Administrative and Linked Datasets

Statistics Canada McMaster Research Data Centre Showcase

Mustafa Ornek

May 22, 2019

Delivering insight through data, for a better Canada

Introduction

- New focus on administrative & linked datasets

- Establishing more partnerships with other federal government departments, provincial governments and institutions

- Incorporating researchers' needs, facilitating custom requests when possible

- Consequently, the administrative and linked data holdings have been increasing exponentially since 2015

What is an administrative dataset?

Advantages:

- Not a survey… not a census

- Has information on the whole population that is covered → Not a sample!

- More accurate than self-reported figures (e.g. income, hospital visits, immigration status & type, social assistance recipients, etc.)

Challenges:

- Not structured as well as Statistics Canada's own datasets
  - Data collection is generally not for the sake of statistical analysis
  - Requires more prep time to understand and clean the data
  - Documentations can be scarce

What is a linked dataset?

- Admin to admin (to admin)

- Census to admin

- Survey to admin

Some of our administrative datasets

- Longitudinal Administrative Databank (LAD)
- Intergenerational Income Database (IID)
- Longitudinal Immigration Database (IMDB)
- Canadian Cancer Registry (CCR)
- Canadian Vital Statistics Database (CVSD)
- A set of ESDC databases:
    - Employment Insurance Status Vector (EISV)
    - Employment Insurance Beneficiaries (EIB)
    - Record of Employment (ROE)
    - Education and Labour Market Longitudinal Linkage Platform (ELMLP)
    - Canada Student Loan Program (CSLP)
    - Canada Disability Savings Program (CDSP)
    - Apprenticeship Grants and Canada Apprenticeship Loan (AG – CAL)

Some of our administrative datasets

- Longitudinal Administrative Databank (LAD)

- Intergenerational Income Database (IID)

- Longitudinal Immigration Database (IMDB)

- Canadian Cancer Registry (CCR)

- Canadian Vital Statistics Database (CVSD)

- A set of ESDC databases:
  - Employment Insurance Status Vector (EISV)
  - Employment Insurance Beneficiaries (EIB)
  - Record of Employment (ROE)
  - Education and Labour Market Longitudinal Linkage Platform (ELMLP)
  - Canada Student Loan Program (CSLP)
  - Canada Disability Savings Program (CDSP)
  - Apprenticeship Grants and Canada Apprenticeship Loan (AG – CAL)

Statistics Canada / Statistique Canada

Canada

Some of our linked datasets

- IMDB – DAD

- CCHS - DAD

- CCHS – IMDB

- CCHS – CVSD

- DAD, NACRS, OMHRS-CVSD: (Discharge Abstract Database - National Ambulatory Care Reporting System - Ontario Mental Health Reporting System linked to Canadian Vital Statistics Death Database)

- CVSD – DAD, NACRS

- CCR-CVSD-T1PMF: (Canadian Cancer Registry linked to Canadian Vital Statistics Death Database and death information from the T1 Personal Master File)

- ERLF: (Extending the Relevance of the Longitudinal Files)

- HSUS-T1FF-CENSUS-NHS: (High System Users linked to T1 Family File - Census of the Population Long-Form - National Household Survey)

Statistics Canada   Statistique Canada

Canada

Need more information?

- See these links:
    - https://crdcn.org/data

    - https://www.statcan.gc.ca/eng/rdc/data
        - → Especially, "**RDC proposals requiring more data related information**" section at the bottom of the page

Thank you

- If you have any questions or comments, you can reach me at:

  mustafa.ornek@canada.ca   or   rdc2@mcmaster.ca