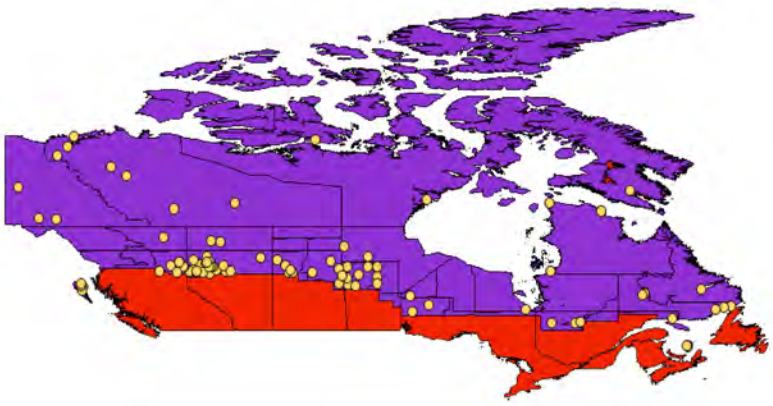


Illustration 15: Raw Output of Joined and Formatted M-Lab and Health Region Data with centroids, NAD83 Geographic Coordinate System

We then added the socioeconomic data to the workspace alongside the geographic data. Both the geographic boundary data and socioeconomic tabular data contain a common field. The reason for this is to perform a tabular join process to relate the socioeconomic data to the geographic data. In this fashion, a new geographic layer is produced with numeric and string values from the socioeconomic data that can be represented spatially. The process for performing this in QGIS is as follows:

1. Load both the geographic boundaries and socioeconomic tabular data, ensuring that a common field exists between both data sets.

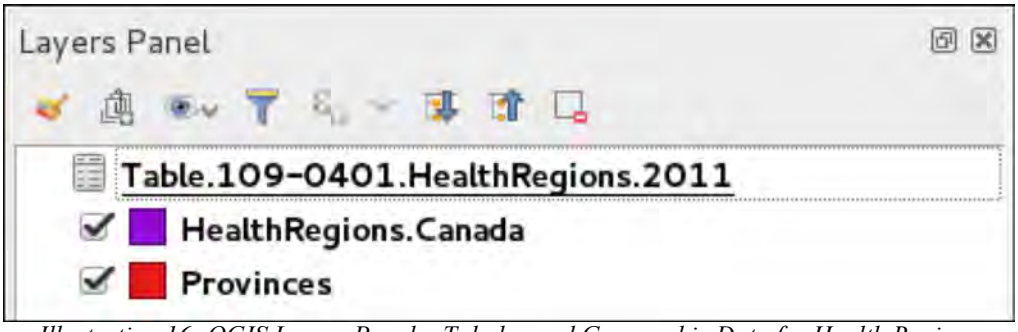


Illustration 16: QGIS Layers Panel – Tabular and Geographic Data for Health Regions

2. In the Processing Toolbox, search for the tool called "Join attributes table" and double-click to open the prompt.

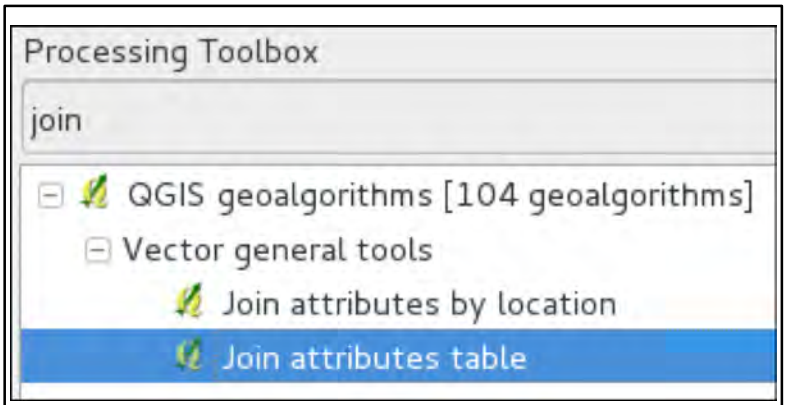


Illustration 17: QGIS Toolbox – "Join attributes table" Function

3. Specify the first input file (the geographic boundary file), and the second input file (the tabular socioeconomic data), the common fields (HR_UID, in this case), and the output file ("OutputLayer.shp") and run the process.

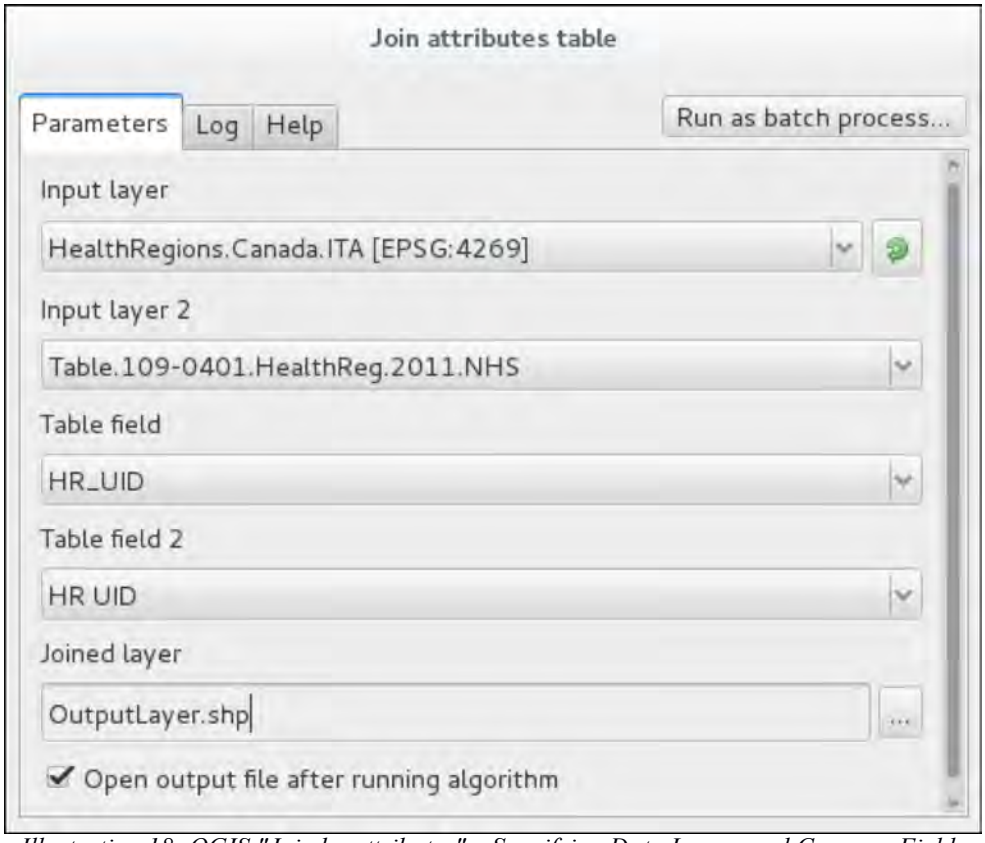


Illustration 18: QGIS "Join by attributes" – Specifying Data Layers and Common Fields

4. Repeat this process as necessary for each compatible table and geographic boundary file, ensuring to save the new file with an appropriate filename.[18]

It is now possible to manipulate the geographic layer to show categorized data and graded values based on the newly joined data. By specifying non-overlapping graduated colour schemes for both the point- and polygon-based data to differentiate data categories, we can differentiate between the data sets. This is possible by accessing the "Properties" option of the joined data layer by right-clicking on the new file and specifying "Graduated" style for the appropriate data column. Additional options such as Binning Mode (Equal Interval, Natural Breaks (Jenks), etc.), Colour Ramp, Legend Style, and Number of Classes can also be specified at this time.

18 For most GIS databases, it is important to not include spaces or use special/non-UTF8 characters when specifying the filename. The reason for this is that many geoprocessing tools will encounter errors as their syntax relies on simplified terminal commands when being compiled. These same rules apply to the file-path and folder names for input/output files.

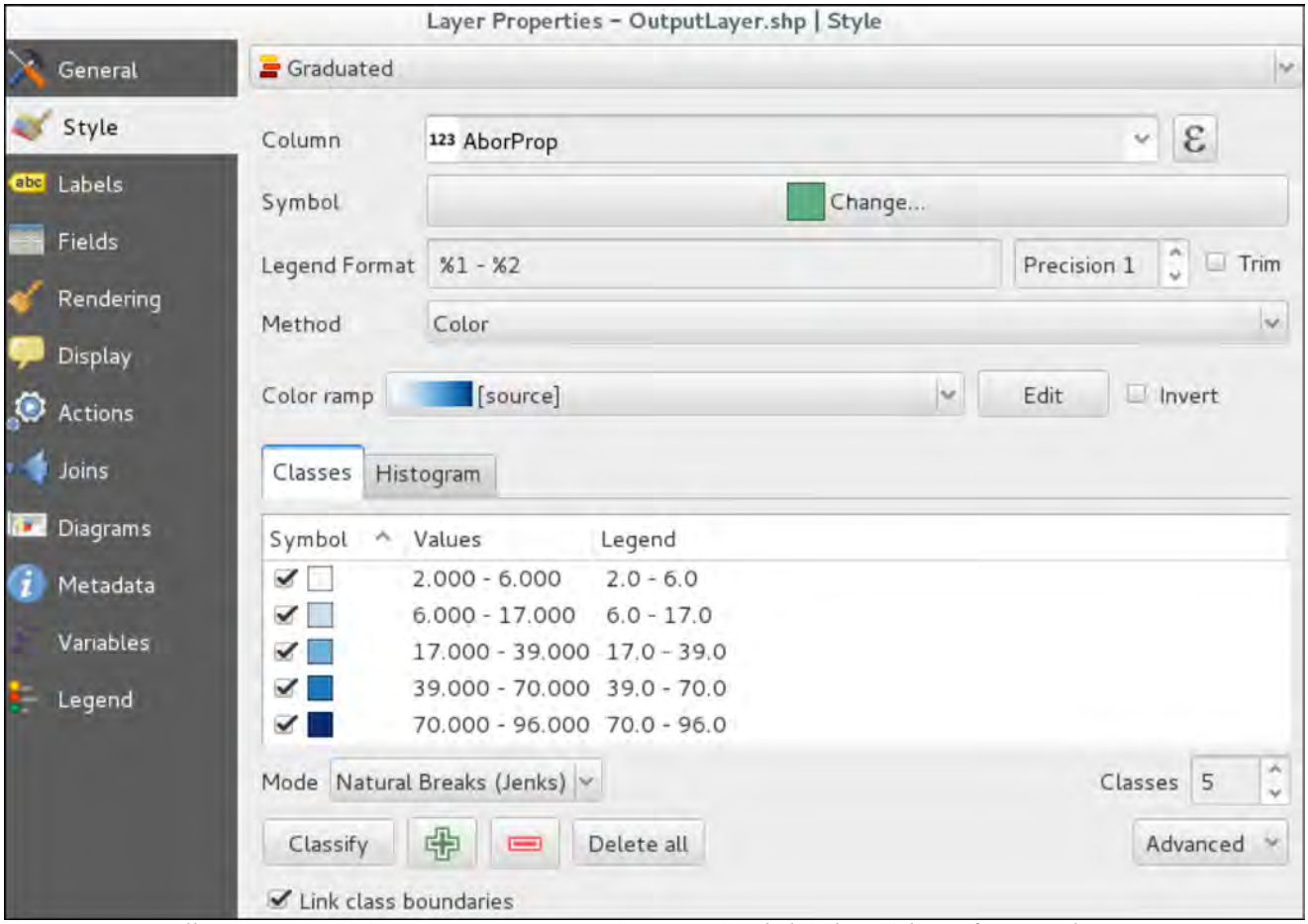


Illustration 19: QGIS Layer Properties – Setting a Graded Colour Scheme for Joined Data

We then transform the geographic data with an appropriate map projection. A projected coordinate system appropriate for Canada as a whole is the Statistics Canada Lambert Conformal Conic projection (EPSG: 3347). This geographic transformation ensures that the relative shapes of each province and territory are preserved and is the standard projection for presenting Canada in maps at the national level.

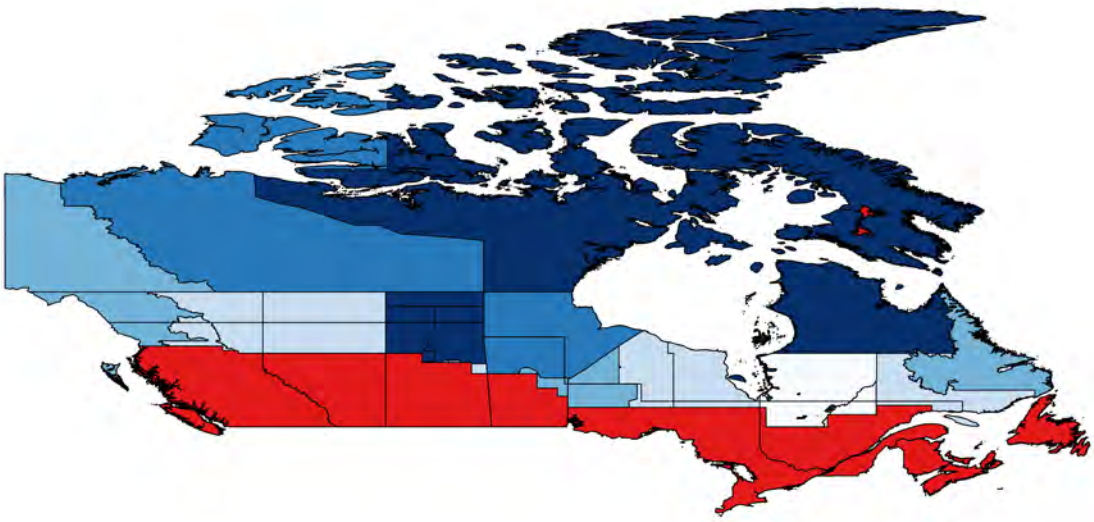


Illustration 20: QGIS - Graded Health Region Data in the NAD83 Geographic Coordinate System (EPSG:4269)

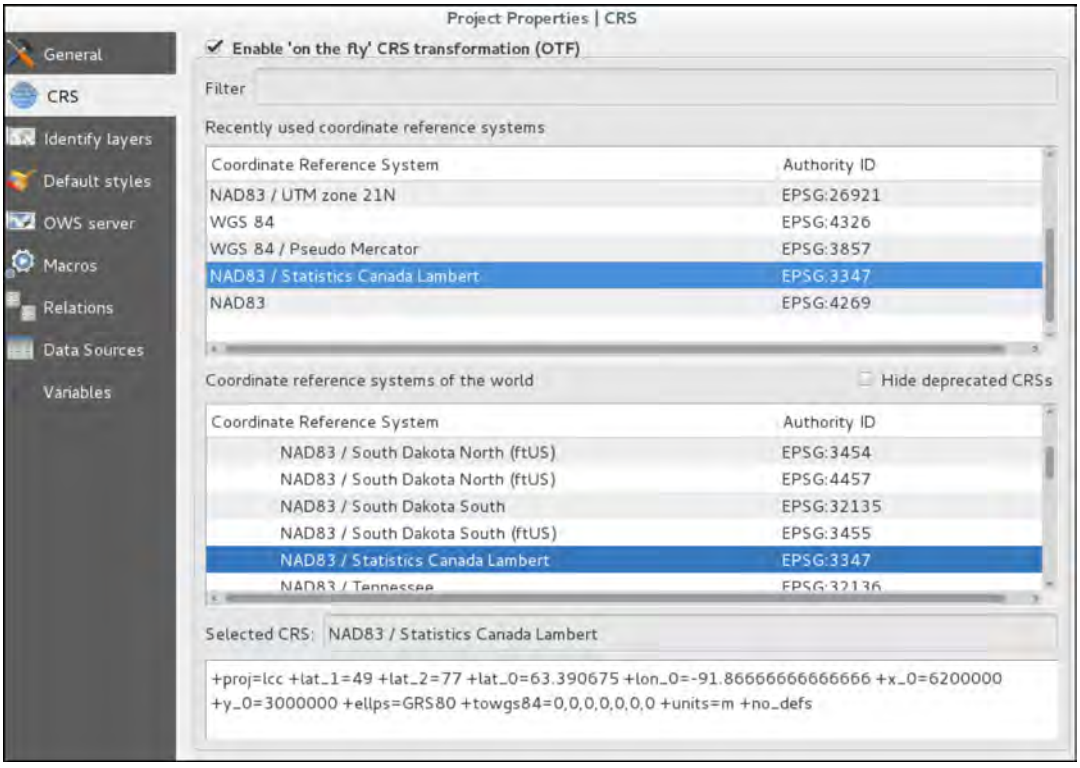


Illustration 21: QGIS Projection Properties – Specifying the Statistics Canada Lambert Projection (EPSG: 3347)

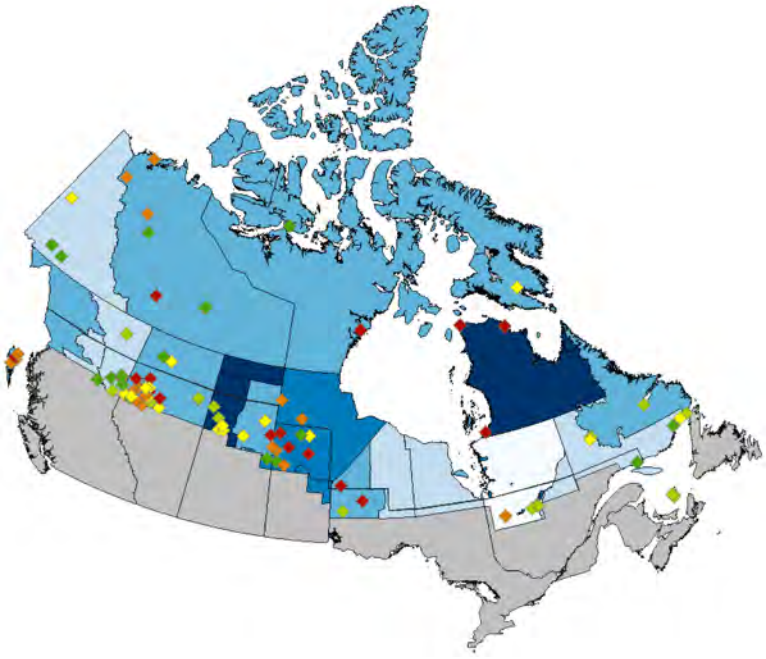


Illustration 22: QGIS – Graded M-Lab NDT Data and Health Region Data in Statistics Canada Lambert Projection (EPSG: 3347)

The final step in the process of static map-making involves adding explanatory elements to the map, as shown in the illustration below. Cartographic elements generally include a title, North arrow, scale bar, and legend. Content descriptions typically include references to data sources, attributions of the developers, software used, the date the map was created and other necessary metadata to correctly interpret the resulting map.

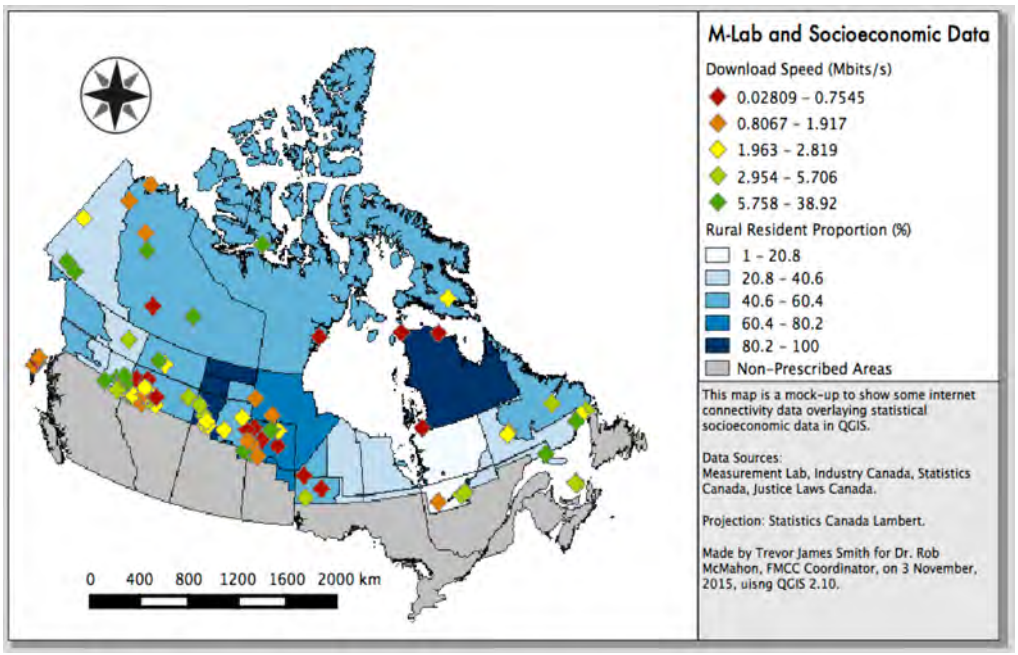


Illustration 23: Chloropleth Map of M-Lab and Health Region data with Cartographic Elements Added

Following the formatting of GIS data with socioeconomic information, it is relatively simple to generate new data to produce other maps - although different types of data call for different methods of processing and presentation (see Steps 2 and 3). For example, the figure below provides an example of two maps illustrating Provincial and Territorial Average Household Expenditure Ratios for Food and Internet Access Spending. Both of these maps illustrate the higher proportion of income spent on Internet Access vs. Total Spending on Communication Services for households in the northern territory of Nunavut. These figures standardize the data between territories and provinces by examining the proportional differences between expenditures, rather than the dollar amounts that vary between provinces by consumer price indexes.[19] This example illustrates how a mapping representation must reflect the characteristics of the available data as well as the communication goal(s) of the community of interest.

19 The data sources for these figures are derived from Statistics Canada's CANSIM database and 2011 Census Boundary Files (Statistics Canada, 2015a, 2015b, 2015c, 2015d).

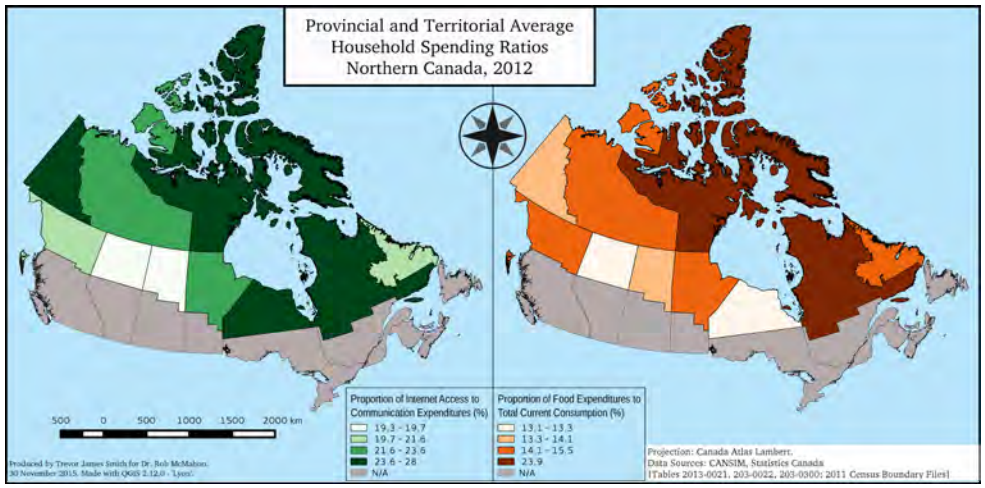


Illustration 24: Sample Map highlighting Provincial and Territorial Average Household Expenditure Ratios for Food and Internet Access Spending for year 2012.

The CARTO Approach

A similar approach to using the QGIS method is to perform the join functions on-line through a web-GIS platform. For this example we show the processing steps required for performing a table join using CARTO, a popular web-GIS platform that offers a simplified set of geoprocessing tools and provides interactive representation capabilities (https://carto.com/). The basic version of this platform is free to use with some restrictions (such as no private data sets and some storage limitations) and does not require the installation of additional software beyond a web browser. We present a methodology for performing a tabular join here and briefly mention a few of the capabilities of the web-GIS platform over traditional GIS methods.

Beginning with the Geographic boundary file and tabular socioeconomic data file with common fields, the data must be uploaded to one's CARTO account. This can be performed by "dragging and dropping" CSV of socioeconomic data and a compressed archive (zip) of the geographic Shapefile into the account dashboard. Upon inspection of the uploaded files, tables will be shown.

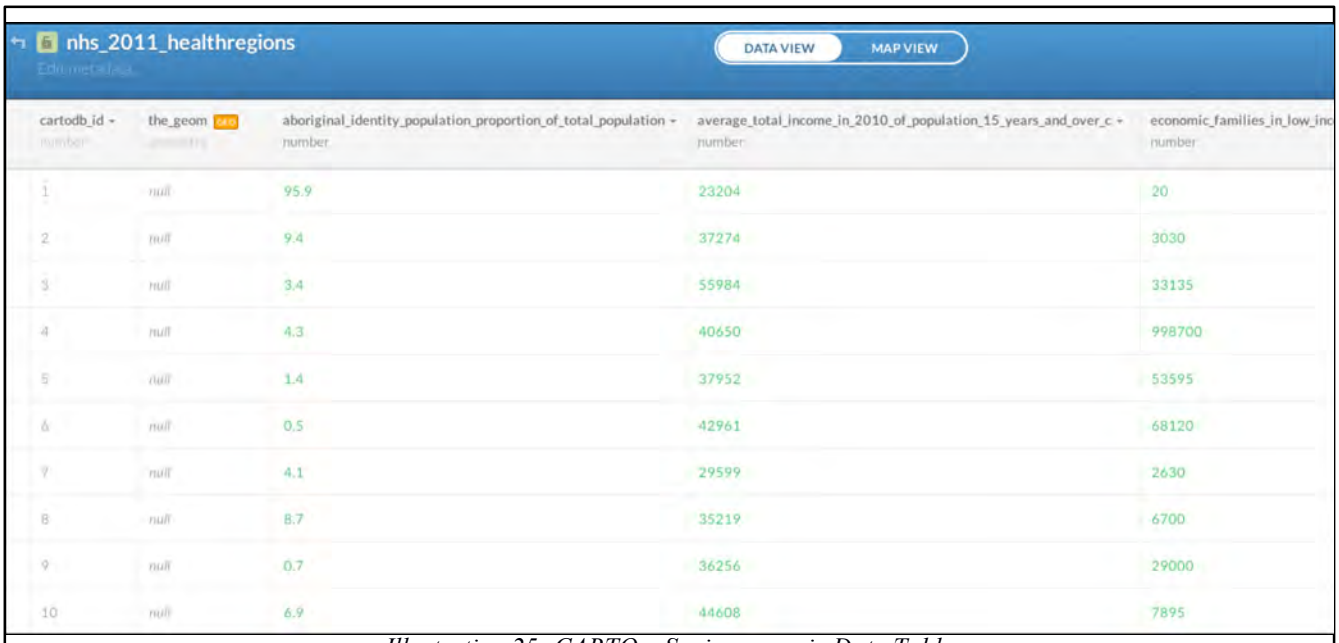


Illustration 25: CARTO – Socioeconomic Data Table

(Note that the Geometry field ("the_geom") is populated with NULL values)

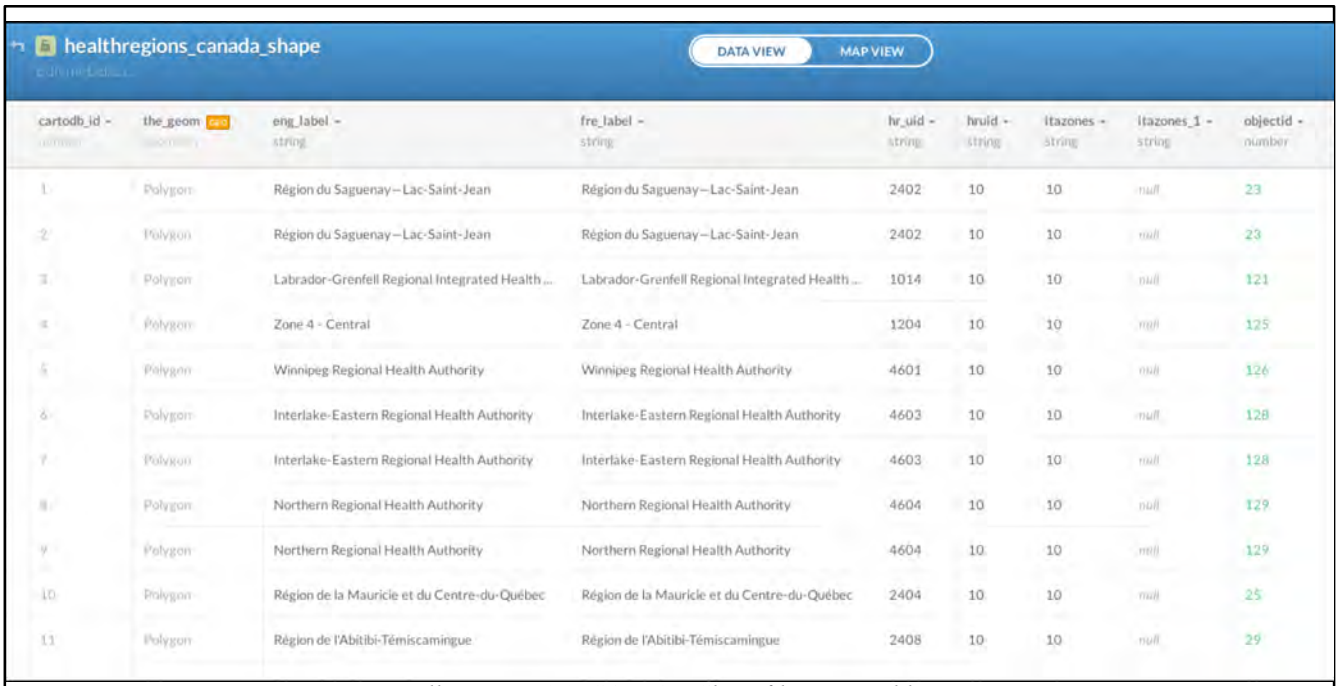


Illustration 26: CARTO – Shapefile Data Table

(Note that the Geometry field ("the_geom") is populated with "Polygon" values)

When focused on a data table, by clicking on the “Edit” function in the top-right, one can select the option to “Merge with dataset”. This will bring up a wizard tool to perform a “Column Join” between the present data set and another data set within the account library. This is functionally the same as performing a table join.

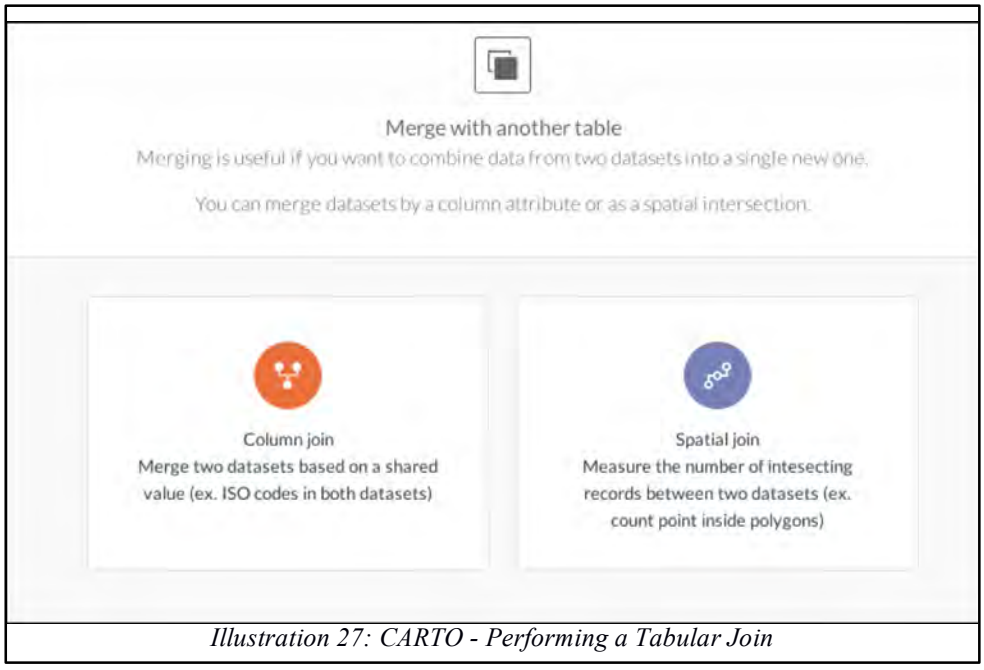


Illustration 27: CARTO - Performing a Tabular Join

After specifying the common field (“HR_UID”) the next option allows for selective data field joining. This option ensures that only the data that one wants to present in the map is appended to the geographic shape. The fields that are enabled (coloured green) from both data sets will be preserved in the joined data set. By clicking to confirm and specifying a new name, the joined data set will be created and added to the account's “Data” section.

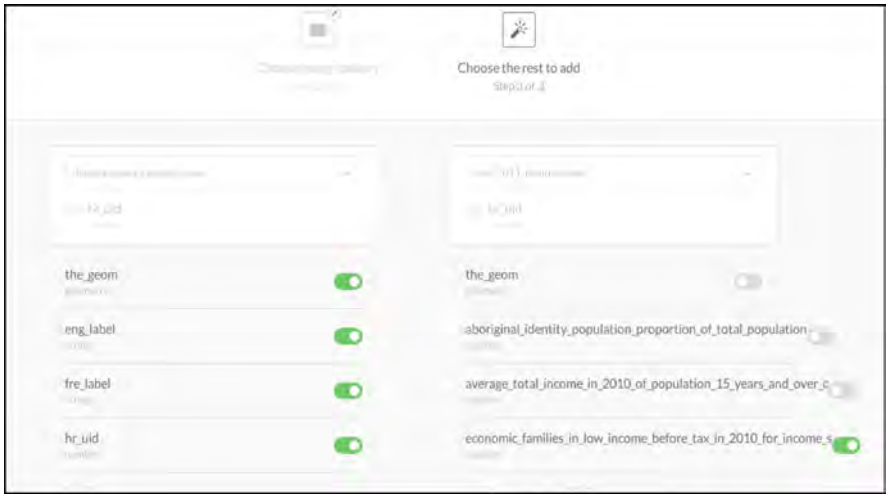


Illustration 28: CARTO - Appending Socioeconomic Data to Geographic Shapefile

The next step is to style the new data set according to the communication needs of the users. By clicking on the “Map View” tab at the top of the new data set, one can set the style and categorizations by clicking the “Map Layer Wizard” tab. Options for categorization styling methods are available with example graphics previewing what the resulting categorization style will look like when translated to the data set. Many of the same options available through QGIS are presented here.

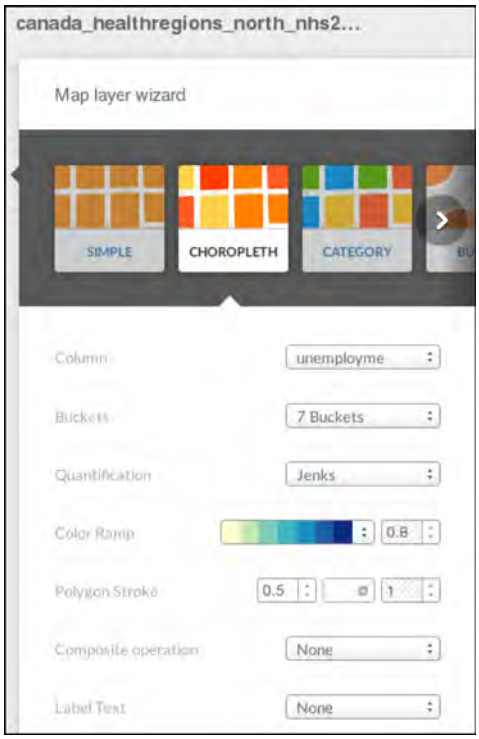


Illustration 29: CARTO – Map Layer Wizard Tab

After selecting the styling options and setting the fields that one wishes to display, the end result will show an interactive map with options to enable panning, location look-ups, and modifiable fields and legends. This end result can be shared via several methods by clicking the “Publish” tab. Methods of embedding the map in a website, on-line applications, or via direct file-sharing are made available to the user.

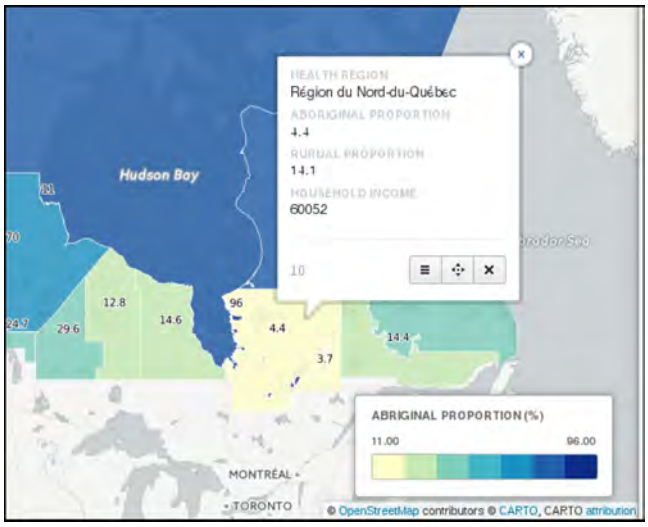


Illustration 30: CARTO - Interactive Health Region Shape with Socioeconomic Data and Legend

Technological Considerations: On-line Data Visualization, Web-GIS Platforms, and Data Ownership

At present, FMCC researchers are considering how best to adapt and integrate web-GIS and other on-line data analysis platforms to support our data visualization goals. Benefits of using these platforms include the ability to dynamically layer and represent many different fields of data for individual geographic objects; to animate changes over time (torque mapping); to continue to simplify GIS work-flows for non-specialists; and to host interactive, modifiable, and shareable maps.

Many community mapping projects provide examples of how web-mapping tools like Google Maps and CARTO (as well as JavaScript libraries such as Leaflet and OpenLayers), and on-line geo-data visualization services (such as PostGIS, GeoNode, and GeoServer) can deliver geospatial data to users and communities more interactively. On-line resources (and support documentation) for these platforms are accessible to users of many differing degrees of specialization, ranging from those with basic knowledge of spreadsheet functions, to professional data scientists and web developers. Depending on the needs of the community of interest, the digital literacy of the database managers, and whether the hosted data is to be made private or publicly available, many cloud-based or open source options can be pursued.

We are now exploring the use of these platforms to simplify the technical expertise needed to work with socioeconomic and geographic data held by third party organizations such as statistical agencies. Easily interpretable user interfaces, such as those seen in CARTO, Google Maps, or MapBox, allow for a more approachable platform for non-specialists to visualize their data. A helpful categorization system presented by Bhargava and D'Ignazio (2015) examines some of these data management tools according to their learn-ability and flexibility. FMCC has performed some initial studies of how reliable these platforms are in terms of their learning curves and GIS capabilities. The figure below illustrates an example of this approach in CARTO.

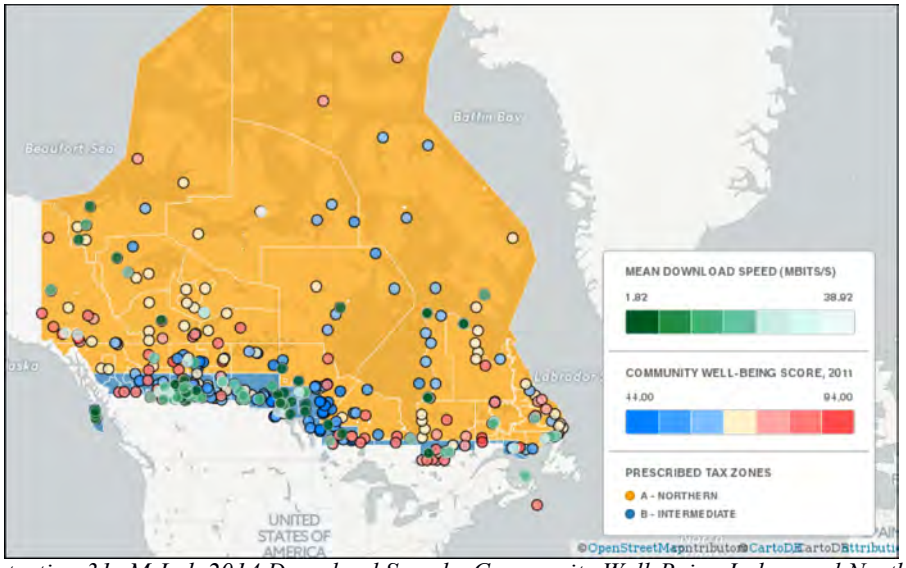


Illustration 31: M-Lab 2014 Download Speeds, Community Well-Being Index, and Northern Census Subdivisions 2011, visualized in CARTO

Although cloud-based platforms tend to present the most user-friendly interfaces, these tools have several disadvantages. These include:

- Limited web platform flexibility (assuming the user has limited knowledge of JavaScript, CSS, and other styling languages)
- Removal of the user from the data management process ("black box")
- The potential for added costs associated with scaling (e.g. additional cloud-based storage)
- Third-party access to (and control of) potentially sensitive or confidential community data

Alternatively, open source JavaScript libraries and geo-data server options afford users complete control over their information and presentation methods. They also provide more affordable data storage limits and knowledge acquisition opportunities. Some disadvantages include a steep learning curve and a requirement of basic expertise in computer programming and data science. Again, solutions for the best platform will depend on community needs, financial and technical resources, and user and manager capacities and skill sets. FMCC will continue to examine these options and work closely with member organizations to aid them in meeting their GIS and web-GIS needs. Development of community capacity for performing GIS and mapping initiatives must be founded within and reflective of community needs and must also be examined critically as it pertains to ownership and control of information.

Building Sustainability in GIS Platforms

These observations point to the need for critical digital literacy in GIS initiatives. Importantly, this literacy must include the deconstruction of GIS practices and technologies, as well as their adoption and adaptation by community-based practitioners. It is not enough to simply engage community partners in GIS mapping initiatives - these participants must also be supported to critically analyze, de-construct and re-construct the social practices and technical tools utilized in GIS. In this way, GIS

initiatives can be taken up and re-shaped at various stages of development and usage - not only at the 'end point' of sharing and explaining completed visuals.

To support this kind of GIS-focused digital literacy, some researchers have proposed an approach influenced by Paulo Freire, and specifically his encouragement of mutually-beneficial knowledge partnerships between educators and learners (Tygel & Kirsch, 2015). For Bhargava and D'Ignazio (2015), this involves producing inclusive, relevant and adaptable GIS tools in partnership with the communities of interest seeking to use them.

The process that we describe in this paper provides several lessons that may inform these digital literacy initiatives. Data advocacy initiatives must incorporate both technical literacy and critical reflection on the social contexts that data is both embedded within and illustrative of. Our work-flow design process benefited from sustained consideration of the social construction of both the community of interest and the data sets that we draw on and the GIS tools we used. Our treatment of these factors supported our efforts to expose and address some of the political choices inherent in the geospatial data presented in our final maps.

We agree with Tygel and Kirsch (2015), who stress that digital literacy must start from the lived reality of involved communities of interest, and extend out from there. Critical digital literacy initiatives work with involved communities of interest to identify ways that data articulates with concrete aspects of their lives. This highlights the need to involve communities of interest in the collection, interpretation and presentation of data through GIS platforms.

However, a key challenge throughout our project has been the distance - both physical and disciplinary - between the involved communities and the GIS designer working to interpret their needs and understandings from afar. Along with a lack of exposure to the lived realities of people in Indigenous communities, the designer found it challenging to produce maps and reports for audiences who lack GIS training. It is very difficult to build a work-flow for communities unless the designer is working directly with them.

GIS and cartography are time-consuming to learn, given the technical complexity of the platforms, work-flow and theory. Efforts are further limited by technical considerations, such as available processing power, data storage, and tools to support data analysis, such as statistical software or data-handling algorithms and organizational capacity limitations such as reliable access to research funding, human technical resources and organizational adaptability (Sieber, 2007).

As a result, it can be challenging to produce accessible, legible and accurate maps. To address this challenge we focused on developing and documenting an accessible methodological approach to act as an explanatory bridge between involved parties. Tygel and Kirsch (2015) discuss how digital literacy involves connecting the thematic focus generated by communities of interest to the codification of data in outputs such as statistics, graphics and tables.

Ideally, such a process may lead to increasingly complex understanding of the topic being illustrated through geospatial data. This approach is reflected in the development of the work-flow process described in this paper, which evolved from constructing a geographic community of interest, to building standardized data sets to illustrate socio-economic and broadband performance indicators, to combining these elements into more complex maps that may be shaped into on-line participatory mapping tools.

We stressed above the necessity of more valid, accurate data on broadband access and availability in Northern and remote regions. We recognize the work being done by parties such as the CRTC and CIRA in this area, and will look to those sources for future visualization projects that utilize the methodology described in this paper.

As a final point, Tygel and Kirsch (2015) stress that digital literacy involves the systematization of process: beyond simply merging data and information about an issue, it is the exercise of theorizing and deeply analyzing an experience of digital literacy. With this in mind, we present this paper as an early-stage effort in systematizing our geospatial data advocacy project.

References

Bhargava, R., & D'Ignazio, C. (2015). Designing Tools and Activities for Data Literacy Learners. Presented at the Data Literacy Workshop: Web Science 2015 Conference, Oxford, Britain. Retrieved from http://www.dataliteracy.eita.org.br/wp-content/uploads/2015/02/Designing-Tools-and-Activities-for-Data-Literacy-Learners.pdf

Canadian Internet Registration Authority. (2015). CIRA Home Page. Retrieved August 19, 2015, from http://cira.ca/

Centre for the North. (2013). *Mapping the Long-Term Options for Canada's North: Telecommunication and Broadband Connectivity.* (pp. 1–85). Canada: The Conference Board of Canada.

CPHA. (2010). *The Impact of Cancelling the Mandatory Long-Form Census on Health, Health Equity and Public Health* (Presentation) (p. 8). Ottawa, Ontario: Canadian Public Health Association.

Creative Commons. (2015). About CC0 — "No Rights Reserved." Retrieved August 24, 2015, from https://creativecommons.org/about/cc0

Dobson, J. E. (1983). Automated Geography. *The Professional Geographer*, *35*(2), 135–143. https://doi.org/10.1111/j.0033-0124.1983.00135.x

ESRI. (1998). ESRI Shapefile Technical Description. Environmental Systems Research Institute.

ESRI. (2015). *ArcGIS Geographic Information System*. Redlands, CA, USA: Environmental Systems Research Institute.

Free Software Foundation. (2015, September 1). What is free software? - GNU Operating System Philosophy. Retrieved October 26, 2015, from https://www.gnu.org/philosophy/free-sw.html

Gandhi, U. (2015a). Basic Vector Styling. Retrieved October 27, 2015, from http://www.qgistutorials.com/en/docs/basic_vector_styling.html

Gandhi, U. (2015b). Digitizing Map Data. Retrieved October 28, 2015, from http://www.qgistutorials.com/en/docs/digitizing_basics.html

Gandhi, U. (2016). Automating Complex Workflows using Processing Modeler. Retrieved November 6, 2016, from http://www.qgistutorials.com/en/docs/processing_graphical_modeler.html

Goodchild, M. F. (1993). Ten Years Ahead: Dobson's Automated Geography in 1993. *The Professional Geographer*, *45*(4), 444–446. https://doi.org/10.1111/j.0033-0124.1993.00444.x

Google. (2016, November 6). What is BigQuery? [BigQuery Documentation]. Retrieved November 6, 2016, from https://cloud.google.com/bigquery/what-is-bigquery

Harris, K. (2015, November 6). Mandatory long-form census restored by new Liberal government. Retrieved January 7, 2016, from http://www.cbc.ca/news/politics/canada-liberal-census-data-

1.3305271

Industry Canada. (2010, October 18). Digital Canada 150 - Home [home page; Home Pages]. Retrieved August 19, 2015, from http://www.ic.gc.ca/eic/site/028.nsf/eng/home

Justice Laws of the Government of Canada. Consolidated Income Tax Regulations, C.R.C., c.945 § 7303.1 161 (2015). Retrieved from http://laws-lois.justice.gc.ca/eng/regulations/C.R.C.%2c_c._945/page-161.html

Keenan, P. B. (2008). Geographic Information and Analysis for Decision Support. In *Handbook on Decision Support Systems 2* (pp. 65–79). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-48716-6_4

McKelvey, F. (2015, July 14). 2014 Results from the Network Diagnostic Tool in Canada as part of CRTC BSO Intervention, file number 8663-C12-201503186. Retrieved from http://spectrum.library.concordia.ca/980168/

McMahon, R., O'Donnell, S., Smith, R., Walmark, B., Beaton, B., & Simmonds, J. (2011). Digital Divides and the "First Mile": Framing First Nations Broadband Development in Canada. *The International Indigenous Policy Journal*, *2*(2). Retrieved from http://meeting.knet.ca/mp19/mod/book/view.php?id=1722&chapterid=2175

Measurement Lab. (2015). M-Lab Home Page. Retrieved August 19, 2015, from http://www.measurementlab.net/

Open Geospatial Consortium. (2015). KML. Retrieved August 19, 2015, from http://www.opengeospatial.org/standards/kml

QGIS Development Team. (2016). QGIS Geographic Information System (Version 2.16). Open Source Geospatial Foundation Project. Retrieved from http://qgis.osgeo.org

Sieber, R. E. (2007). Spatial data access by the grassroots. *Cartography and Geographic Information Science*, *34*(1), 47–62.

Snow, J. (1855). *On the Mode of Communication of Cholera* (2nd ed.). England, United Kingdom: John Churchill.

Statistics Canada. (2015a). Boundary Files, 2011 Census. Statistics Canada. Retrieved from https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2011-eng.cfm

Statistics Canada. (2015b). *National Household Survey indicator profile, Canada, provinces, territories, health regions (2014 boundaries) and peer groups, every 5 years (number unless otherwise noted)* (Census Table No. 109-0401).

Statistics Canada. (2015c). *Survey of household spending (SHS), household spending, regions and*

provinces, by household income quintile (Census Table No. 203-0022). Retrieved from http://www5.statcan.gc.ca/cansim/a26

Statistics Canada. (2015d). *Survey of household spending (SHS), household spending, three territories and selected metropolitan areas* (Census Table No. 203-0030). Retrieved from http://www5.statcan.gc.ca/cansim/a26

Tomlinson, R. F. (1962). Computer Mapping: An Introduction to the Use of Electronic Computers In the Storage, Compilation and Assessment of Natural and Economic Data for the Evaluation of Marginal Lands. In *Agricultural Rehabilitation and Development Administration* (pp. 1–9). Ottawa, Ontario: Canada Department of Agriculture.

Tygel, A., & Kirsch, R. (2015). Contributions of Paulo Freire for a critical data literacy. In *Data Literacy: Web Science 2015*. Oxford, Britain: EITA Cooperative. Retrieved from http://www.dataliteracy.eita.org.br/wp-content/uploads/2015/02/Contributions-of-Paulo-Freire-for-a-critical-data-literacy.pdf

Wikipedia. (2015a, August 24). Throughput. Retrieved August 24, 2015, from https://en.wikipedia.org/wiki/Throughput

Wikipedia. (2015b, September 25). Web Map Service. Retrieved September 25, 2015 from https://en.wikipedia.org/wiki/Web_Map_Service

Annex I – Glossary

Choropleth: An often-used style applied in thematic mapping and quantitative spatial data representation to denote the differences of intensity for statistical variables across geographically defined regions using progressive shades of a colour or a colour spectrum.

Comma-Separated Values (CSV): A tabular data file format that stores records in plain-text and is compatible with a wide array of spreadsheet software applications. Ideal for storing and loading individual data tables into GIS platforms, SQL databases, as well as for analysis in powerful statistical languages such as R, MATLAB, and Python.

Free and Open Source Software (FOSS): Software that conforms to four (4) essential freedoms: 1) "The freedom to run the program as you wish, for any purpose", 2) "The freedom to study how the program works, and change it so it does your computing as you wish", 3) "The freedom to redistribute copies so you can help your neighbor", and 4) "The freedom to distribute copies of your modified versions to others" (Free Software Foundation, 2015; "The Free Software Definition")

Geographic Information System (GIS): A suite of tools used to generate, visualize, query, analyze and modify spatial data. The fundamental idea for GIS was conceived in the 19th century (Snow, 1855) and later developed for land inventory purposes with the rise of computation (Tomlinson, 1962). Numerous desktop GIS platforms are available and nearly all web mapping platforms use some form of GIS in their delivery systems. The industry standard software for GIS analysis is 'ArcGIS' (ESRI, 2015), while its open source desktop equivalent is (arguably) 'QGIS' (QGIS Development Team, 2016).

Keyhole Markup Language (KML; file extensions: .kml, .kmz): Proprietary format for spatial data developed by Keyhole Inc. and a part of the Open Source Geospatial Consortium (OGC) implementation standard (Open Geospatial Consortium, 2015). The format is popular for its ease of integration with Web Feature/Map Services (WFS/WMS) and is the main spatial data format used by most Google applications.

Shapefile (file extensions: .shp, .shx, .dbf): Proprietary format for spatial data developed by ESRI (ESRI, 1998). Supports the representation of vector objects (points, lines, polygons) within a geographic or projected coordinate system, as well as querying of object attributes in a database management system.

Throughput: a measurement of internet quality based on a personal computer's ability to receive and retransmit data from a remote internet server through Transmission Control Protocol (TCP/IP) in as short a time as possible. A high value typically denotes a high internet bandwidth capacity, a reliable internet connection pathway, and a lack of infrastructural bottlenecks such as slow computer hardware, "Throttling" (intentional transfer speed constrictions), and network latency (a function of physical/geographic distance between a personal computer and server) (Wikipedia, 2015a).

Web Mapping Services and **Web Feature Services** (WMS/WFS): Online protocols for delivering interactive maps powered by GIS databases through the internet. Geographic data is housed in an online geodatabase and visualized according to user demands. Some client-side software systems that employ WMS/WFS include ArcGIS/QGIS, Google Earth/Maps, Tableau, CARTO, HERE, and MapBox (Wikipedia, 2015b).

Annex II – BigQuery SQL Statements

Upload Measurements:

“SELECT

STRFTIME_UTC_USEC((INTEGER(web100_log_entry.log_time) * 1000000), '%F %H:%M') AS day_timestamp,

web100_log_entry.connection_spec.local_ip,

web100_log_entry.connection_spec.remote_ip,

connection_spec.client_hostname,

connection_spec.client_geolocation.country_name,connection_spec.client_geolocation.city,

connection_spec.client_geolocation.postal_code,connection_spec.client_geolocation.latitude,

connection_spec.client_geolocation.longitude,

8 * web100_log_entry.snap.HCThruOctetsReceived/web100_log_entry.snap.Duration AS uploadThroughput,

web100_log_entry.snap.SegsRetrans/web100_log_entry.snap.DataSegsOut AS packetRetransmitRate,

web100_log_entry.snap.SumRTT/web100_log_entry.snap.CountRTT AS avgRTT,

web100_log_entry.snap.MinRTT,

web100_log_entry.snap.SndLimTimeCwnd/8 * (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd) AS congestionLimitedStateTimeShare,

web100_log_entry.snap.SndLimTimeRwin/8 * (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd) AS receiverLimitedStateTimeShare,

web100_log_entry.snap.SndLimTimeSnd/8 * (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd) AS senderLimitedStateTimeShare

FROM

[plx.google:m_lab.2014_01.all],[plx.google:m_lab.2014_02.all],[plx.google:m_lab.2014_03.all],[plx.google:m_lab.2014_04.all],[plx.google:m_lab.2014_05.all],[plx.google:m_lab.2014_06.all],[plx.google:m_lab.2014_07.all],[plx.google:m_lab.2014_08.all],[plx.google:m_lab.2014_09.all],[plx.google:m_lab.2014_10.all],[plx.google:m_lab.2014_11.all],[plx.google:m_lab.2014_12.all]

WHERE

connection_spec.client_geolocation.country_name = 'Canada' AND
IS_EXPLICITLY_DEFINED(web100_log_entry.snap.Duration) AND
IS_EXPLICITLY_DEFINED(web100_log_entry.snap.HCThruOctetsReceived) AND
web100_log_entry.snap.HCThruOctetsReceived >= 8192 AND IS_EXPLICITLY_DEFINED(project)
AND project = 0 AND IS_EXPLICITLY_DEFINED(web100_log_entry.is_last_entry) AND
web100_log_entry.snap.Duration >= 9000000 AND web100_log_entry.snap.Duration < 3600000000”

Download Measurements:

“SELECT

STRFTIME_UTC_USEC((INTEGER(web100_log_entry.log_time) * 1000000), '%F %H:%M') AS day_timestamp, web100_log_entry.connection_spec.local_ip,

web100_log_entry.connection_spec.remote_ip,connection_spec.client_hostname, connection_spec.client_geolocation.country_name,connection_spec.client_geolocation.city, connection_spec.client_geolocation.postal_code,connection_spec.client_geolocation.latitude, connection_spec.client_geolocation.longitude,

8 * web100_log_entry.snap.HCThruOctetsAcked/ (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd) AS downloadThroughput,

web100_log_entry.snap.SegsRetrans/web100_log_entry.snap.DataSegsOut AS packetRetransmitRate,

web100_log_entry.snap.SumRTT/web100_log_entry.snap.CountRTT AS avgRTT,

web100_log_entry.snap.MinRTT, web100_log_entry.snap.SndLimTimeCwnd/8 * (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd) AS congestionLimitedStateTimeShare, web100_log_entry.snap.SndLimTimeRwin/8 * (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd) AS receiverLimitedStateTimeShare,

web100_log_entry.snap.SndLimTimeSnd/8 * (web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd + web100_log_entry.snap.SndLimTimeSnd) AS senderLimitedStateTimeShare

FROM

[plx.google:m_lab.2014_01.all],[plx.google:m_lab.2014_02.all],[plx.google:m_lab.2014_03.all],[plx.google:m_lab.2014_04.all],[plx.google:m_lab.2014_05.all],[plx.google:m_lab.2014_06.all],[plx.google:m_lab.2014_07.all],[plx.google:m_lab.2014_08.all],[plx.google:m_lab.2014_09.all],[plx.google:m_lab.2014_10.all],[plx.google:m_lab.2014_11.all],[plx.google:m_lab.2014_12.all]

WHERE

connection_spec.client_geolocation.country_name = 'Canada' AND
IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeRwin) AND
IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeCwnd) AND
IS_EXPLICITLY_DEFINED(web100_log_entry.snap.SndLimTimeSnd) AND
IS_EXPLICITLY_DEFINED(web100_log_entry.is_last_entry) AND
IS_EXPLICITLY_DEFINED(web100_log_entry.snap.HCThruOctetsAcked) AND
IS_EXPLICITLY_DEFINED(project) AND project = 0 AND
IS_EXPLICITLY_DEFINED(connection_spec.data_direction) AND connection_spec.data_direction =
1 AND web100_log_entry.snap.HCThruOctetsAcked >= 8192 AND
(web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd +
web100_log_entry.snap.SndLimTimeSnd) >= 9000000 AND
(web100_log_entry.snap.SndLimTimeRwin + web100_log_entry.snap.SndLimTimeCwnd +
web100_log_entry.snap.SndLimTimeSnd) < 3600000000 AND
IS_EXPLICITLY_DEFINED(web100_log_entry.snap.CongSignals) AND
web100_log_entry.snap.CongSignals > 0”