



Pilot Project to Access and Leverage Government-Collected Business Data

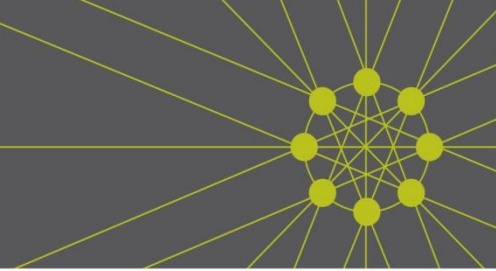
Building an Open Cloud-based Synthetic-Data
Sandbox

PSSDC - Charlottetown, PEI
October 5, 2017



Innovation, Science and
Economic Development Canada

The Opportunity



There is momentum around data:

- Minister Bains announced a Data Vision for Canada in August 2017
- Modernization of Statistics Canada 2017
- Federal/Provincial/Territorial (FPT) Open Government Initiatives
- Goal to adopt the business number (BN) as an identifier across the federal government by 2020

We know there is an untapped potential of data:

- *Data is available:* FPT governments collect a lot of business data in the course of operations
- *Data can be integrated:* BN adoption presents an opportunity to link this data
- *Data can be analyzed:* By building capacity in analytics, intelligence can be derived from linked data.
- *Value can be created:* Linked data can help us better understand business and client needs, spur innovation and improve efficiency.

BETTER USE OF DATA CAN IMPROVE SERVICE AND PROGRAM DELIVERY

The Project



A cloud-based, scalable sandbox for policymakers and researchers to use to explore business data and experiment outside the firewall Statistics Canada must maintain for privacy and confidentiality purposes.

Input

- Variables and business data collected by F-P/T programs that use the BN as a unique identifier.
- This data is (would be) reported to Statistics Canada and placed behind a firewall.
- BN adoption enables improved data integration and validation (i.e. improved quality).

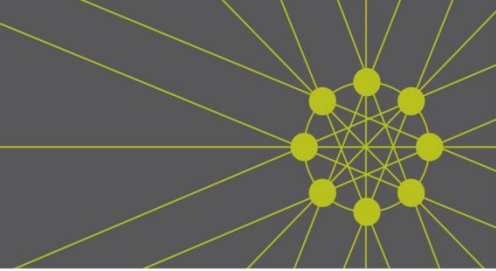
Process

- Leveraging Statistics Canada's techniques, data collected would be made "synthetic" to maintain privacy and confidentiality and then placed in the "sandbox".
- Users (e.g. policymakers, program officers, researchers, civil society) would know what data variables are available for analysis.
- Users could experiment using data variables and the synthetic data to create analytic models or questions.
- Upon request, Statistics Canada could apply these models or ask the question to the real data behind the firewall.

Output

- Results based on the real data would be released to the research consistent with standard practice.

Expected Outcomes



There is value to having accessible data organized by BN to enable analytics and improve policy, service and program outcomes.

Enhanced Access to Data

- Integration of data from a variety of sources and levels of Government.
- Open access to government-collected business data variables and synthetic data.
- Government data is easier to find and use, which results in greater transparency.
- Could encourage adoption of the BN.

More Analytics and Data-Driven Decision-Making

- Analytics is leveraged to identify relationships between data variables and derive intelligence.
- Better problem identification and use of analytics to solve these problems.
- Capacity building: increased number of users that are digitally literate and have capacity for data analytics and data visualization.
- This tool is used to foster research and analysis and improve policymaking and program delivery.

Maximized Program Outcomes

- Better insight of clients profiles and needs.
- Greater understanding of policy and program impacts.
- More efficient, innovative and impactful services that could exceed delivery objectives.

Existing Use of the BN



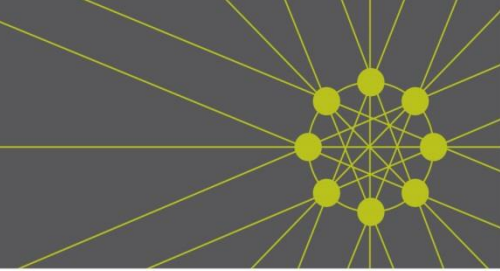
Federal Level

- The Statistics Canada Business Register contains data variables associated with the BN (e.g. Revenues, GST sales, number of employees).
- In 2016, Canada Revenue Agency launched a real-time BN Validation web service for federal early adopters.
- The Job Bank (Employment and Social Development Canada) is currently the only Government service that has fully adopted the BN as a unique identifier.
- 21 Government services are currently working towards leveraging the BN, including:
 - Dairy Farm Investment Program (Agriculture and Agri-food Canada)
 - International Mobility Program (Immigration, Refugees and Citizenship Canada)
 - Buy and Sell (Public Services and Procurement Canada)
 - Issuing Radio/Spectrum Licenses (Innovation, Science and Economic Development Canada)

Provincial Level

- Six provinces have adopted the BN as a single identifier so far (BC, MB, NB, NS, ON and SK).
- PEI and AB have been making strides towards its adoption.

Job Bank Use Case



- Given the innovative nature of this pilot project, we sought out a partner in developing a possible use case to think through how the tool might be used.
- Job Bank is a long-time user of the BN and agreed to help out so we could learn more about its experience and the possibility for it to be the first use case for this pilot project.
- The Job Bank shared its data dictionary and its current analytic strategy.
- Early analysis suggests that wider use of the BN, with data housed together at Statistics Canada, would allow:
 - Job Bank to learn more about the businesses that are using Job Bank
 - Analysis of linked data to determine if firms that use the Job Bank report employment growth (e.g. through record of employment data)
 - For the design of an analytical model that could help assess whether public grants and contributions have an impact on the number of jobs posted by firms through the Job Bank.

Challenges



Legal, Privacy and Confidentiality Requirements

- Ensure compatibility with *Statistics Canada Act* and other legal requirements associated with the data Statistics Canada possesses.
- Ease of use: utilize appropriate technology and techniques to provide users with tools and apps to assist in experimentation.

Bringing the Data Together

- Data custodians remain hesitant to share the business data they collect.
- A limited number of programs have adopted the BN.
- Common data definitions and data standards are needed to improve data linkages.
- Need to work across all levels of governments to amalgamate business data.

Scaling the Project

- Ensure built-in scalability of the underlying cloud-based platform for future increased usage (e.g. more data, more users).
- Secure access to funding to maintain and scale the project.
- Capture lessons learned and constantly improve the sandbox.

Next Steps



Accounting for existing challenges and constraints, the pilot approach will bring together people, technology and data to experiment and ultimately expand on best practices and lessons learned.

Next Steps Include:

- Further develop and implement the Job Bank use case.
- Engage with technology experts.
- Create of a User Advisory Group (UAG), including early BN adopters, policymakers and researchers, to advise the project team on the technology, the data and user interface.
- Secure one (or many) provincial partner to contribute data and test this tool.
- Organize a brainstorming session at the ISED Innovation Lab with ISED, Statistics Canada, technology experts and the UAG to refine the idea and its implementation.
- Identify and secure financial partners for the project.
- Identify first wave of data variables to include and construct the synthetic data.
 - Launch the cloud platform, adding analytic tools as available.
 - Launch the sandbox.
 - Conduct beta test.
 - Learn, expand and improve the tool.