Statistics Canada / **Statistique Canada**

Future Data in the RDCs

www.statcan.gc.ca

Telling Canada's story in numbers

Donna Dosman
Acting Director Microdata Access Division

Friday, March 24, 2017

Canada

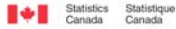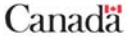**Statistics Canada** / **Statistique Canada**

Canada

Agenda

- StatCan Challenges and Opportunities
- RDC Data Pilots
- RDC Data Collection
- Upcoming RDC Data
  - Administrative Data
  - Linked Data
  - Other Data
- Where to find more information

2

06/04/2017

4/6/2017

Statistics Canada   Statistique Canada

Canadä

Agenda

- RDC Data Collection
- Upcoming RDC Data
  - Administrative Data
  - Linked Data
  - Other Data
- Where to find more information

3

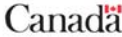06/04/2017

Statistics Canada   Statistique Canada

Canadä

Statistics Canada's Challenges and Opportunities

- Increase effort to find target population
  - Reduction of landlines in households
  - Response burden
  - Reduction in response rate
- Increase costs to obtain desired sample size
- Timeliness of data availability
- Proliferation of data providers and data sources
- New technologies
- Changing privacy lens

4

06/04/2017

Statistics Statistique
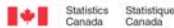Canada Canada

Canadä

RDC Pilot Projects

With new types of data come data pilots

- What is a pilot?
- Why have a pilot?
- Closed vs open pilots?

5   06/04/2017

Statistics Statistique
Canada Canada

Canadä
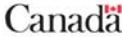
RDC Data Collection

- Household and individual level social, health and economic data
- De-identified respondent level data
- Nearly 400 cycles of social survey and administrative data
  - Cross-sectional → One-time collection and multiple cycles
  - Longitudinal → survey and administrative data
  - Linked data files → survey to administrative and administrative to administrative

6   06/04/2017

Statistics Statistique
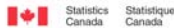Canada   Canada

Canadä

RDC Data Collection: Household Survey Data Subjects

Range of subjects of survey data dating from 1980s to current

- Health, Immunization, Smoking
- Family, Social Identity, Care Giving, Victimization, Time Use, Retirement patterns
- Internet use
- Volunteering
- Education
- Justice
- Labour, Employment, Income
- Aboriginal People
- Immigration
- Census 1911 to current

7   06/04/2017

Statistics Statistique
Canada   Canada

Canadä

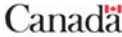# RDC Data Collection: Administrative data files

- Canadian Cancer Registry (1992-2013)
- Vital Statistics Birth Database (1974-2014)
- Vital Statistics Death Database (1974-2014)
- Permanent Resident Landing file (2004-2013)
- Uniform Crime Reporting Survey (2006-2015)
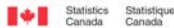- Ontario Ministry of Community and Social Services (2003-2015)

8   06/04/2017

Statistics Statistique
Canada Canada

Canadä

RDC Data Collection: Linked data

- Canadian census mortality and cancer follow-up study
  - Brings together 1991 Census, historic tax summary file (mobility), cancer incidence and mortality
- 1991 Canadian Census Health and Environment Cohort (CanCHEC)
  - Updated Canadian census mortality and cancer follow-up study
- 1996 and 2006 Canadian Birth-Census Cohort (CanBCC)
- 2006 Census linked to Discharge Abstract Database (DAD)

9

06/04/2017

Statistics Statistique
Canada Canada

Canadä

Upcoming Administrative Data

10

06/04/2017

Statistics Statistique
Canada Canada

Canadä

Employment Insurance Beneficiary Claim microdata

- ESDC and STC are collaborating to develop analytic microdata and documentation from the EI beneficiary claim records.
- The microdata will include detailed weekly (status vector) claim data and other claimant information from all available records from 1997 through 2016.
- Project milestones:
  - March 2017:
    - development of a test file and preliminary documentation
    - preparation of record linkage applications.
  - 2017-2018:
    - Development of an analytic microdata and final documentation
    - Submission of record linkage applications for longitudinal person ID, and possibly for family tax data linkage
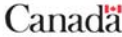
11   06/04/2017

Statistics Statistique
Canada Canada

Canadä

EI Status Vector contents

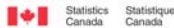| Claimant & dependents	Weekly status	Last job & employer
• Age & gender • Parent type • Family Supplementary benefit • Child birth or adoption date • Economic Region • Postal code (first three digits) • Most recent claim type & status • Claim termination date & reason • Apprentice status • Language • Aboriginal group • Disability type • Visible minority status • Social assistance recipient • Self-Employment Assistance	• Benefit type & amount • Reason for partial or nil benefit • Training • Found work • Earnings & hours worked • Disqualifications • Disentitlement	• Last job occupation (NOC) • Insurable earnings & hours • Last employer industry (SIC/NAICS) • Employer ID (anonymous) • Reason for job separation

12   06/04/2017

4/6/2017

Statistics Canada / Statistique Canada — Canada

The Value of EI Status Vector microdata

- Fills data gaps with detailed weekly benefits and labour market activity at small geographic levels for all EI beneficiaries since 1997
- Will allow analysis of:
  - How EI beneficiaries' respond to changes in program regulations
  - Differences in outcomes for subpopulations
  - Detailed geographic analysis of community effects
- Depending on record linkage approval, researcher may be able to:
  - Create longitudinal histories of all EI beneficiaries since 1997
  - Analyze changes in EI benefit recidivism over time
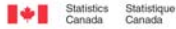 - Study differences between subpopulations in the recidivism of EI benefits

13   06/04/2017

Statistics Canada / Statistique Canada — Canada

Postsecondary Student Information System

- PSIS is a data holding of all public college and university enrolments and graduates by Program/credential type and field of study for each school year.
- Socio-demographic characteristics:
  - Already included: age; sex; student status in Canada (Canadian or international); personal identifiers; province/territory of study
  - Could be imported from other sources: mother tongue, knowledge of official languages, and immigrant status
- The reference year for the data will be 2017-18 and will be available Dec 2019.
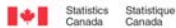- By March 2018, partial PSIS data including – the Maritimes and BC where the data have been verified.

14   06/04/2017

Statistics   Statistique
Canada   Canada

Canadä

Registered Apprenticeship Information System

- RAIS compiles data on the number of registered apprentices taking in-class and on-the-job training in trades that are either Red Seal or non-Red Seal and where apprenticeship training is either compulsory or voluntary.
- It also compiles data on the number of provincial and interprovincial certificates granted to apprentices or trade qualifiers (challengers).
- Socio-demographic characteristics:
  - Already included: age, sex, personal identifiers
  - Could be imported from other sources: mother tongue, knowledge of official languages, and immigrant status
- RAIS in RDCs not before December 2018

15   06/04/2017

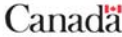Statistics   Statistique
Canada   Canada

Canadä

Ministry of Health Long Term Care Data

- Follow up on a McMaster Data Pilot conducted 2008-2016
- Negotiations have been underway with MOHLTC for the past 2 years
- Phased approach over several years to bring in up to 19 new data sets
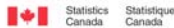- Nearing the signature of agreement for Phase 1

16   06/04/2017

Statistics Statistique
Canada Canada

Canada

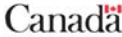## MOHLTC Data: Key data sets of interest

- **OHIP Claims Extract File Database**
- **CIHI Discharge Abstract Dataset (DAD) (i.e. day procedure and inpatient- DAD-DP and DAD-IP for relevant year)**
- **CIHI National Ambulatory Care Reporting System (NACRS) Master Database**
- Registered Persons Database (RPDB)
- Home Care Database (HCD) (from OACCAC)
- Client Profile Database" (CPRO)
- Resident Assessment Instrument-Home Care (RAI-HC)
- Corporate Provider Database (CPDB)
- Client Agency Program Enrolment Database (CAPE)
- Contract Financial Management (CFM)
- Decision Support System (DSS) (to access - Family Health Team (FHT))
- DB2 (to access - Architected Payments System database)

17   06/04/2017

Statistics Statistique
Canada Canada

Canada

Upcoming Linked Data

18   06/04/2017

Statistics Statistique
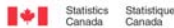Canada Canada

Canadä

Longitudinal Immigration Database (IMDB)

- Record linkage between administrative immigration data and annual tax files
  - Immigrant landing file: Immigrants who have landed in Canada since 1980
  - Non-permanent resident files since 1980:
    - Temporary foreign workers
    - International students
    - Refugee claimants
  - Annual T1FFs: Includes tax files since 1982
  - Amalgamated Mortality Database (AMDB)and annual tax files

87% of immigrants who landed from 1980-2013 linked to at least one tax record from 1982-2013

19

06/04/2017

Statistics Statistique
Canada Canada

Canadä
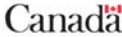
Extending the relevance of discontinued longitudinal files

- Discontinued longitudinal files
  - Youth in Transition Survey (YITS)
  - National Population Health Survey (NPHS)
  - Survey of Labour and Income Dynamics (SLID)
  - National Longitudinal Survey of Children and Youth (NLSCY), and
  - Longitudinal Survey of Immigrants to Canada
- Linked to outcome variables from Cancer, Mortality and Tax administrative files
- Currently conducting validation work of the linkages
- Linked data to be piloted during 2017-18

20

06/04/2017

Statistics Statistique
Canada   Canada

Canada

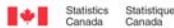## 2006 Census Linked to Discharge Abstract Database

- 2006 Census
  - Short form – used for record linkage
  - Long-form – used for validation and analysis
    - 20% representative sample of the Canadian household population
    - Demography, labour market, income, education, language, disabilities, housing, immigration, ethno-cultural, Aboriginal identity, Registered Indian
- Discharge Abstract Database (DAD) (CIHI):
  - DAD 2005/06 through 2008/09 used for pre-processing
  - DAD 2006/07 through 2008/09 used for record linkage
  - Census of discharges from acute care hospitals (~3 million records/yr) (excludes Quebec)
  - Clinical diagnostic and intervention information, limited demographic
- T1Personal Masterfile (T1PMF)
  - T1 tax returns - historical
  - Annual place of residence (postal codes) to tract mobility over time
  - *No income data included*
- 21 Data are available now in the RDCs   06/04/2017

Statistics Statistique
Canada   Canada

Canada

2000-2011 CCHS-Mortality and Hospital Linked Data

- The primary purpose is to examine mortality and hospital outcomes associated with key risk lifestyle and socioeconomic risk factors.
- Widespread access later in 2017
  - Discharge Abstract Database (DAD) (CIHI):
    - DAD 2005/06 through 2008/09 used for pre-processing
    - DAD 2006/07 through 2008/09 used for record linkage
    - Census of discharges from acute care hospitals (~3 million records/yr) (excludes Quebec)
    - Clinical diagnostic and intervention information, limited demographic
  - T1Personal Masterfile (T1PMF)
    - T1 tax returns - historical
    - Annual place of residence (postal codes) to tract mobility over time
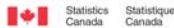   - *No income data included*

22   06/04/2017

Upcoming Other Data

06/04/2017
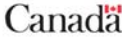
University and College Academic Staff Survey

- UCASS is an annual survey 1937 to 2012 to obtain a national picture of the socioeconomic characteristics and earnings of Full-time university staff
- Cancelled in 2012, but some data continued to be collected outside Statistics Canada
- Recently reinstated and 2016-17 data to be released next spring
- Will expand to also include public college academic staff and part-time academic staff (future)
- Should be available in RDCs in 2017

06/04/2017

Statistics Statistique
Canada Canada
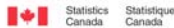
Canadä

Canadian Health Survey on Children and Youth

- Lack of consistent data on children less than 12
- A key objective of the 2015 CCHS redesign was to study options to address this gap
- Most efficient option is a stand-alone survey using the Canadian Child Tax Benefit as a sampling frame
- Pilot test was conducted in fall 2016
- Data file release is planned for fall 2017
- Consultations underway for Cycle 1 content

25

06/04/2017

Statistics Statistique
Canada Canada

Canadä

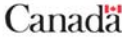## Canadian Health Survey on Children and Youth

- Three cycles are planned (pending funding)
  - Cycle 1
    - Collection Sept. 2018 – June 2019; Data file: ~early 2020
  - Cycle 2
    - Collection Sept. 2021 – June 2022; Data file: ~early 2023
  - Cycle 3
    - Collection Sept. 2024 – June 2025; Data file: ~early 2026
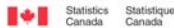- Collection by electronic questionnaires (internet) and telephone interviews

26

06/04/2017

Statistics Statistique
Canada Canada

Canadä

Biobank data coming to RDCs

- Fatty Acid reference ranges
  - Develop fatty acid reference ranges and examine associations with chronic disease.
- Measles and Varicella immunity
  - Measuring immunity of Canadians to measles and varicella and assessing risk of epidemics
- Measurement of metals and trace elements
  - For biomonitoring, developing reference ranges, and to associate levels of contaminants with health outcomes

27   06/04/2017

Statistics Statistique
Canada Canada

Canadä
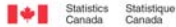
Biobank data coming to RDCs

- Genotyping CHMS Cycles 1 to 4
  - Initial study will identify genetic determinants of environmental toxins and the influence on metabolic disease.
  - This genotyping can then be used for other studies.
  - CHMS would be the largest genome-wide genotyped cohort in Canada and amongst the largest in the world
  - Creation of a national platform of genotype data from about 13,000 Canadians for the better understanding of the biologic determinants of disease

28   06/04/2017

4/6/2017

Statistics Canada / Statistique Canada   Canada

CHMS Biobank
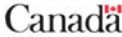
Details on:

Approved studies – completed

"Genetic modifiers of folate, vitamin B-12, and homocysteine status in a cross-sectional study of the Canadian population"

Approved studies – in progress

Approval process and how to access biospecimens

Can be found at :

http://www.statcan.gc.ca/eng/help/microdata/biobank

29   06/04/2017

RDC Data Collection

30   06/04/2017

Statistics   Statistique
Canada   Canada

Canadä

Labour Data

- Labour Force Survey (LFS)
  - 1976-2015
  - Ongoing monthly survey measuring the current state of the Canadian labour market.
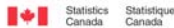 - Used to calculate national, provincial, territorial and regional employment and unemployment rates and to study wages and occupations
  - Data collected from respondents for 6 months
- Survey of Labour and Income Dynamics (SLID)
  - 1993-2011 (longitudinal available up to/including 2010 and cross-sectional)
  - Understanding the economic well-being of Canadians: What economic shifts do indiv/families live through, how does it vary with changes in paid work, family make-up, receipt of gov't transfers or other factors?
  - Data collected from respondents over 6 year period

31   06/04/2017

Statistics   Statistique
Canada   Canada

Canadä
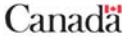
Workplace Data

- Workplace and Employee Survey (WES)

  - 1999-2006

  - Explores issues relating to employers and their employees
  - Sheds light on:
    - relationships among competitiveness, innovation, technology use and human resource management on the employer side
    - technology use, training, job stability and earnings on the employee side
  - More detail on the business and industry than other social surveys

32   06/04/2017

Statistics **Statistique**
Canada Canada

Canadä

Income Data

- Canadian Income Survey (CIS)
  - 2012-2014
  - Ongoing annual cross-sectional survey
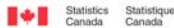 - Provides information on income and income sources of Canadians, and individual and household characteristics
  - Gathers information on labour market activity, school attendance, disability, support payments, child care expenses, inter-household transfers, personal income, and characteristics and costs of housing
  - Household, demographic and geography data from LFS and Tax data used for income and income sources

33   06/04/2017

Statistics **Statistique**
Canada Canada

Canadä

Income Data cont.

- Longitudinal and International Study of Adults (LISA)
  - Wave 1 (2012) and Wave 2 (2014)
  - Collects information about jobs, education, health and family.
  - Contains information on household income and direct measures of literacy, numeracy and problem-solving skills and income data from annual income tax files provided by CRA and other income sources
  - At Wave 1, LISA was integrated with the Programme for the International Assessment of Adult Competencies (PIAAC

34   06/04/2017

Administrative Income Data
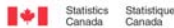
- Longitudinal Administrative Databank (LAD)
  (widespread RDC access in 2017)
  - 1982 to 2014
  - File augmented annually with new data
  - Longitudinal file designed as a research tool on income and demographics
  - Comprises a 20% sample of the annual T1 Family File and some immigration data from the Immigration Landing File
  - Variables have been harmonized across the years where possible and individuals can be linked year to year starting with 1982 data
    - Ethnic diversity and immigration
    - Household, family and personal income
    - Income, pensions, spending and wealth
    - Labour market and income
    - Personal and household taxation

35   06/04/2017

Statistics Statistique
Canada Canada   Canada

Administrative Income Data

- Ontario Social Assistance Data
  (widespread RDC access in 2015)
  - 2003 to 2013.
  - Administrative records from the income and employment program designed to support to single adults and families who are in financial need.
  - Data include information on the benefit unit (family) and recipient, transactional information and skills of recipient.
  - Researchers will also have access to monthly time-series data on Ontario Works and Ontario Disability Support Program (ODSP) provincial caseloads from 1990 to 2013
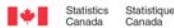
36   06/04/2017

Statistics Statistique
Canada Canada

Canadä

Immigration, Refugees and Citizenship Canada Data

- Permanent Resident Landing File
  - Widespread access
  - File contains approximately 2.75 million records corresponding to all individuals who landed in Canada during 2003 – 2013
  - Variables include occupation, skill levels, NOC code (2006 and 2011)
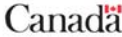
37   06/04/2017

Statistics Statistique
Canada Canada

Canadä

Administrative Data Linked to Survey Data

- Canadian Birth-Census cohorts
  - Widespread access
  - Long-form Census from 1996 and 2006
  - Long-form Census information linked to vital statistics data on births, stillbirths and infant deaths in Canada,
  - Objective is to provide socio-economic information about the household with infant mortality data
- Canadian Census Health and Environment Cohorts (CanCHEC)
  - Widespread access
  - Long-form Census 1991 and 2001
  - Cohort of census year linked with Mortality and Cancer data along with historical postal codes
  - Objective is to provide data to examine mortality and cancer outcomes by census characteristics and geography. Additional environmental measures (such as air pollution) can be added to the file.

38   06/04/2017

Statistics Statistique
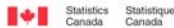Canada Canada

Canada

Administrative Justice Data

- **Uniform Crime Reporting (UCR) Survey**
  - Widespread access
  - Data are from 2006-2015
  - Designed to measure the incidence of crime in Canadian society and its characteristics. Data are collected directly from police services and extracted from administrative files

- **Homicide Survey**
  - Currently limited access
  - Data are from 1961-2011
  - A census administrative survey completed for each police-reported homicide incident occurring in Canada.

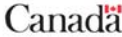39   06/04/2017

Statistics Statistique
Canada Canada

Canada

Administrative Justice Data cont.

- **Hate Crime** (Uniform Crime Reporting Survey Modules)
  - Currently limited access
  - Data are from 2009, 2010 and 2011
  - Identifies criminal incidents reported by police as being motivated by hate based on race, national or ethnic origin, language, colour, religion, sex, age, mental or physical disability, sexual orientation or any other similar factor (such as occupation or political beliefs)

- **Integrated Criminal Courts Survey (ICCS)**
  - Currently limited access
  - Data are from 2005/06 – 2001/12
  - Designed to collect statistical information on adult and youth court cases involving Criminal Code and other federal statute offences in Canadian courts, and their characteristics.

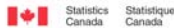40   06/04/2017

Statistics Canada · Statistique Canada   Canadä

Information on New Data Availability

- **English Twitter Account:**
  Handle: @StatCan_eng
  URL: https://twitter.com/StatCan_eng

- **French Twitter Account:**
  Handle : @StatCan_fra
  URL: https://twitter.com/StatCan_fra

41   06/04/2017

Statistics Canada · Statistique Canada   Canadä

Information on the RDC Program and Data Availability

Statistics Canada RDC
- http://www.statcan.gc.ca/eng/rdc/index

Canadian Research Data Centre Network
- http://www.rdc-cdr.ca/

Up-to-date list of RDC data
- http://www.statcan.gc.ca/eng/rdc/data

List of all Statistics Canada data
- http://www23.statcan.gc.ca/imdb-bmdi/pub/indexth-eng.htm

42   06/04/2017