

Evaluating the Fitness for Use of Alternative Data Sources and Integrating Them into the Retail and Service Industries Programs

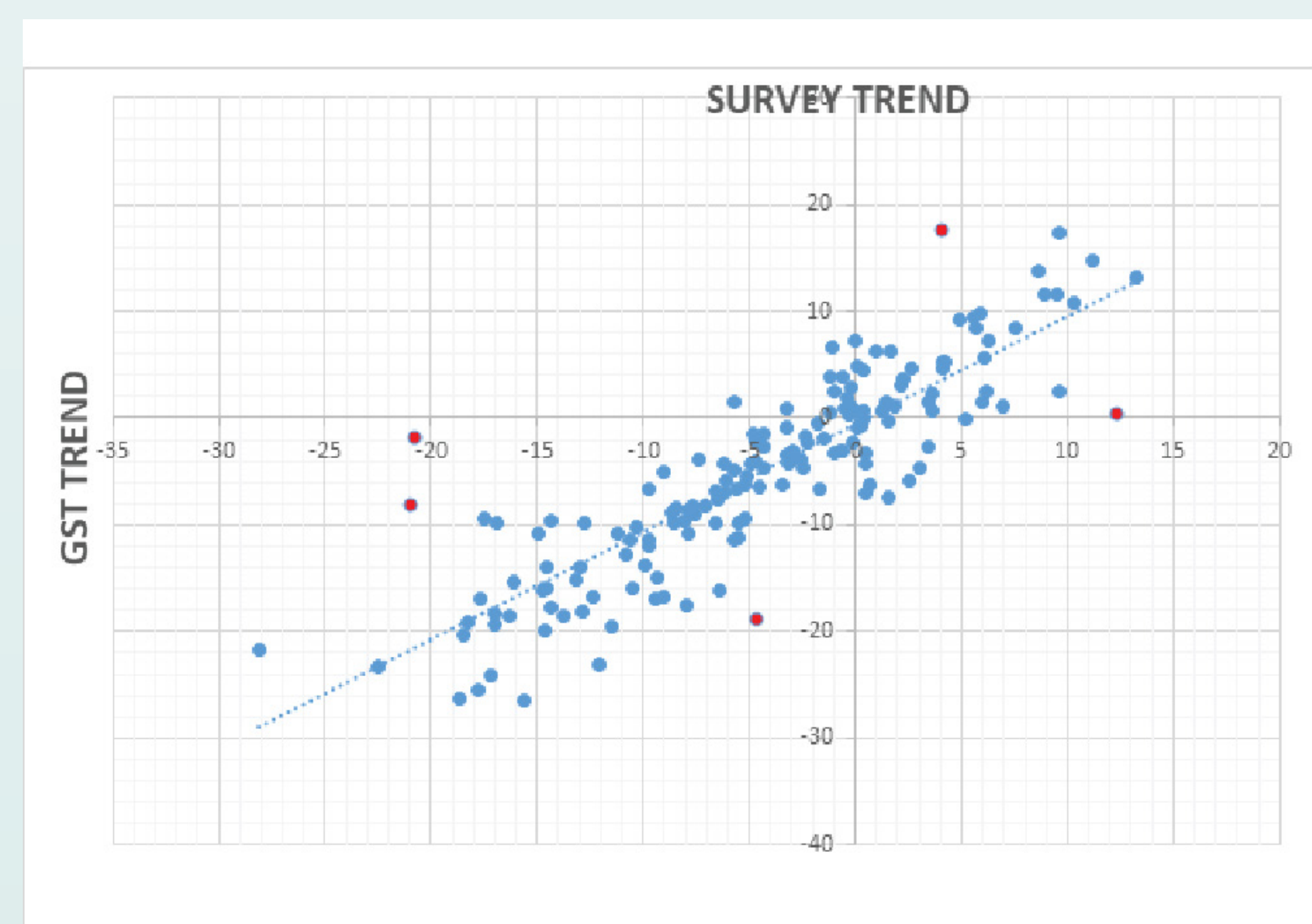
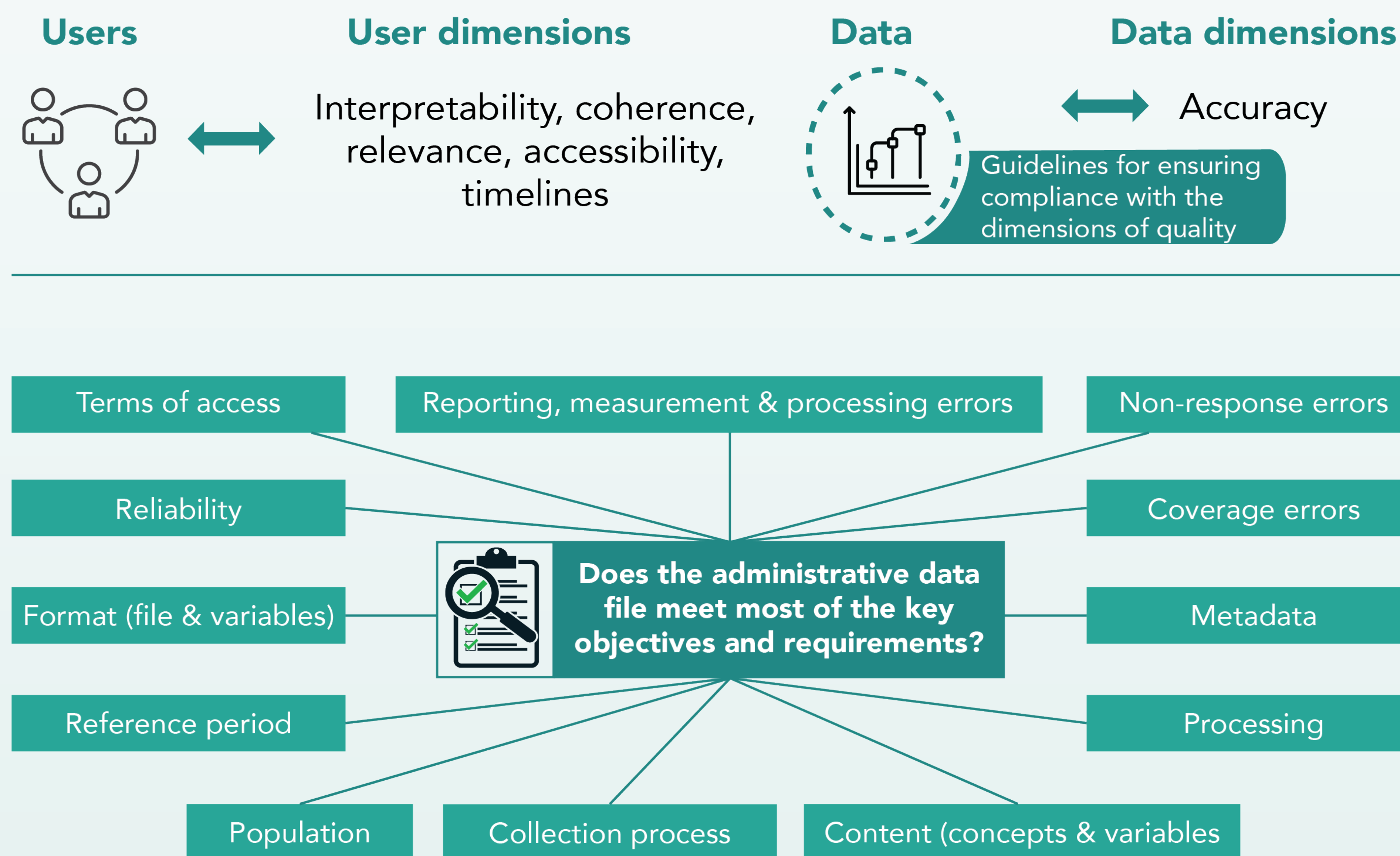
Mathieu Thomassin and Kyle Virgin
– Statistics Canada

Abstract

- The Administrative Data Inventory lists more than **1,500** administrative data sources StatCan has in house or is in the process of negotiating
- Statistics Canada uses corporately recognized quality assurance processes to assess the fitness for use of administrative data in its statistical programs
- For each dimension, a list of elements is established and managers determine the importance depending on intended use and assess the level of fitness and appropriateness

Introduction

- For several decades, the Monthly Survey of Food Services and Drinking Places has collected between 2,600 and 3,500 questionnaires and produced estimates using traditional sampling and estimation methods
- This initiative aimed to replace the majority of survey data and the sample design process through a more simple and direct tabulation of goods and services tax (GST) data following an assessment of the quality of this alternative data source
- But what is quality in a multi-data source environment? The sources being combined have different types of errors: issues with input data and errors due to processing



Method

Replacing survey data with goods and services tax data in the Monthly Survey of Food Services and Drinking Places



- Terms of access and reliability were sound with this data source based on historical experience
- Collaborated proactively with data providers to improve the relevance, timeliness and quality of administrative data holdings for statistical use
- Adapted statistical processes to minimize errors:
 - imputation of GST
 - calendarization
 - sub-national allocation of the GST data
 - estimation (adjustments to take care of the coverage issues)
 - negotiate chain agreements with major respondents
- Validated new method against previous method (e.g. level and trend), while bearing in mind the sources of sampling and non-sampling error in each

Results

- Reduces collection costs (\$225,000)
- Reduces response burden
- Creates better precision
- Produces estimates closer to its annual counterpart
- Produces more granular data to publish more domains
- Reduces sampling and processing costs

Conclusion

- Our methodologists were asked to propose one "global" quality indicator for estimates at the cell level
- The use of websites or other administrative data should be used to better identify seasonal closures and improve the quality of estimates, particularly in Prince Edward Island and, to a lesser extent, in Nova Scotia
- More efforts are required to keep the Business Register updated for classification and business status
- We must pay special attention to shocks that could have an impact on GST data (COVID-19) and react!