



Statistics
Canada

Statistique
Canada

Canada



Statistics Canada
www.statcan.gc.ca

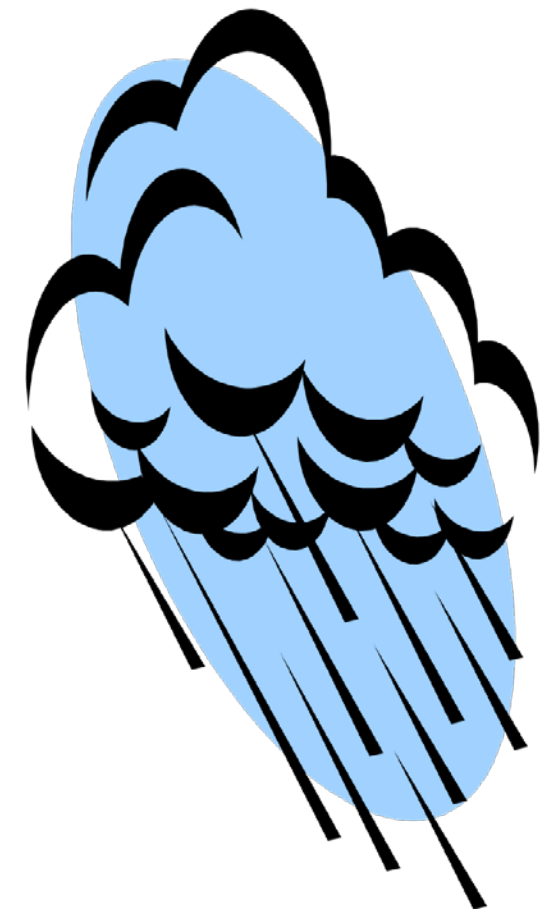
Opportunities: Linking Administrative and Survey Data for Statistical Purposes

Panel on New Data: Sources, Governance and Infrastructure, Analysis
Conference on Population Change and Life Course: Taking Stock and Looking to the Future

March 20, 2015

Current environment

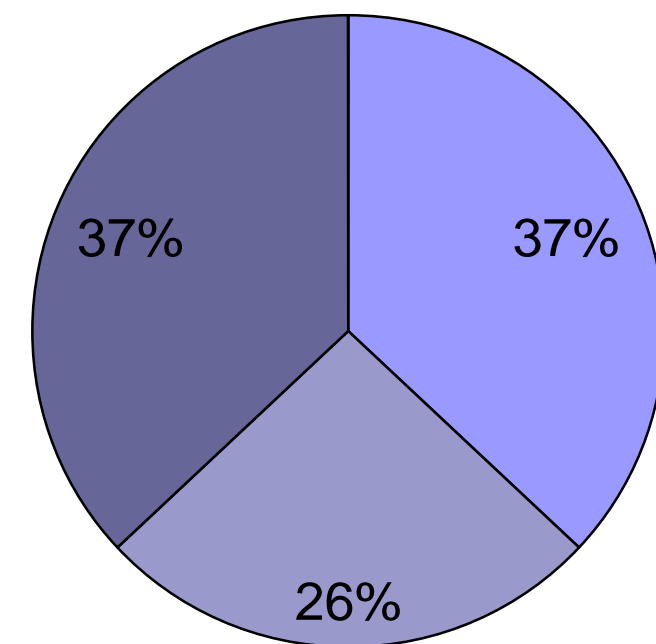
- Declining budgets
- Falling survey response rates
- Respondent burden
- Privacy concerns
- Need to look at alternatives



Use of administrative data has grown over time...

- Statistics Canada has been using administrative data in its programs since 1918
- Currently over 500 administrative files used to produce statistics
- 40% of Statistics Canada’s statistical programs are based in whole or in part on them
- Current uses include:
 - Direct replacement in business and household surveys to reduce respondent burden
 - Maintenance of Statistics Canada’s business and address registers
 - Benchmark economic and social statistics

Sources of Administrative Data



■ Federal
 ■ Prov/Terr
 ■ Other

Note: “Other” includes data from municipalities, non-profit organizations and private sector sources.

Record Linkage Advances our Knowledge

Examples of health research:

- Socio-economic differences in mortality outcomes
- Health outcomes of immigrants – mortality, hospital use, birth outcomes
- Risk factors associated with hospital utilization and mortality
- Long-term income and employment among cancer survivors
- Impact of long-term exposure to air pollution
- Impact of blood transfusions on long-term mortality

Data Linkage “What STC brings to the table”

Expertise

- leader in the development of methodologies and software for record linkage (GRLS, G-Link)

Legislative framework

- each record linkage must be in accordance with the Statistics Act and Privacy Act
- benefits must clearly outweigh the invasion of privacy

Transparency

- annual Report to Parliament
- summary of all approved record linkages available on website

Accessibility

- to linked data via Research Data Centres
- no charge to university researchers.....just some paperwork!

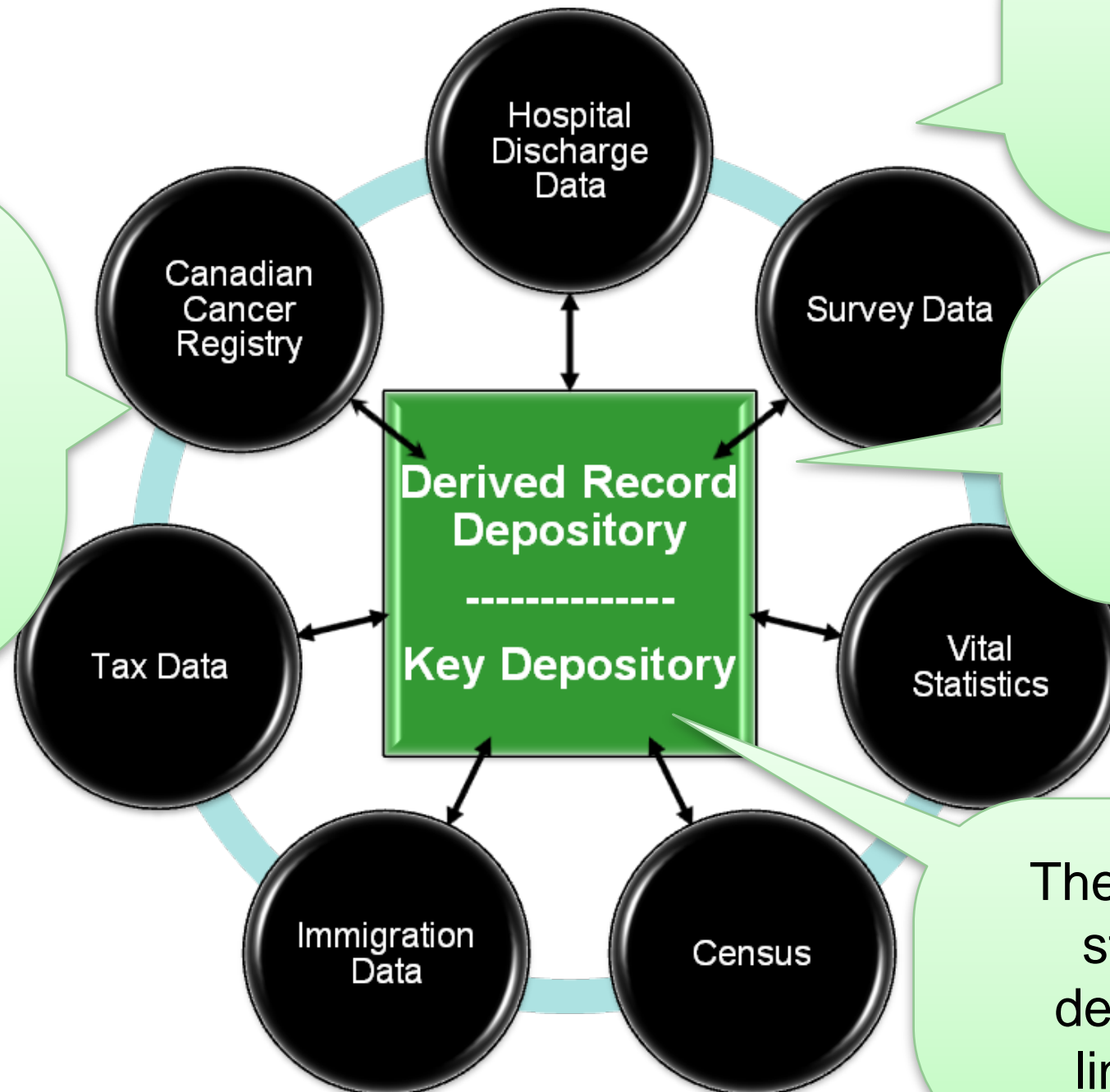
What is the SDLE?

- A new strategy for broad-based socio-economic record linkage work
- An environment for record linkage activities
- Strong governance, adherence to policy and privacy requirements
- It is NOT a large linked data base of integrated survey information
- Two major components: derived record depository and key registry



SDRL Environment

Identifiers of the datasets are linked to the DRD



Only identifiers of datasets are brought into the environment

DRD is built through successive record linkages

The results are stored in a depository of linked keys

Opportunities

- Efficiently link data across multiple sectors in the social domain (i.e. health, justice, education) to complement surveys and data development
- Extend longitudinal surveys by linking respondents to longer term outcomes

National indicators of re-contact

- The re-contact project aims to provide on-going, **high quality indicators** of re-contact with, and **pathways through**, the Canadian criminal **justice system** from policing through corrections
- The availability of nationally comparable and robust indicators of re-contact are essential for:
 - evidence-based **decision-making**;
 - effective **justice policy review**; and
 - to assist in addressing the information needs of criminal justice agencies in evaluating justice **system effectiveness** and **resource utilization**

Re-contact Future

- Plan for future evaluations aimed at:
 - Studying the **profiles of individuals** having repeated contact with the criminal justice system in Saskatchewan
 - Evaluate the potential for **data integration**—in the context of **other social domain data** (e.g. health, education, labour and income statistics)
 - Provide a more **complete understanding** of the **factors associated with repeat contact** with the justice system and assist policymakers and justice administrators in developing **crime control strategies** and **programs** aimed at ensuring **better outcomes for at-risk populations**

Furthering Analysis

- Using the linked keys – get access to data files containing analysis variables
- Extract analysis files
- Derive variables
- Summarize/collapse events/transactions – derive variables
- Challenges to address
 - Longitudinal cohort file
 - Events/transaction data
 - File size; data structure changes over time; unknown quality
 - Getting access to information