TDR TriUniversity DATA RESOURCES

Centres de données de recherche
Research Data Centres
CDR RDC

March 2006

The DLI and RDC Program: Valuable Research Resources for WIHIR Researchers

- Sandra Keyes, DLI Representative, UW
- Pat Newcombe-Welch, Statistics Canada Data Analyst, SWORDC
- G. Keith Warriner, Co-director, South-Western Ontario Research Data Centre

March 8, 2006

Outline

DLI

- What is it?
- Local access points

Microdata Files

- PUMF’s, Synthetic Files, Master Files
- Where to get access?
- Which file to use?
- Implications for analysis

RDC

- What is a Research Data Centre?
- Where is it?
- How does one access data?
- What can you access?

Data Liberation Initiative

DLI

- Provides Canadian academic institutions access to Statistics Canada data files and databases for teaching and research.
- 10 years old
- Institutions pay an annual fee

Local access to DLI files

Institutions have local DLI rep

- Usually associated with a data service or centre (LDC)
- Act as liaison between DLI and local data service

Local Services

- Provide access to metadata (codebooks…)
- Assist with finding data sets and variables
- SPSS and SAS system files from the raw data files that are distributed through DLI
- Assist with statistical packages
- Provide information about the use of Remote Job Submission and RDC’s

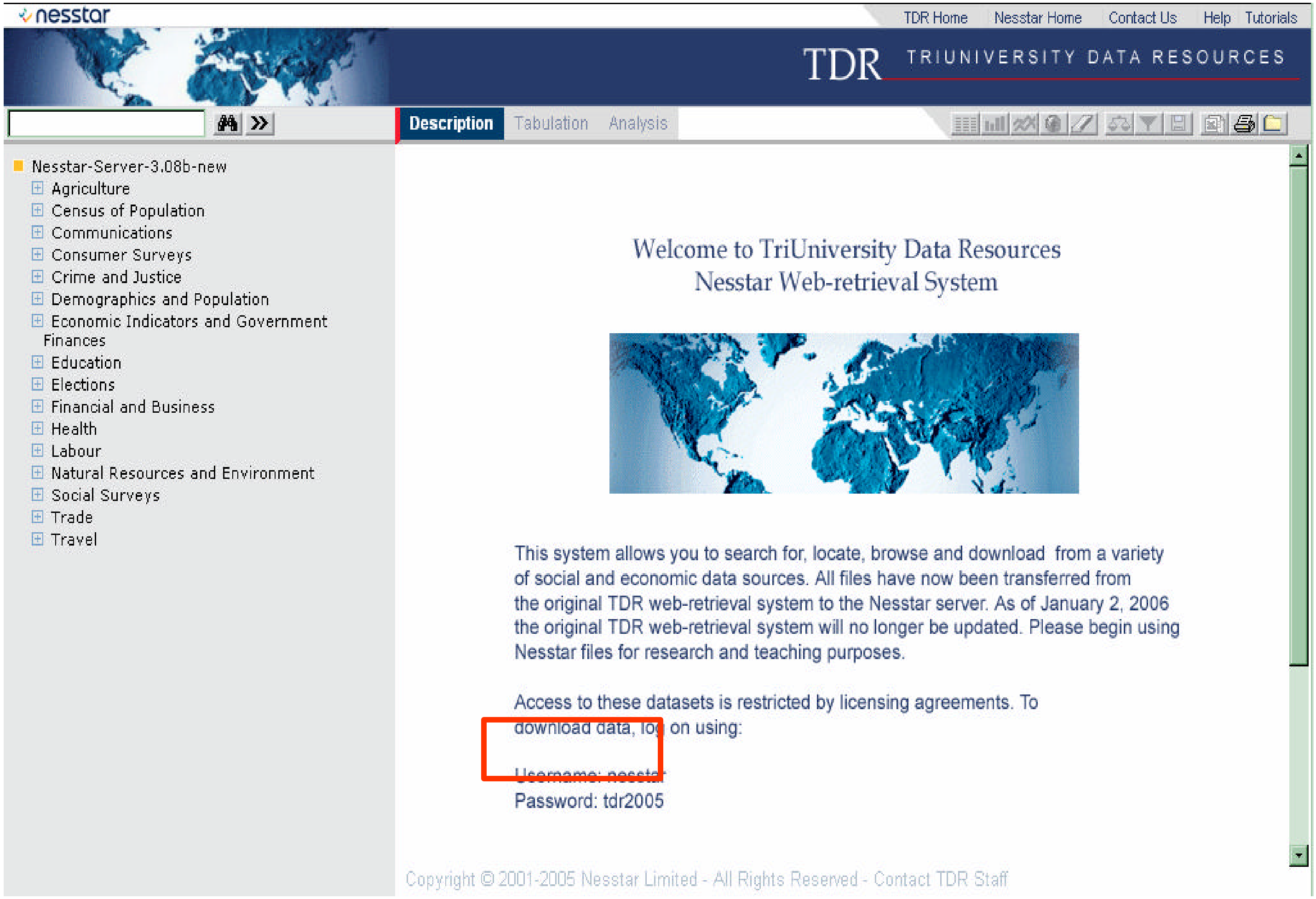
Microdata Files

- Hundreds of data sets
- Census, GSS, SCF, NLSCY, YITS, WES, SHS, SLID,CCHS, HIUS, LFS…
- Information on age, gender, income, internet use, marital status, education, ethnic origin, home ownership, VCR’s, attitudes, health …

Overview of Nesstar

- The new interface for these Surveys
- Search or browse surveys
- Subset the survey
- Download the data for analysis
- Access the documentation

Overview of Nesstar



- Nesstar-Server-3.08b-new
  - ⊞ Agriculture
  - ⊞ Census of Population
  - ⊞ Communications
  - ⊞ Consumer Surveys
  - ⊞ Crime and Justice
  - ⊞ Demographics and Population
  - ⊞ Economic Indicators and Government Finances
  - ⊞ Education
  - ⊞ Elections
  - ⊞ Financial and Business
  - ⊞ Health
  - ⊞ Labour
  - ⊞ Natural Resources and Environment
  - ⊞ Social Surveys
  - ⊞ Trade
  - ⊞ Travel

Welcome to TriUniversity Data Resources
Nesstar Web-retrieval System

This system allows you to search for, locate, browse and download from a variety of social and economic data sources. All files have now been transferred from the original TDR web-retrieval system to the Nesstar server. As of January 2, 2006 the original TDR web-retrieval system will no longer be updated. Please begin using Nesstar files for research and teaching purposes.

Access to these datasets is restricted by licensing agreements. To download data, log on using:

Username: nesstar
Password: tdr2005

Copyright © 2001-2005 Nesstar Limited - All Rights Reserved - Contact TDR Staff

Types of Microdata Files

Public Access Microdata Products (LDC)

- Public use anonymized microdata (PUMF)
- Synthetic Files

Confidential Microdata Products (RDC)

- Master Files

PUMF

PUMF Microdata

- raw data organized in a file where the records or lines in the file are observations of a specific unit of analysis and the information on the lines are the values of variables
- requires some form of processing or analysis to be used
- Shouldn’t be able to identify individuals

Synthetic Microdata

Synthetic Files

- Looks like master files
- Lots of observations (maybe)
- Lots of variables
- Little grouping or capping of categories
- Lots of geographic detail

- **Not ‘real’ observations**

Synthetic Microdata

Synthetic Files

- Author Divisions ‘may’ create it
- Most relevant when dealing with new Panel Data, e.g....
    - NLSCY, NPHS & CCHS

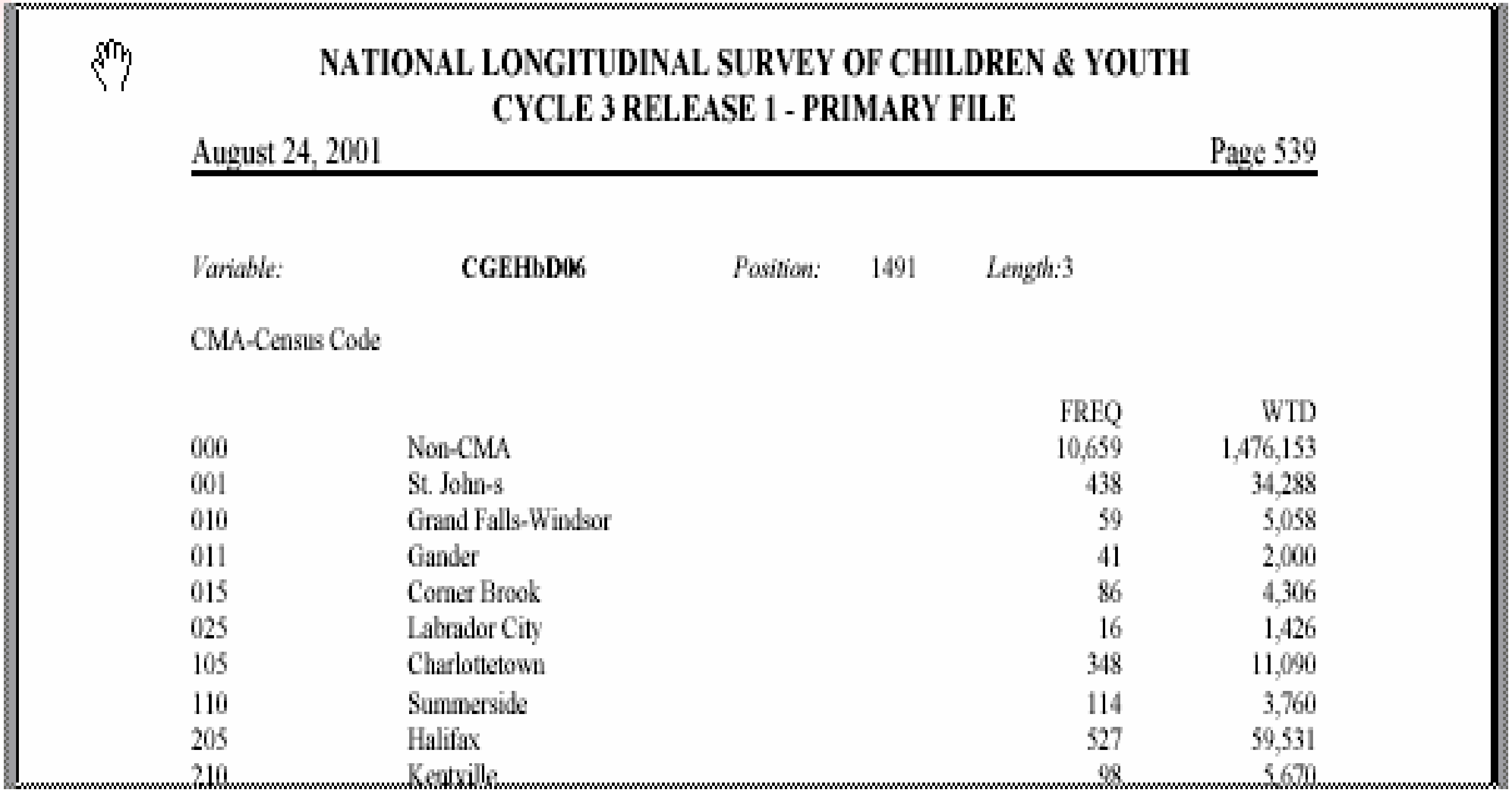
Confidential Microdata

Master Files

- files contain the fullness of detail captured about the unit of observation.
- information in these files could identify the individual who provided the original information and, therefore, are considered confidential.

Confidential Microdata

Master File – geography



NATIONAL LONGITUDINAL SURVEY OF CHILDREN & YOUTH
CYCLE 3 RELEASE 1 - PRIMARY FILE

August 24, 2001 Page 539

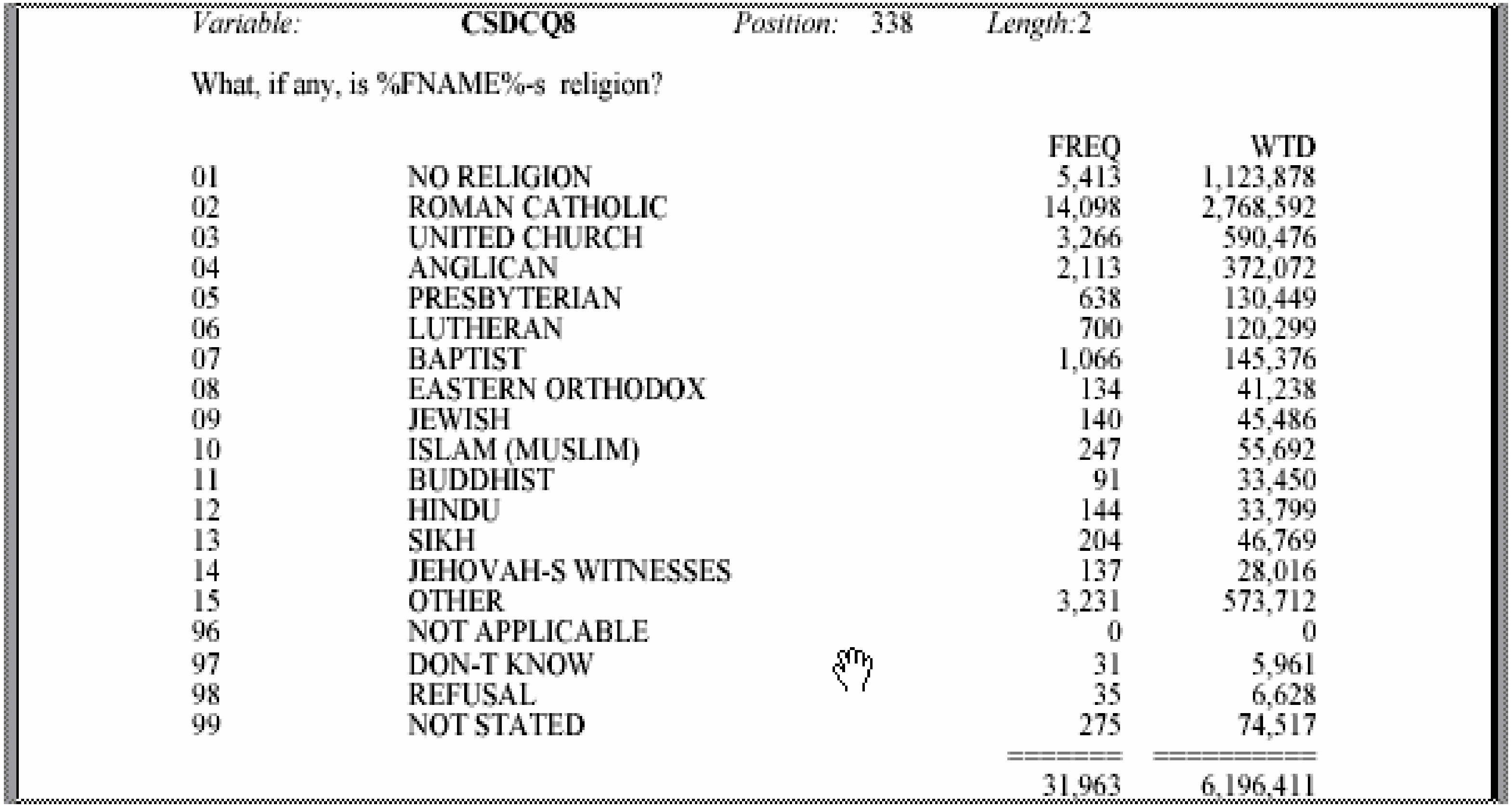
Variable: **CGEHbD06** *Position:* 1491 *Length:*3

CMA-Census Code

		FREQ	WTD
000	Non-CMA	10,659	1,476,153
001	St. John-s	438	34,288
010	Grand Falls-Windsor	59	5,058
011	Gander	41	2,000
015	Corner Brook	86	4,306
025	Labrador City	16	1,426
105	Charlottetown	348	11,090
110	Summerside	114	3,760
205	Halifax	527	59,531
210	Kentville	98	5,670

Confidential Microdata

Master File - fullness of data



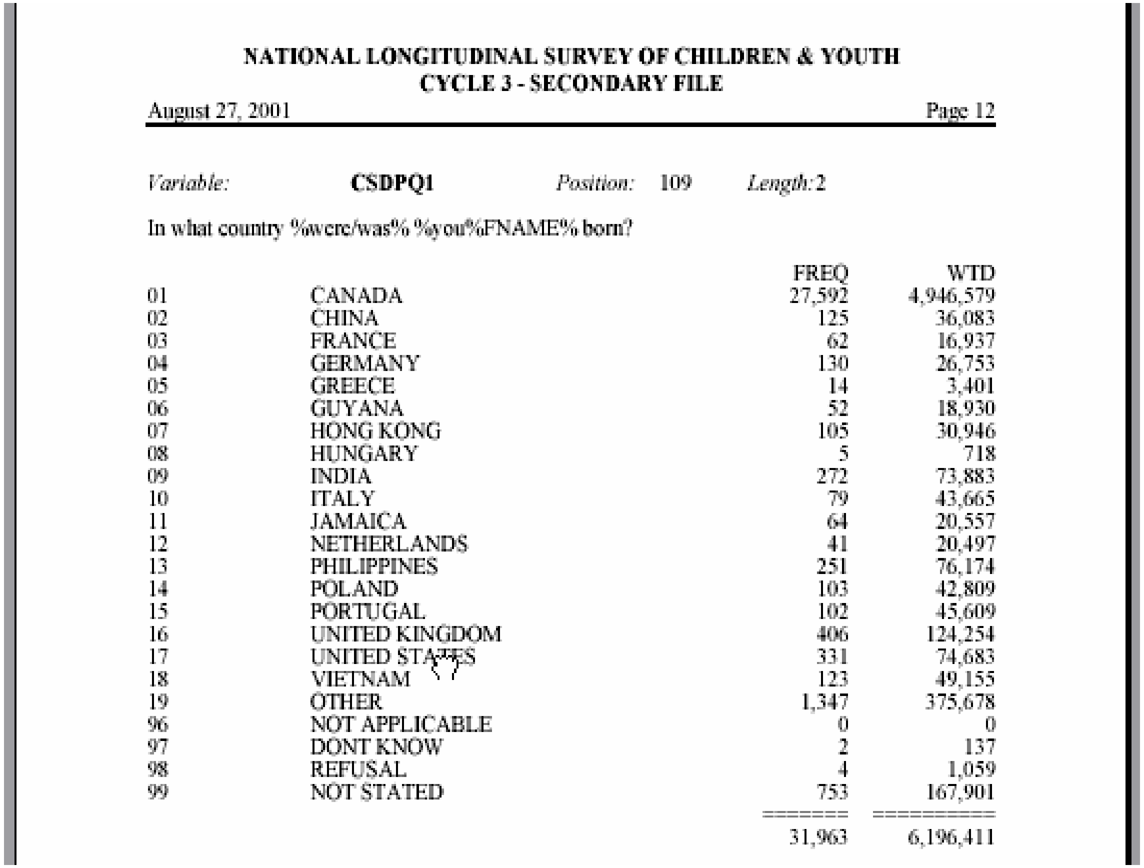
Variable: **CSDCQ8** *Position:* 338 *Length:*2

What, if any, is %FNAME%-s religion?

		FREQ	WTD
01	NO RELIGION	5,413	1,123,878
02	ROMAN CATHOLIC	14,098	2,768,592
03	UNITED CHURCH	3,266	590,476
04	ANGLICAN	2,113	372,072
05	PRESBYTERIAN	638	130,449
06	LUTHERAN	700	120,299
07	BAPTIST	1,066	145,376
08	EASTERN ORTHODOX	134	41,238
09	JEWISH	140	45,486
10	ISLAM (MUSLIM)	247	55,692
11	BUDDHIST	91	33,450
12	HINDU	144	33,799
13	SIKH	204	46,769
14	JEHOVAH-S WITNESSES	137	28,016
15	OTHER	3,231	573,712
96	NOT APPLICABLE	0	0
97	DON-T KNOW	31	5,961
98	REFUSAL	35	6,628
99	NOT STATED	275	74,517
		======	=========
		31,963	6,196,411

Confidential Microdata

Master File - fullness of data



NATIONAL LONGITUDINAL SURVEY OF CHILDREN & YOUTH
CYCLE 3 - SECONDARY FILE

August 27, 2001 Page 12

Variable: **CSDPQ1** *Position:* 109 *Length:*2

In what country %were/was% %you%FNAME% born?

		FREQ	WTD
01	CANADA	27,592	4,946,579
02	CHINA	125	36,083
03	FRANCE	62	16,937
04	GERMANY	130	26,753
05	GREECE	14	3,401
06	GUYANA	52	18,930
07	HONG KONG	105	30,946
08	HUNGARY	5	718
09	INDIA	272	73,883
10	ITALY	79	43,665
11	JAMAICA	64	20,557
12	NETHERLANDS	41	20,497
13	PHILIPPINES	251	76,174
14	POLAND	103	42,809
15	PORTUGAL	102	45,609
16	UNITED KINGDOM	406	124,254
17	UNITED STATES	331	74,683
18	VIETNAM	123	49,155
19	OTHER	1,347	375,678
96	NOT APPLICABLE	0	0
97	DONT KNOW	2	137
98	REFUSAL	4	1,059
99	NOT STATED	753	167,901
		=======	==========
		31,963	6,196,411

Where do we get Access?

PUMF

- Obtain from DLI
- Available through LDC - TDR
- Analyze where it is convenient
- Can use a variety of analysis software, including SAS, SPSS, Stata, HLM, LISREL, etc.

Where do we get Access?

Synthetic Files

- If created by 'Author Division'
- Available through LDC - TDR
- Work locally with the file
- Build SAS and SPSS setups

Where do we get Access?

Master File

- Restricted access governed under the Statistics Act;
- Research Data Centres - **SWORDC**

Southwestern Ontario Research Data Centre

Located at the University of Waterloo

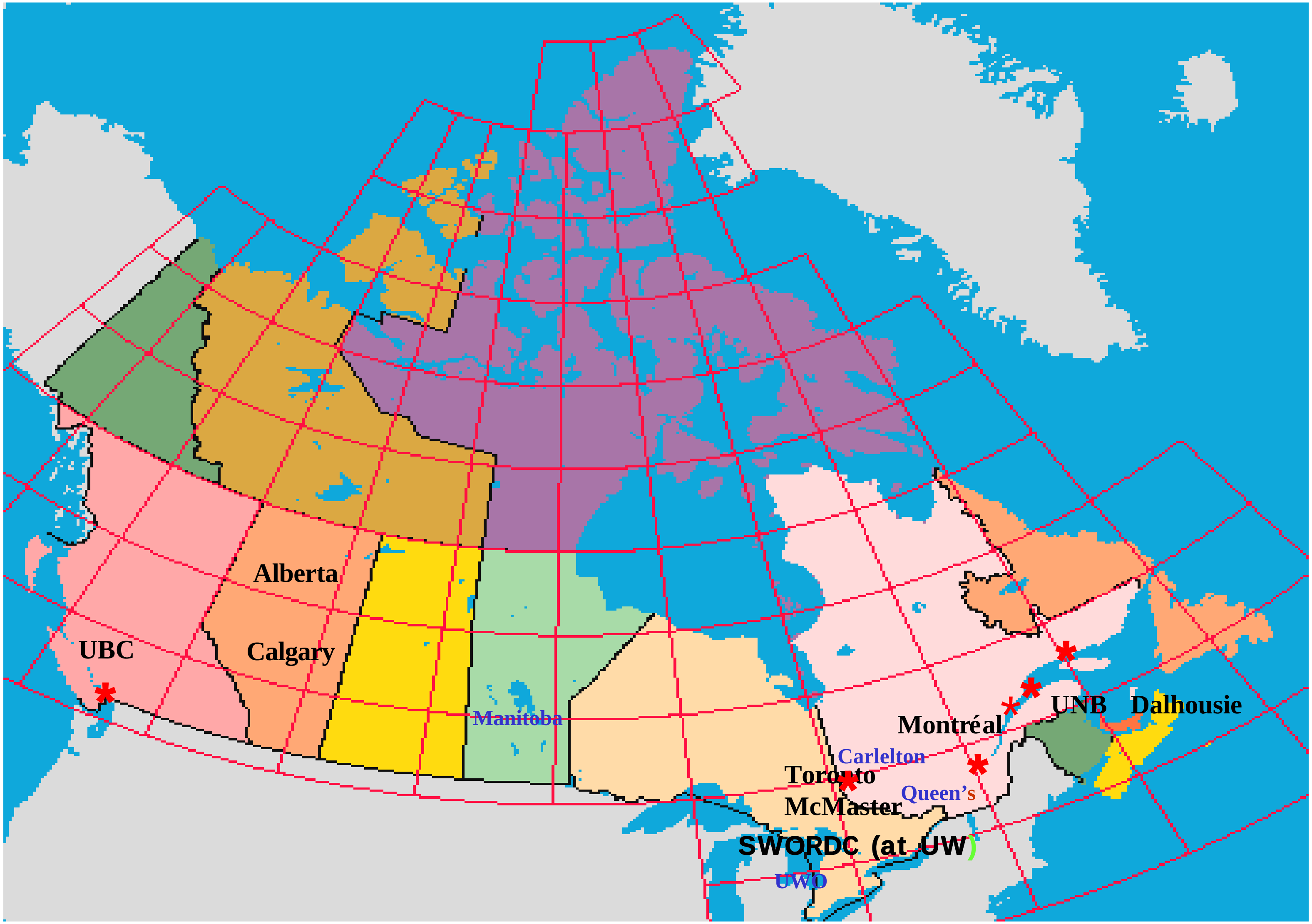
Context

- New longitudinal surveys for which it is not practical to produce public use microdata files
- An increasing need for detailed microdata to analyse crucial social and socio-economic issues
- RDCs Recommended by joint SSHRC/Statistics Canada Task Force - Canadian Initiative on the Social Sciences

What is a Research Data Centre?

- Secure Statistics Canada environment in a university setting
- Houses Statistics Canada microdata files
- Staffed by a Statistics Canada employee at all times
- Operates under the provisions of the Statistics Act
- Access limited to researchers with approved projects and "sworn-in" under Statistics Act as "deemed employees"

Locations


UBC
Alberta
Calgary
Manitoba
Montréal
UNB
Dalhousie
Carlelton
Toronto
Queen's
McMaster
SWORDC (at UW)
UWO

Fosters Relevant Research

- Secure computing environment
- On-site Statistics Canada analyst - support
- Links to specialists in Ottawa
  - analytical support
  - data support
- RDCs link with resources in the host institution

Data Holdings

- National Longitudinal Survey of Children and Youth (NLSCY)
- Survey of Labour and Income Dynamics (SLID)
- National Population Health Survey (NPHS)
- Youth in Transition Survey (YITS)
- Workplace and Employee Survey (WES)
- Canadian Community Health Survey (CCHS)
- Ethnic Diversity Survey (EDS)
- General Social Surveys (GSS)
- Longitudinal Survey of Immigrants to Canada (LSIC)
- Survey of labour and Income Dynamics (SLID)
- Aboriginal People's Survey (APS)
- Health Services Access Survey (HSAS)

http://www.statcan.gc.ca/english/rdc/whatdata.htm

Data Holdings

- Additional
  - If a proposal which demonstrates the need for data unavailable via Public Use Micro Files (PUMFs) is approved by SSHRC, the data will be made available in the RDC

Recent Health-Related Projects

- **Human And Social Capital Indicies As Predictors Of Health Compromising Behaviours In 14-15 Year Old Youth (NLSCY)**
- **Association Between Indicators Of Health And Well-being And Over-weight Or Obesity In Adolescents Aged 12-15y (NLSCY)**
- **Effects of Unemployment on Individual Health Status (NPHS)**
- **Depression And Incident Diabetes (NPHS)**
- **Longitudinal Study Of Family Relationships As Mediators Of Children's Schooling Outcomes (NLSCY)**
- **The Biochemistry Of Florine In Well Water In Oxford County, Ontario (CCHS)**
- **Problem Behaviour At Ages 10-17 As A Predictor Of Identity Decision-making Styles At Age 16-17 (NLSCY)**
- **Psychometrics of Sense of Coherence Scale (NPHS)**

Health-Related Projects (cont'd)

- **Exploring Risk Factors Affecting Mortality (NPHS)**
- **Adolescent Sexuality And Sexual Health: Transition Into Sexual Adulthood (NPHS/NLSCY)**
- **Determinants Of Unintentional Child Injuries: Child, ,Parent And Environmental Characteristics (NLSCY)**
- **The Impact Of Provincial Health Care Expenditures On The Health Of Canadians (NPHS)**
- **Socioeconomic Status And Accute Non Fatal Injury Among Canadian Adolescents (NPHS)**
- **The Development Of A Social Bond: Differences And Similarities For Early Aggressive Children Compared To Other Children (NLSCY)**
- **Modeling Health Production Function And Health Inequality Based On Nonparametric Quantile Regression (NPHS)**
- **Family Structure And Child Outcome: A Longitudinal Study Of The Sexual Behaviour Of Canadian Children (NLSCY)**

Health-Related Projects (cont'd)

- **Can We Explain Leisure-time Physical Activity By Family Demands, Gender And Employment Status? (CCHS)**
- **Does Pregnancy Trigger Smoking Cessation? Inferring Time Order From Longitudinal Data (NPHS)**
- **An Examination Of The Impact Of Municipal Smoking Restrictions On Behaviour Of Current Smokers (CCHS)**
- **Effects Of Physically Active Leisure, Social Support, Work Stress And Chronic Stress On Mental And Physical Health: A Longitudinal Perspective (NPHS)**
- **Child Health Care Utilization In Canada (NLSCY)**
- **Canadian Adolescent's Mental Health And Participation In Physically Active Leisure: Cross-sectional And Longitudinal Perspectives (NPHS)**
- **Prediction Of Smoking Cessation And Other Health Behaviours (NPHS)**

Access to the Research Data Centres

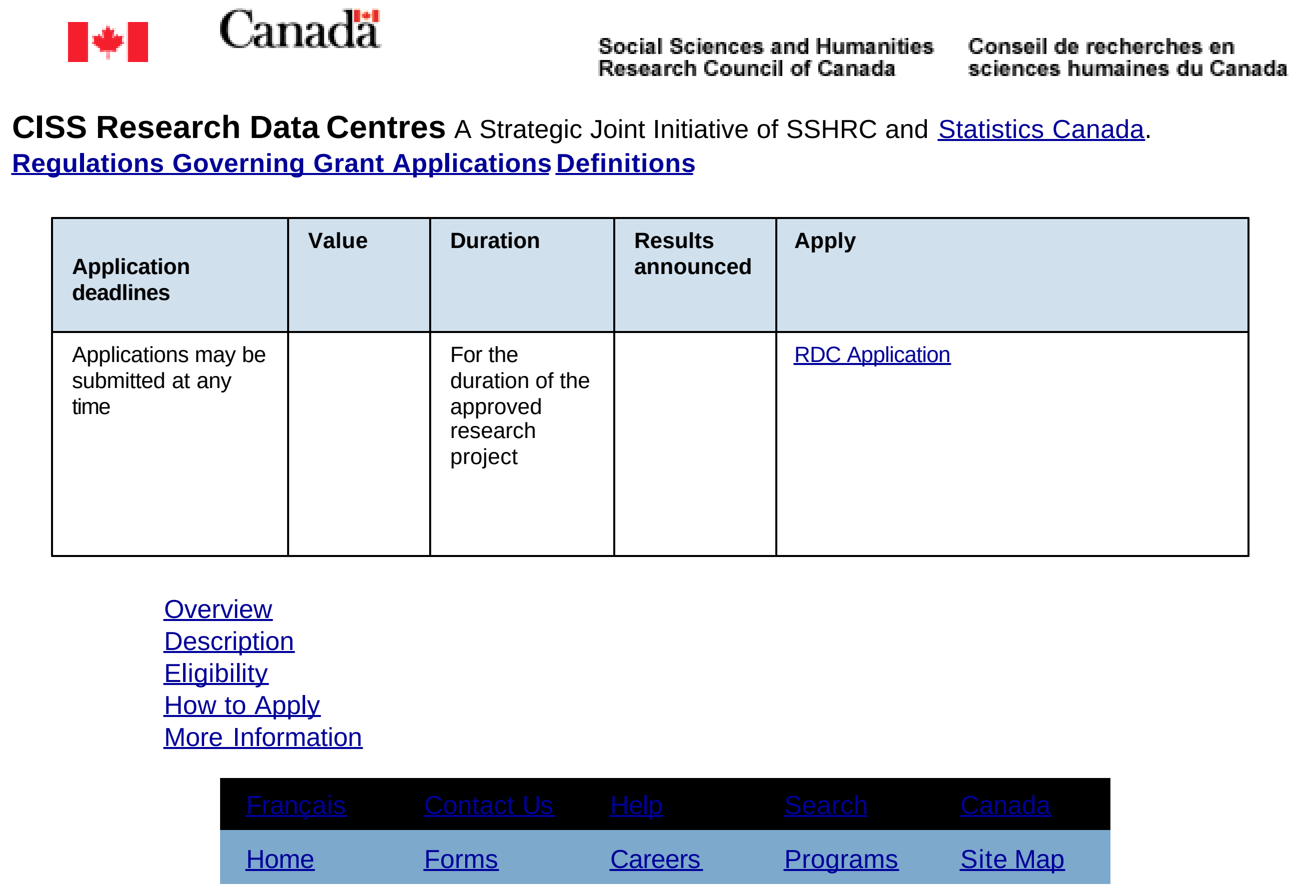
- Project proposal
- Proposal evaluation - SSHRC
- Security clearance - enhanced reliability check
- Orientation session and “oath of office”
- Researcher agrees to provide publicly available report that falls within Statistics Canada’s mandate

Evaluation Criteria

The proposal must successfully demonstrate:

- Clearly defined project objectives
- Need for detailed microdata file rather than PUMFs
- That the data will support the proposed research
- Scientific merit of the project, suitability of analytical and statistical methods
- That the applicant and co-researchers have the experience, qualifications, and expertise to successfully complete the proposed project

Project Proposal



Canada

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

CISS Research Data Centres A Strategic Joint Initiative of SSHRC and Statistics Canada.

Regulations Governing Grant Applications **Definitions**

Application deadlines	Value	Duration	Results announced	Apply
Applications may be submitted at any time		For the duration of the approved research project		RDC Application

Overview
Description
Eligibility
How to Apply
More Information

Français	Contact Us	Help	Search	Canada
Home	Forms	Careers	Programs	Site Map

Confidentiality and Security

💻 Stand-alone computing system which has no link outside of Statistics Canada

✓ Access limited to researchers with approved projects who have enhanced reliability check and are “sworn-in” under the Statistics Act

🔒 Physical security of site equivalent to that in Statistics Canada.

Confidentiality and Security (cont'd)

📖Orientation session and handbook

☺Statistics Canada analyst on-site

👓Disclosure analysis on all products leaving the RDC

Getting Started

- 1st stop is still the LDC and the PUMF
- This file has the easiest access
- Probably meets the needs of most patrons
- Not as administratively burdensome as synthetic or master file
- If more detail is needed, refer to the Master File Documentation

Implications for Analysis

Public Use Microdata (PUMF)

- Valuable content for a tremendous amount of research
- Suppressed observations
- Suppressed variables
- Suppressed Content
  - Gross Geography
  - Collapsed categories
  - Capped variables
- Where issues arise is when smaller area geography is desired; rare subpopulations are being studied; or the variables that are needed have been used to anonymize respondents
- Licensed product: agree to certain terms of use
- No linkage to multiple units of analysis, except for a few exceptions (e.g., GSS Time Use and Family)

Implications for Analysis

Synthetic Files

- Looks like master files
- Lots of observations (maybe)
- Lots of variables
- Little grouping or capping of categories
- Lots of geographic detail

Precautions

- Results not authentic – but may be close in the aggregate for some synthetic files
- Use for testing analysis setups only
- Still need the master files for publishable results

Implications for Analysis

Master File

- All observations
- Has the most variables with the most detail
  - Lots of geography and personal characteristics
  - Little grouping or capping of categories
- Restricted access: only available to authorized Statistics Canada employees, which includes 'deemed employees'
- Use of the analysis is controlled through a contract
- Includes linkage variables across files within a study, e.g., NLSCY linkage among the files for different units of analysis (kids, parents, teachers…)

More Information?

Contacts

- Sandra Keys, DLI Librarian, UW

  skeys@library.uwaterloo.ca

- Keith Warriner, Co-director, SWORDC

  wnrr@uwaterloo.ca

- Pat Newcombe-Welch, Data Analyst, SWORDC

  panewcombe@uwaterloo.ca

- SWORDC website:

  http://tdr.uoguelph.ca/DATA/WWWDOCS/SWORDCSITE/splash.html

http://tdr.uoguelph.ca/DATA/WWWDOCS/SWORDCSITE/splash.html