



Statistique  
Canada

Statistics  
Canada

# Marrying demand for statistical information with disclosure control: The Canadian experience in developing an automated dissemination tool in an open-data world

---

March 9, 2018

---

Zixin Nie (Statistics Canada)  
Claude Girard (Statistics Canada)



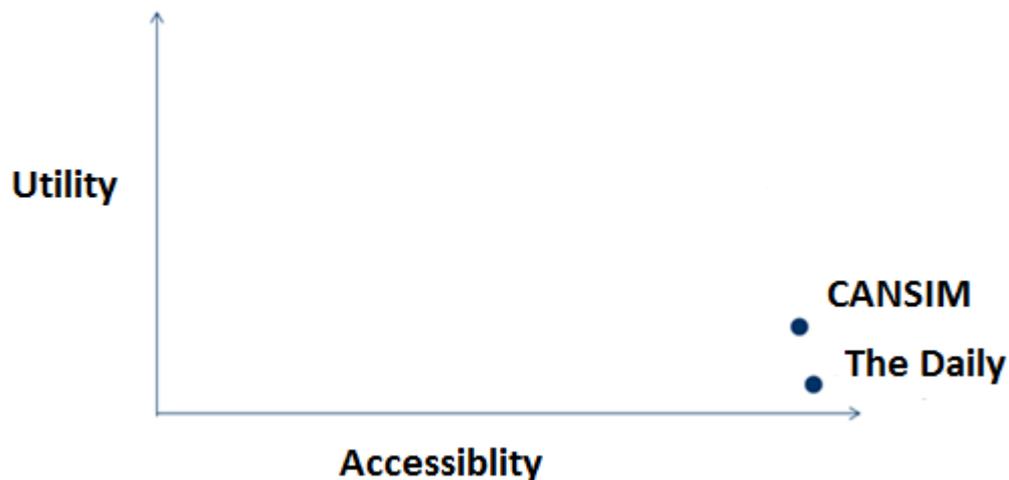
# Outline

- Evolution of Data Access
- The Generalized Tabulation system (GTAB)
- Results and the current state of GTAB



# Evolution of Data Access

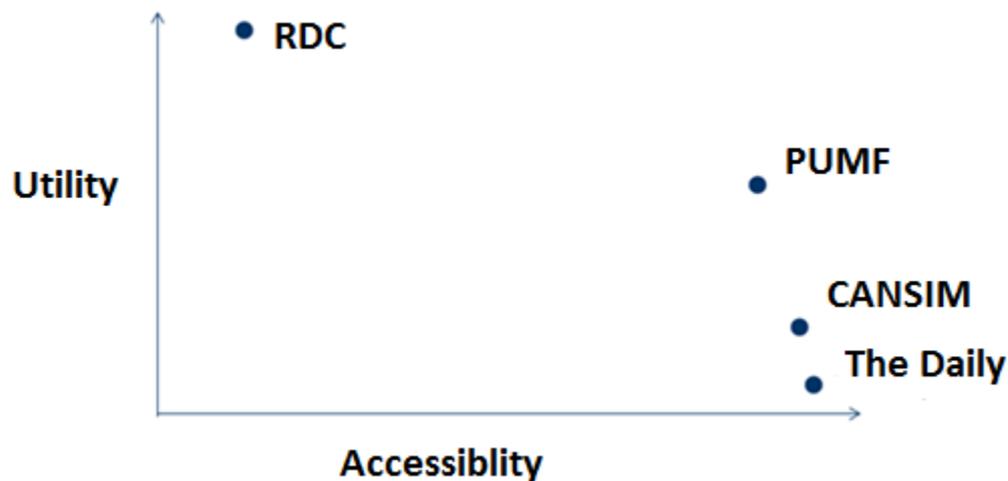
- Statistics Canada has a mandate to gather data and report upon the findings to all Canadians
- Main vehicles of dissemination are the Statistics Canada Daily Report (*The Daily*) and CANSIM (system for viewing official tables)
  - Published products of aggregate statistics
  - Relatively high-level overviews





# Evolution of Data Access

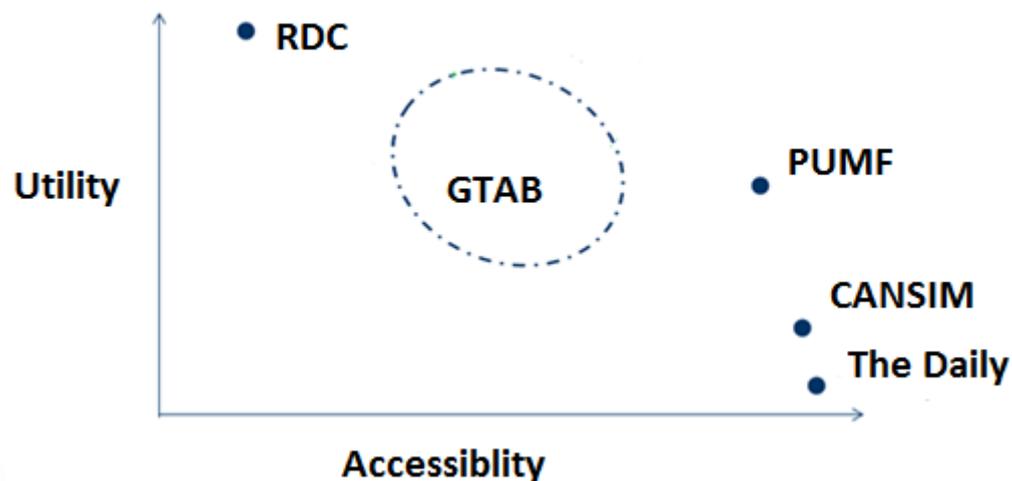
- Main modes of directly accessing Statistics Canada microdata by external users are Public Use Microdata Files (PUMFs) and Research Data Centres (RDCs)
- PUMFs are modified microdata files that minimise the risk of disclosure of confidential information
- RDCs give users direct access to unmodified STATCAN microdata, but require special permission to access





# Creation of the Generalized Tabulation System

- Generalized Tabulation System (GTAB), was borne from new needs arising from greater demand for data access





# What is the GTAB system?

- Generalized System
- Next generation tabulation and dissemination tool
- Corporate tabulation tool
  - Social, health, and labour statistics
- Direct pipeline from microdata to publishable tables



# Broader vision of GTAB

- Standardization of practices within Statistics Canada
  - GTAB not a system designed to replicate all functionality of previous systems
  - Move towards more standard practices across surveys in microdata structure, estimation, dissemination, confidentiality
  - Main benefit: making published products easier to comprehend through similar structure
  - Another benefit: skills obtained through working on one survey can be easily transferred to working in other areas



## Broader vision of GTAB

- Create an easy to use system for users without programming experience
  - Other systems used at STATCAN require experience with coding in SAS
  - GTAB needs to be accessible to users who do not have extensive coding experience
  - GUI – simple enough for most users to learn and use quickly



# The GTAB Framework

Single Microdata File



Calculation Engine

- Calculates statistics
- Calculates precision measures



Assign Quality Indicators based on Precision Measures



Apply Confidentiality



Create final output for dissemination

- Generalized process to dissemination
- Streamlined pipeline from microdata to publishable tables
  - Takes final microdata files as input (with survey weights and replicate weights)
  - Create table specifications
    - i.e. Select domain variables to cross, statistics to be calculated
  - Outputs can directly be disseminated



# The GTAB Framework

Single Microdata File



Calculation Engine

- Calculates statistics
- Calculates precision measures



Assign Quality Indicators based on Precision Measures



Apply Confidentiality



Create final output for dissemination

- GTAB will automatically
  - Calculate precision measures using replicate weights (Rao-Wu-Yue bootstrap weights)
  - Assign standardized quality indicators based on coefficient of variation
  - Apply confidentiality rules
- Data flow is linear, we do not pass information back to previous steps
- Cannot combine multiple files for calculation within GTAB



GTAB - Tabulation Tool (GRID TEST Environment)

Application: Report Page

Save all Close all unmodified Close page

Page path: Main GMTS - GTAB Methodology Testing Survey 2008 - 2008 Jobs: zxiismalcats **Table: M1stats**

Main **Table: M1stats**

Table name: M1stats Rename ... Save Results Discard changes

Masterfile and weight

SAS input

Variables of interest

Selected analysis variables:		Selected domain variables:	Domain hierarchy:	
Name	Constant dollar			
INR_Q032		CAT		

Const. \$ year: 2015\$

Statistic specifications

GIN(INR_Q032)	+ Add statistics ...	<input type="checkbox"/> Add WEIGHTED-FREQUENCY	+ Add quantile ...
MEAN(INR_Q032)	+ Add ratio ...	<input type="checkbox"/> Add PERCENT-DISTRIBUTION	+ Add moving averages ...
P10(INR_Q032)	+ Add proportions ...		+ Add higher-order statistics ...
P20(INR_Q032)			+ Add higher-order statistics over domain hierarchies ...
P50-MEDIAN(INR_Q032)			+ Add higher-order statistics over domain variables within a domain hierarchy ...
P75(INR_Q032)			
P90(INR_Q032)			
TOTAL(INR_Q032)			

Outputs and formats

Messages:

- 2017/06/29 3:39:10 PM To continue you may open the Jobs page.
- 2017/06/29 3:20:22 PM Finished SAS connection test: No connection.
- 2017/06/29 3:20:22 PM SAS connection attempt timed out.

Copy Clear



# GTAB Functionality Development Process

- Data providers (clients) approach GTAB team for dissemination
- Demand VS supply assessment: Clients' needs VS GTAB's functionalities
- Standardization: Significant business case must be made before turning yet-unfulfilled needs into new system specifications



# GTAB Functionality

- Statistics that GTAB can currently calculate
  - Level 1 statistics: Mean, percentiles, median, total, weighted frequency
  - Level 2 statistics: Gini coefficient
  - Level 3 statistics: Proportions and ratios
  - Level 4 statistics: Moving averages over time
  - Level 5 statistics: Level change, percentage change, significance tests (Global, base value, sequential, sequential over time)
  - Quantiles, both as domain variables and as bound statistics
  - All calculated statistics use survey weights
- Precision measures
  - Variance, standard error, coefficients of variation, confidence interval bounds



# GTAB Functionality

- Confidentiality rules
  - Each statistic currently available in GTAB has its own set of confidentiality rules
  - Rules are applied equally, regardless of subject matter
  - Tested through simulation studies on fake and real data, vetted by experts, and approved through management
  - ACRound, suppression based on minimum counts, rounding of final outputs
  - Parameter-driven
  - Rules are automatically applied to outputs when requested



## Current state of GTAB

- GTAB can calculate ~90% of statistics found in published tables for social, health, and labour statistics
- Numerous surveys are transitioning their dissemination to GTAB, such as Canadian Community Health Survey, Education Surveys, Tourism and Travel surveys, and Labour Force Survey
- Census moving dissemination to GTAB
- New functionality in constant development to meet new business requirements



# Advantages to using GTAB

- Dissemination becoming more standardized
  - Tables from many surveys now use standard confidentiality rules, standard quality indicators, and standard methods for calculating statistics and precision measures
  - Skills obtained when disseminating for one survey are now useful for many different surveys
  - Users of published STATCAN data on CANSIM now have information presented in a more uniform fashion, increases usability of data
- Promotion of improved methods for reporting quality of estimates, such as publication of confidence intervals
- Promotes better practices internally when creating pre-dissemination files



## Advantages to using GTAB

- Provides precision estimates for a variety of statistics using replicate methods (bootstrapping)
- Increased timeliness for custom tabulations
- Engine used in GTAB system is also being used to power other systems for automated dissemination



## Conclusion

- Demands from users of Statistics Canada data drove need to develop new system for dissemination and confidentiality
- Developed GTAB as an easy-to-use tabulation system
  - Automated calculation of statistics, application of confidentiality rules, and creation of quality indicators
  - Rigorous approved methods for standardization
- Adoption of GTAB has resulted in more standardization in disseminated products and practices within STATCAN



# Future Developments

- Modernization initiative
- Open-data initiative
- Cloud-based storage



Thank you for attending!  
Merci de votre attention!

Questions?

Zixin Nie: [zixin.nie@canada.ca](mailto:zixin.nie@canada.ca)

Claude Girard: [claudio.girard@canada.ca](mailto:claudio.girard@canada.ca)