



Statistics
Canada

Statistique
Canada

Data Centres- current and future state

www.statcan.gc.ca



Telling Canada's
story in numbers

Sylvie Michaud

November 2017

Canada

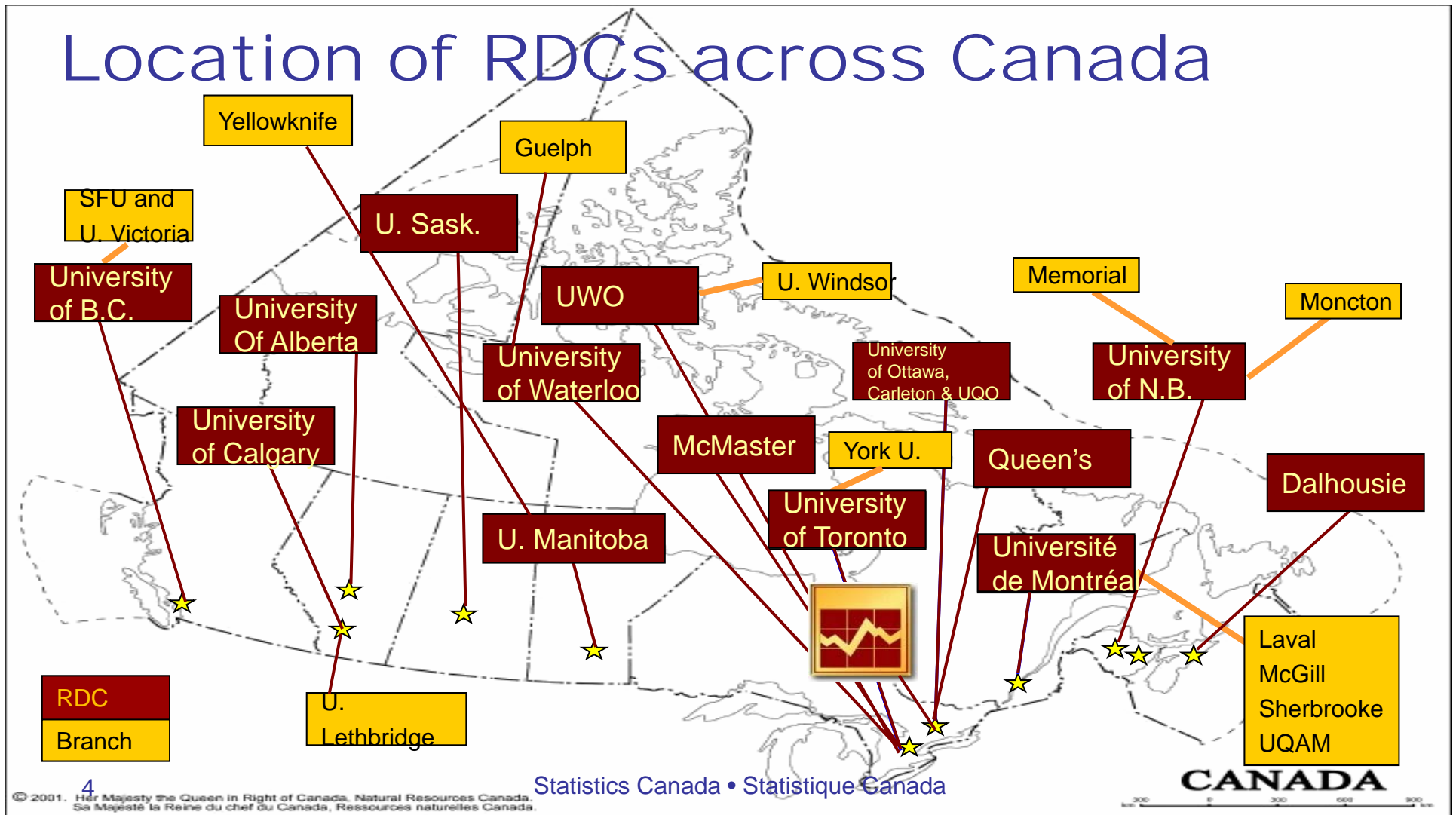
Current state

- In-take: « administrative » data is transferred to STC
- Apply steps of Generalised Statistical Business Process Model (GSBPM) in STC
 - Developed a model for record linkage
- Access to data is provided: continuum of methods
 - from infographics, aggregate statistics to anonymised micro-data files
 - For social statistics (Research Data Centres) to business statistics (Centre for Development of Economic Research)

What is a Research Data Centre?

- Secure Statistics Canada environment on a university campus
- Operates under the provisions of the Statistics Act
- Houses Statistics Canada social microdata files
- Managed by Statistics Canada employees at all times
- Access limited to researchers with approved projects and “**sworn-in**” under Statistics Act as “deemed employees”
- All researchers have direct access to the data
- Data are de-identified – all personal identifiers are removed

Location of RDCs across Canada

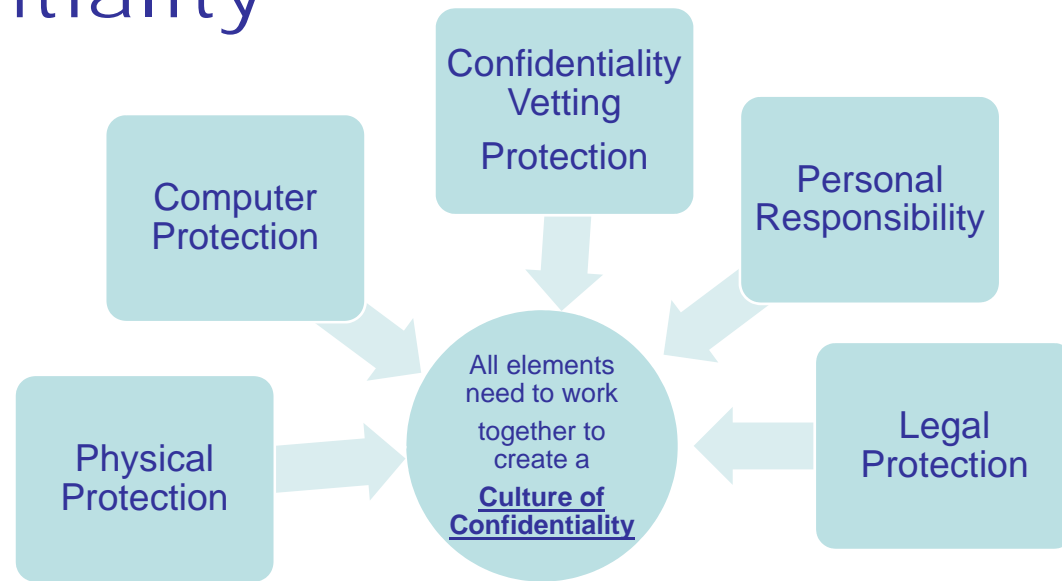


What is the Centre for Development of Economic Research

- Secure environment on Statistics Canada's premise
- Houses Statistics Canada business microdata files
- Researchers are deemed employees
 - Data are de-identified – all personal identifiers are removed
 - Researchers do not have direct access to the data;
 - they have access to synthetic data to test their program
 - An interface allows STC employees to run the same program on real data and provide outputs

August 2017

Measures for protecting data confidentiality



Inspections and audits are conducted regularly to ensure adherence to the security procedures

The challenge

- Storage and compute requirements are increasing
 - Digital content is expected to double every two years; we are starting to see the impact
- Solutions to many problems will likely come from the integration of data sources
 - possibility of duplication of large datasets
- There are pressure to provide more access to data

Path forward

- Better management of data from the in-take
 - IM/IT discussion
- Examining the use of the cloud
 - Based on levels of sensitivity of the data
- Research on protection of confidentiality
- Virtual data access instead of brick and mortar