Rural Matrix Process Application: Initial Data Analysis

Authors: Sarah-Patricia Breen & Danika Hammond
Version: March 25, 2026

Background

We applied the Rural Matrix process (as described in the Rural Matrix: Summary brief) to the *2026-01-31_Census Tables First Nation custom tabulations* data set. Analysis focused on population (Total reserve population and On reserve population, 2023) and Statistics Canada Index of Remoteness scores (Remoteness Index 2021) to investigate for clusters that could be used as the basis of a matrix. Investigations also explored geographic zone code, as well as income. The following note summarizes and discusses our findings.

Findings

1. Selecting the Unit

The first decision in applying the Rural Matrix process is to determine a unit type. For the BC Rural Matrix, the census subdivision was chosen as the most commonly used unit in government programs and as a well known and understood unit.

The custom tabulations data set includes 636 First Nations. From our understanding of the data set, these entries do not directly correspond with a consistent Statistics Canada geographic unit at they include on and off reserve population. Researchers used these entries as provided. Additional research using the census subdivision unit was not explored.

2. Selecting More than One Measure

The Rural Matrix process recommends selecting more than one measure for defining rurality in order to serve as proxies for understanding differences across rural as well as between rural and urban. The BC Rural Matrix selected distance (remoteness) and size (population).

For the first measure, the data set provided offered two measures of remoteness for consideration: Statistics Canada's Index of Remoteness and a Geographic Zone Code.

To better understand the overlap between these two remoteness measures, the Geographic Zone Codes (1 accessible and 4 being remote) were compared to the five classifications of the Index of Remoteness suggested by Statistics Canada (easily accessible, accessible, less accessible, remote and very remote).

The following matrix table shows the number of First Nations with each type of score on both remoteness measures.

Geo Zone Code						
1 – accessible	6	45	45	56	32	24
2	73	142	59	14	2	1
3	22	3	1	0	0	2
4 – remote	74	14	3	3	0	15
	Very Remote	**Remote**	**Less Accessible**	**Accessible**	**Easily Accessible**	**NA**
	Index of Remoteness					

If the comparison demonstrated overlap it would indicate that using either measure would yield similar results. However, this was not the case. The First Nations that are determined to be the most accessible using the Geographic Zone Code were had a relatively even distribution between remote and easily accessible using the Index of Remoteness, including 6 First Nations being considered very remote using the Index. On the opposite end, 3 First Nations are considered remote using the Geographic Zone Code are Easily Accessible using the Index of Remoteness.

As an example, Penelakut Tribe is Zone 4, the most remote using the Geographic Zone Code but is accessible using the Index of Remoteness. Penelakut is in the Gulf Islands and has ferry access 10 times a day, an example of a location that isn't

accurately classified by either system. While restricted to ferry access, they have relatively easy access to major urban centres, making the most remote classification inaccurate, but classifying them as accessible is also inaccurate.

Both variables have limitations. The Geographic Zone Code cannot be translated into an accessible to remote scale as it purely measures the type of access (road access) and the distance to a service centre; and the service centre may be relatively small. However, the Index of Remoteness still fails to accurately account for ferry and plane access dependent communities. In BC we rectified this problem by classifying communities one measure higher in remoteness if they lacked year-round road access. A similar approach could be used here using the Geographic Zone Code 4.

Key Finding: continue exploration using the Index of Remoteness

Having chosen the Index of Remoteness as one measure to build a matrix, determining the appropriate thresholds was next on our process.

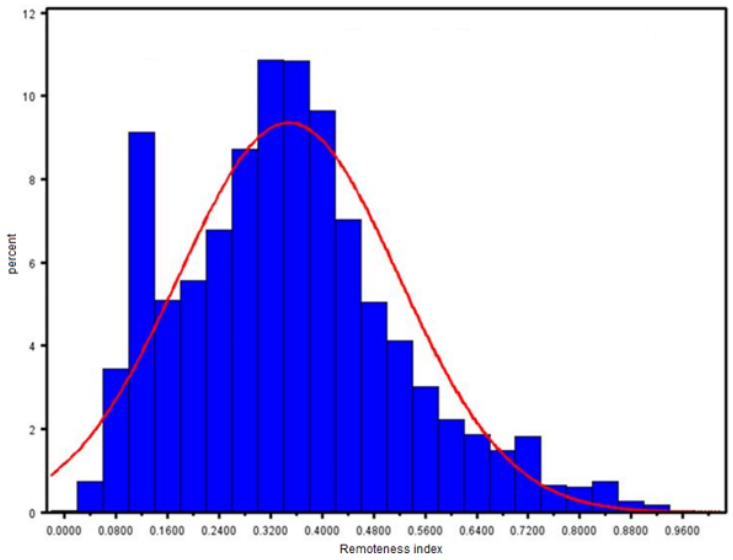
Statistics Canada classifies the index of remoteness into five groups based on a manual cluster approach, identifying natural breaks in the data to create mutually exclusive categories. This is opposed to an equal interval approach (<0.2, 0.2-0.39, 0.4-0.59, 0.6-0.79) which would result in very few communities being in one category and the majority being in another, or a quantile classification approach which requires an equal number of communities in each category.

When looking at Canadian census subdivisions (CSDs), there is a range between each category with 11.92% of CSDs being very remote and 27.51% being less accessible. The range is larger for the First Nations communities and weighted differently, with only 5.72% of First Nations being easily accessible and 34.34% being remote.

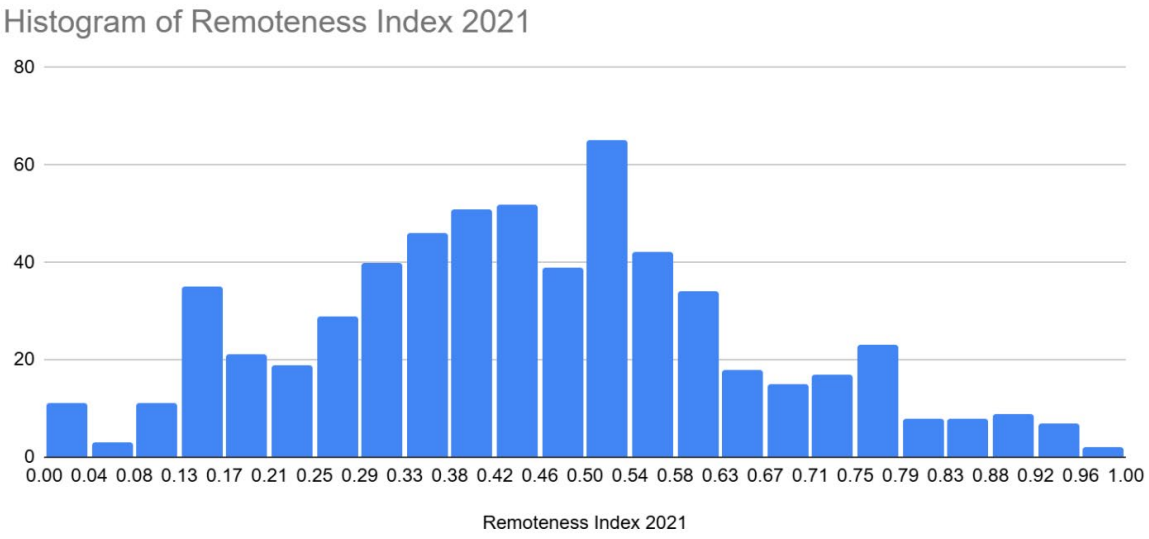
	IFSD	IFSD	All Canadian CSDS
Easily Accessible	34	5.72%	15.10%
Accessible	73	12.29%	21.48%
Less Accessible	108	18.18%	27.51%
Remote	204	34.34%	23.98%
Very Remote	175	29.46%	11.92%
Total	594		

The two graphs below show the distribution of index values for all census subdivisions in Canada and all First Nations. The variation between the distributions can be seen while comparing the graphs.

Distribution of remoteness index values of census subdivisions in Canada



Source: The data for this chart are from the data table from the remoteness index of Canadian census subdivisions (Alasia et al. 2017).



Given the variation in distribution, there are two potential options for creating classification thresholds using the Index of Remoteness.

Option one is to use Statistics Canada's existing manual classification thresholds as described above. This option is simple and captures the relativity of remoteness of all communities. However, it is using a classification system based on all Canadian CSDs, which represents a wider range of communities and different geographic unit than that reflected in the data set.

Option two is to create a new manual classification system using only the First Nations provided in the data set. This option is more complicated, but creates thresholds based on the relativity of remoteness between the First Nations in the data set.

Remoteness is a relative, not absolute concept. If the desire is to compare the First Nations communities only against themselves, it is recommended to go with option two, if there is a desire to incorporate the relativity of all Canadian communities, it is recommended to go with option one.

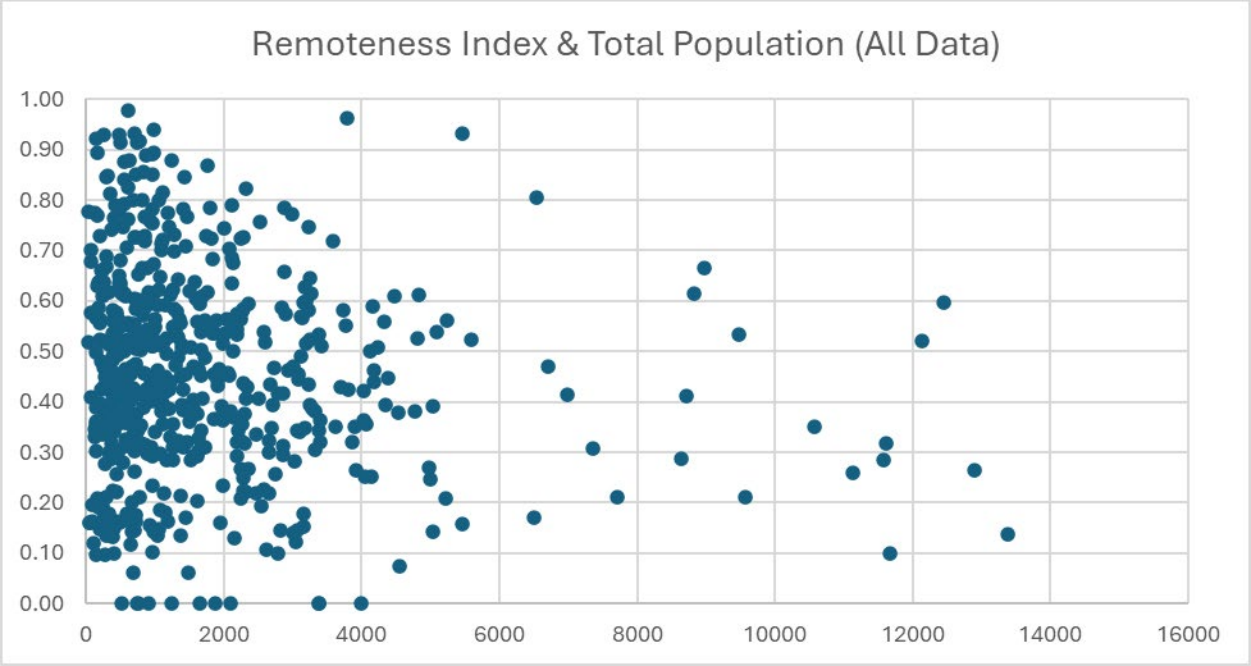
Key Finding: create new classifications for the Index of Remoteness based on the data set.

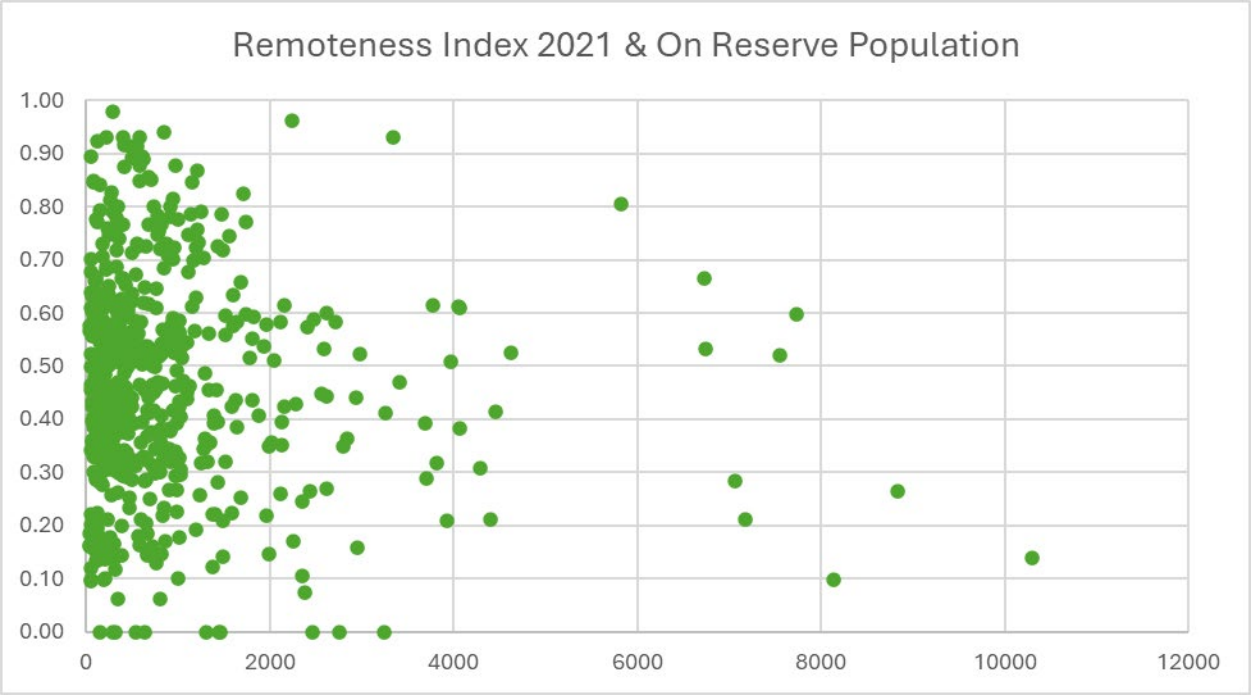
When selecting the second measure, population is the most commonly and typically most easily used measure for defining rurality. The data set included multiple population measures (i.e., On reserve, off reserve and total), and all First Nations have relatively low populations in comparison to municipalities. We compared Total Population and On Reserve Population and finding Total population had larger variation in data (min. 41 – max. 25,401) we determined to start our exploration with Total Population.

Key Finding: continue exploration using total population.

3. Cluster Analysis

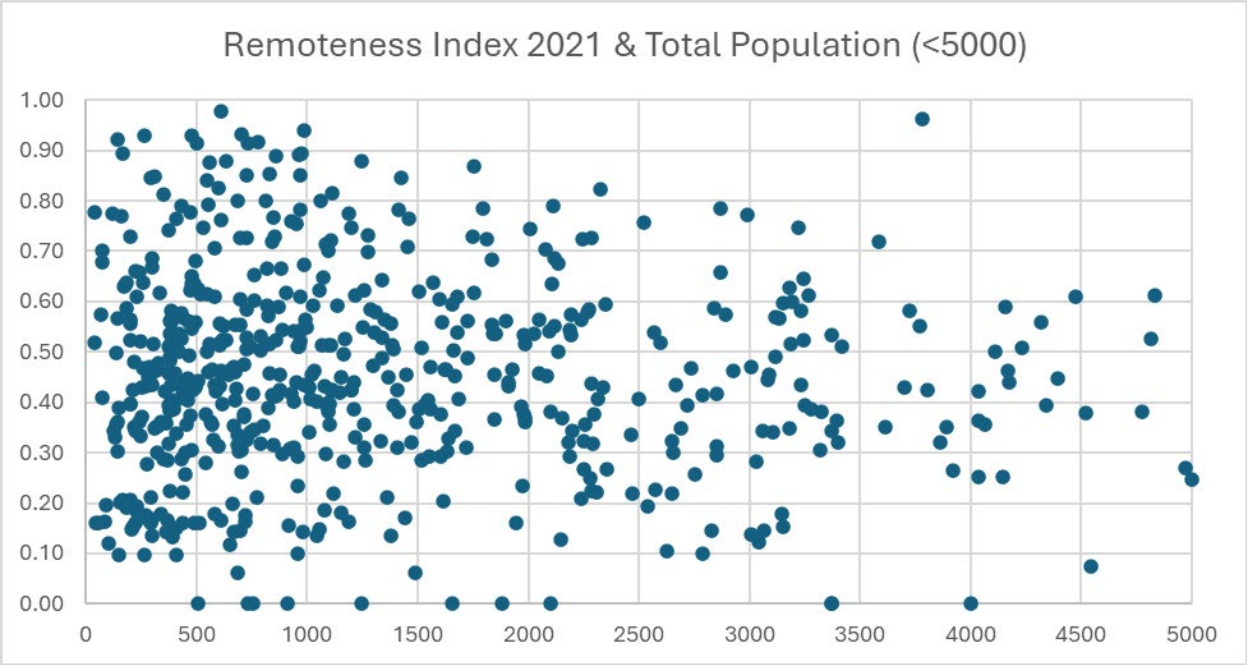
There were no visually obvious clusters corresponding to total population and remoteness. Using Total population, the largest population is 25,401, but this entry does not have a corresponding remoteness score. A comparison with On reserve population shows a similar distribution.

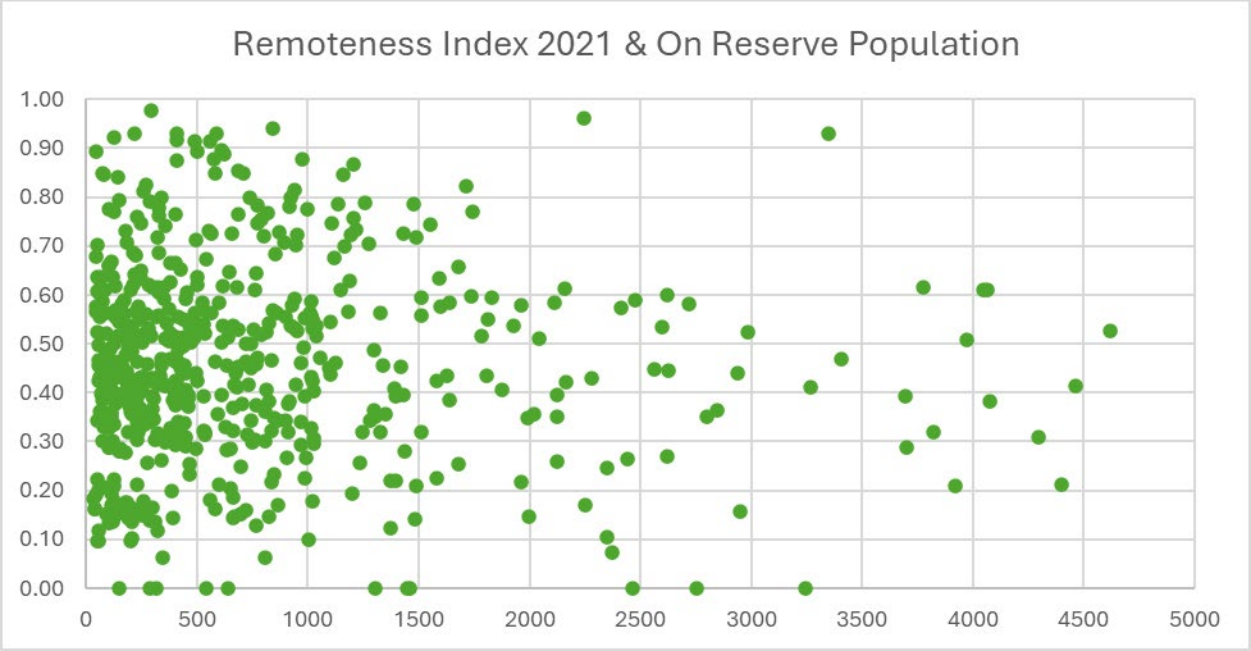




Using Total population, the majority (94%) of entries have a population of less than 5,000, and 49% have a population of <1,000. The benchmarks of 5,000 and 1,000

are useful for consideration for rurality. 5,000 is often used as threshold for additional requirements and responsibilities, as well as access to funding. 1,000 is the Statistics Canada threshold for being defined as rural. Again, a comparison with On reserve population shows a similar distribution.

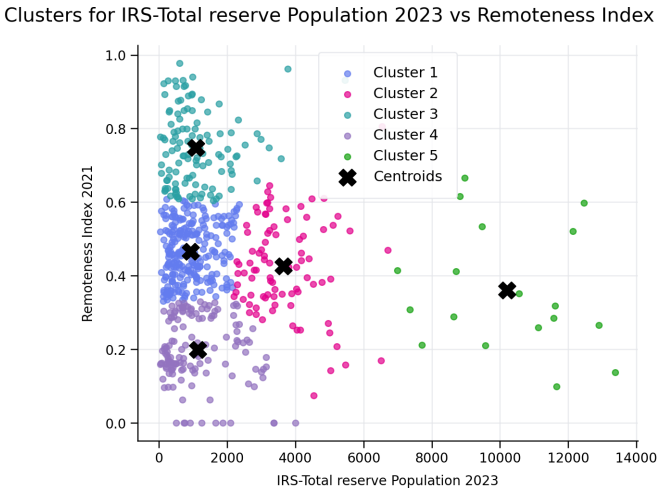




Key Finding: The data set is relatively similar in terms of size, but with is a wide range of remoteness scores.

Further cluster investigations were conducted using AI and SPSS.

Example #1: K-means cluster analysis using AI. Identified 5 clusters, with 1, 3, and 4 being similar in size but different clusters by remoteness score, and 2 and 5 being differentiated by both size and remoteness.



Example #2: Hierarchical Coding via SPSS identified two clusters. Cluster #1 included 24 First Nations (centre of cluster population of 9,163, IR of 0.39). Cluster#2 included 577 First Nations (centre of cluster population of 1379, IR of 0.45). This suggests significant difference between the small number of large ones and the majority small ones.

Given the lack of variation in population, we explored the potential of an alternative second variable. As the aim of the Income Assistance Program is to address income inequality, we explored median household income. Unfortunately, given that many of the entries were missing this data, there was insufficient data to identify obvious connections or clusters.

Key Findings:

- ***The identified clusters suggest different possibilities but non are particularly strong.***
- ***Identified clusters do not align with the remoteness index thresholds set by Statistics Canada.***

4. Trial Matrix Creation

We created a trial matrix using population and remoteness, using the AI generated K-means results as a starting point. Given the small size of cluster 5 (10) and 2 (92), they were combined. The remoteness score classifications identified do not align with those suggested by Statistics Canada.

SMALL	**Pop1**	38	58	39	16
	Pop2	23	49	42	30
	Pop3	40	84	59	26
LARGE	**Pop4**	27	37	28	5
		RI1 LEAST REMOTE	**RI2**	**RI3**	**RI4 MOST REMOTE**

Key Finding: The matrix identified communities that do not follow expectations based on previous applications of the matrix. This

substantiates that there is greater variation in the data set based on remoteness than size.

Discussion

Based on the above, the application of the Rural Matrix approach using only population and remoteness may not be directly applicable to First Nations in isolation when they're removed from the broader context of other communities.

- **Classification based on Population may not be suitable:** there is too little variation in population for this variable to be effective, outside dividing between small and very small. In the cases of non First Nations communities, the thresholds of 5,000 and 1,000 have practical applications, which substantiate their use. However, it is unclear if population acts as a proxy for financial and administrative access and capacity in the same way for First Nations as it does for non-First Nations Communities.
- **Classification based on Remoteness has potential applicability:** different classifications from what Statistics Canada recommends should be used to account for the variation in a large number of similarly sized communities. Geographic zone codes could be used to identify where remoteness score may be mis-aligned with reality.
- **The creation of meaningful groups of First Nations for program improvements should not be based exclusively on factors of rurality:** measures of population and remoteness are critical for rurality, as they measure against a broad data set of all communities. When looking at First Nations in isolation from other communities, the variation within rural is reduced. While most First Nations are rural, our findings suggest that when looking at First Nations in isolation, rurality may not be the most effective factor to apply in order to create meaningful groups.

Revisiting the options identified in the Rural Matrix Summary, the findings indicate the following:

1) **Classification of First Nations communities using remoteness only:** it is possible to create a simple classification based on remoteness and geographic zone. Remoteness threshold could be tailored at the provincial or regional level to account for differing contexts. However, this approach would lack the nuance of having an additional variable.
2) **Treat classification of First Nations the same as municipalities and apply both population and remoteness:** beyond separating the largest from smallest communities, creating thresholds based on population is not meaningful approach for this application.
3) **Exclude and explore a more suitable approach:**
    - **Option #3A:** Investigate other available Income Assistance program variables as the second variable.
    - **Option #3B:** Identify and explore additional factors that can be used to create groups applicable to the Income Assistance program. This would likely focus on qualitative factors, including the culture and history of colonization unique to each First Nation, as well as unique provincial considerations.