



Statistics Canada
www.statcan.gc.ca

Dataset Builder Tool

**Canadian Research Data Centre
Network**

Statistics Canada

NADDI 2014



Statistics
Canada Statistique
Canada

Canada

Outline

- Background of project
- Converted metadata
- Process of project
- Tools developed
- Licensing agreement
- Future metadata plans
- Benefits of Dataset Builder Tool
- Demonstration of tool

Background



- In 2008 CRDCN received CFI funding
- Project completed April, 2013

- Objectives
 - Retrospective conversion of the metadata for the StatCan surveys in the RDC's to the DDI-Lifecycle standard
 - Develop an editor tool to produce DDI-Lifecycle metadata
 - Develop a search, discover and build tool for researchers

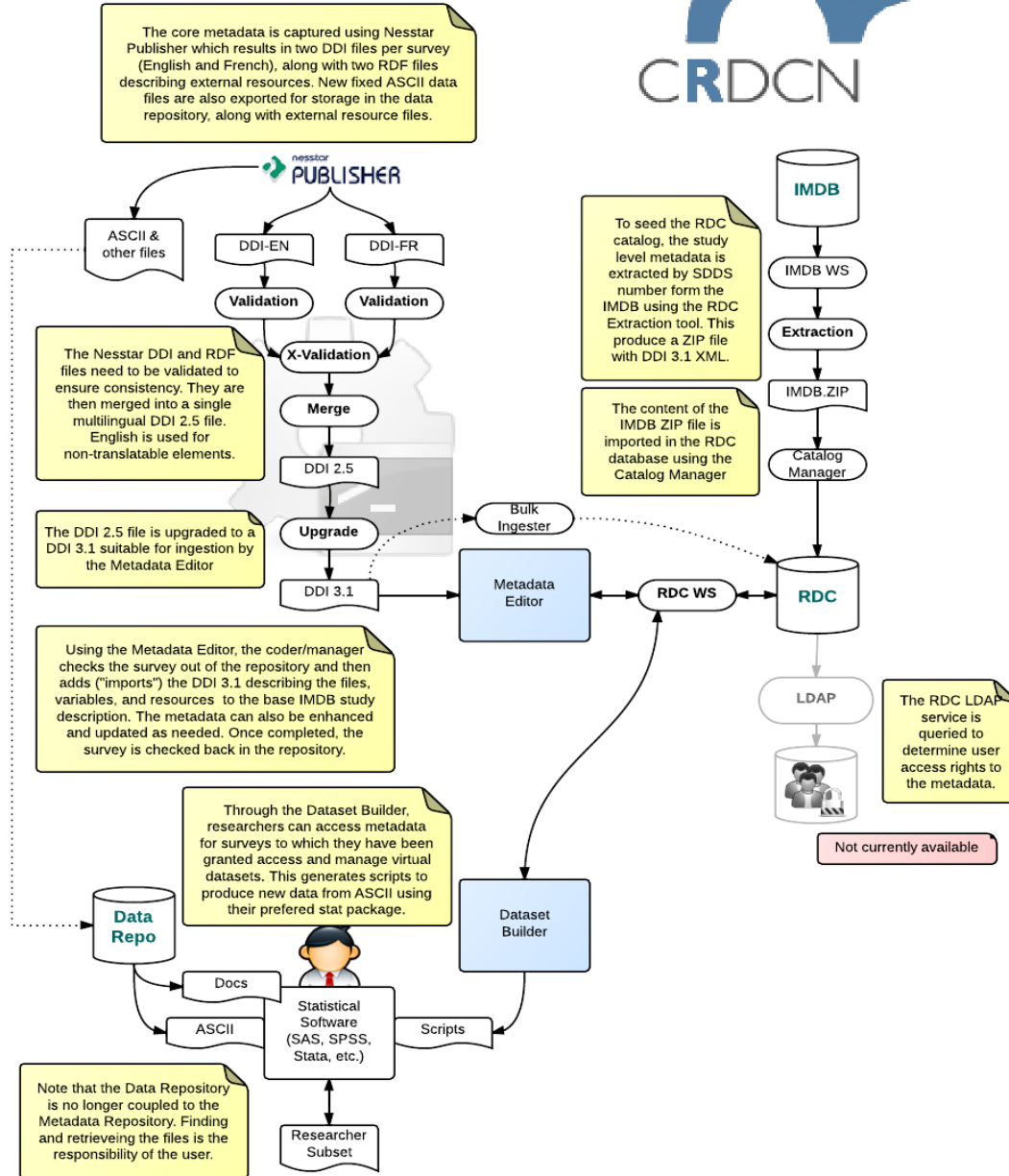
Converted Metadata



- Coded 100,000+ variables
- 200,000+ variables in both English and French

- Coded into DDI-Codebook:
 - 243 cycles of data
 - 39 surveys

- Upgraded into DDI-Lifecycle:
 - 223 cycles of data
 - 38 surveys



Tools Developed



- Tools used in coding/editing process
 - DDI registry/catalog
 - DDI base Editing suite and enhanced editor
 - Import/export tools
 - Quality assurance utility
 - IMDB – DDI-L conversion tool

- Dataset Builder Tool

Licensing Agreement



- Copyright © 2013 Metadata Technology North America Inc.

<http://www.mtna.us> | info@mtna.us

- Open source license: GNU Lesser General Public License

Future Metadata Plans



- **Statistics Canada is:**
 - Short term
 - Coding 60 new cycles of data
 - Placing the DDI – Codebook metadata in Nesstar on DLI site in a separate collection
 - Longer term
 - Developing a metadata strategy in which DDI may be named as one of the standards
 - Developing a common tool which exports metadata (codebook) into xml
- **CRDCN plans to:**
 - Eventually host the dataset builder on the CRDCN website

Benefits of Dataset Builder Tool



- Provides a single tool for researchers to search and discover survey metadata
- Survey metadata can be used to create better research proposals
- Ability to subset data and create the associated codebook/data dictionary documentation
- Makes research process more efficient by saving time and effort

Contact Information



- Metadata Technology North America
<http://www.mtna.us> | info@mtna.us
- Pamela Moren, Statistics Canada
pamela.moren@statcan.gc.ca
- Donna Dosman, Statistics Canada
donna.dosman@statcan.gc.ca



Demo of Dataset Builder

Dataset Builder: Homepage

The screenshot displays the 'Report center' application window. The left pane shows a 'Catalog Explorer' with a tree view where 'Census of Population - 2006' is selected. The main pane shows the 'Overview' tab for this dataset, including an abstract, purpose, and client information. The right pane shows a 'Shopping Cart' with an empty table.

Report center

Catalog Explorer Filter

Search the catalog

- Variable Explorer
- Group
 - Personal
 - Master
 - Census of Population
 - Census of Population - 1996
 - Census of Population - 2001
 - Census of Population - 2006**
 - Study Unit
 - Concept
 - Universe
 - Question
 - Classification
 - Variable
 - Data Set
 - Organization
 - Other Materials

- Labour Force Survey
- National Population Health Surve
- Canadian Community Health Sur
- Survey of Financial Security
- Survey of Household Spending
- Health Promotion Survey
- Survey of Graduates of Post-seco
- Survey of Displaced Workers
- Adult Education Survey
- International Adult Literacy Surve
- National Private Vehicle Use Surve
- Longitudinal Survey of Immigrant
- Employment Insurance Coverage
- National Survey of Giving, Volunt

Overview Variables

Census of Population - 2006

"Census of Population - 2006".

Abstract

Statistics Canada conducts the Census of Population in order to develop a statistical portrait of Canada and Canadians on one specific day. The census is designed to provide information about people and housing units in Canada by their demographic, social and economic characteristics.

The Census of Population is a reliable basis for the estimation of the population of the provinces, territories and local municipal areas. The information collected is related to more than 80 federal and provincial legislative measures and provides a basis for the distribution of federal transfer payments. The census also provides information about the characteristics of the population and its housing within small geographic areas and for small population groups to support planning, administration, policy development and evaluation activities of governments at all levels, as well as data users in the private sector.

Clients: Federal government, provincial and territorial governments, municipal governments; libraries; educational institutions; researchers and academics; private industry; business associations labour organisations; ethnic and cultural groups; private citizens; public interest groups.

Purpose

The census is designed to provide information about people and housing units in Canada by their demographic, social and economic characteristics.

System Information

Shopping Cart Outline

New set Delete Rename Generate scripts

Name	Label	Study
------	-------	-------

Total : 0

Dataset Builder: Variable Basket

The screenshot displays the 'Report center' application window, specifically the 'Variables' tab. The interface is divided into several sections:

- Left Panel (Catalog Explorer):** A list of survey datasets with checkboxes. Two surveys are selected: 'Survey of Financial Security - 2005' and 'Survey of Household Spending - 2002'.
- Search Bar:** A search box containing the text 'income', with 'Options' and 'Show' buttons to its right.
- Variables List:** A table showing search results for 'income'. It includes columns for 'Study', 'Name', and 'Label'. The first entry is 'Survey of Financial Security - 2005' with variable 'TTINC27' and label 'Ef - total income'.
- Navigation:** 'Page 1 of 2' with navigation buttons (First, < 5, << Previous, 1, Next >>, 5 >, Last) and a 'Show 100 / page' dropdown.
- Variable Detail Panel:** A form for the selected variable 'TTINC27'. It includes fields for 'Name' (TTINC27), 'Label' (Ef - total income), and checkboxes for 'Is Time', 'Is Geographic', and 'Is Weight'. There are also fields for 'Description' and 'Response Unit'.
- Right Panel (Shopping Cart):** A table with columns 'Name', 'Label', and 'Study'. It is currently empty, with a 'Total: 0' at the bottom.

Study	Name	Label
Survey of Financial Security - 2005	TTINC27	Ef - total income
General Social Survey - Social Support - 1996 (Cycle 11)	J41	Estimate of your total personal inc...
Survey of Financial Security - 2005	ATINC27	Family unit - after-tax income
Survey of Financial Security - 2005	FDITX42	Fed income tax
Survey of Financial Security - 2005	G142	Guaranteed income supplement
General Social Survey - Social Support - 1996 (Cycle 11)	J38G	Guaranteed income supplement/s...
Census of Population - 2006	CQPPB	Income : Canada/Québec Pension ..
Census of Population - 2006	EFINC PP	Income : Economic family income..

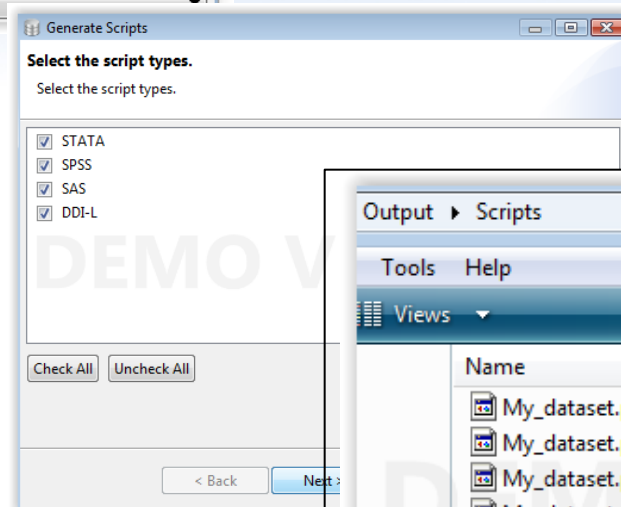
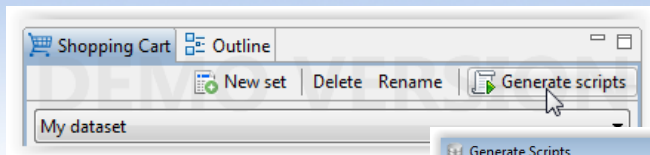
Dataset Builder: Creating New Dataset

The screenshot shows the 'New Dataset' dialog box in the Dataset Builder application. The dialog has a title bar with 'New Dataset' and a close button. Inside, there is a label 'Enter Dataset Name:' followed by a text input field containing 'My dataset'. At the bottom right, there are two buttons: 'OK' and 'Cancel'. A mouse cursor is pointing at the 'OK' button. In the background, a portion of the 'Shopping Cart' window is visible, showing a 'New set' button and a table with columns 'Name', 'Label', and 'Study'.

The screenshot shows two windows from the Dataset Builder application. The 'Variables' window is in the foreground, displaying a search for 'income' with 139 entries. It has a search bar, 'Options' and 'Show' buttons, and a table of results. The 'Shopping Cart' window is in the background, showing a list of variables. The 'Shopping Cart' window has a title bar with 'Shopping Cart' and 'Outline', and buttons for 'New set', 'Delete', 'Rename', and 'Generate scripts'. A dropdown menu shows 'My dataset' selected. Below it is a table of variables.

Name	Label	Study
LOINEFB	Income: Low inco...	Census of Population - 2006
AFTAXI	Income: After-tax i...	Census of Population - 2006
J38G	Guaranteed income...	General Social Survey - Social Supp
J38F	Income from basic ...	General Social Survey - Social Supp
IN2D063x	Total income Cana...	Longitudinal Survey of Immigrants
IN2Q032x	Source income in C...	Longitudinal Survey of Immigrants
WA_Q09	Total income - all h...	Maternity Experiences Survey - 2006
TTINC42	Total income (befo...	Survey of Financial Security - 2005
WNETWPT	Net worth income ...	Survey of Financial Security - 2005
CTID	Combined income ...	Youth in Transition Survey, 15 year

Dataset Builder: Generating Scripts



A screenshot of the 'Output > Scripts' window. The window has a search bar and a 'Views' dropdown menu. Below the menu is a table listing generated files. The table has columns for Name, Date modified, Type, and Size. The files are organized into two groups: SAS System Programs and SPS Files. The SAS System Programs group contains five files named 'My_dataset.physInst_1...' through '5...', all with a date modified of 29/03/2014 5:35 PM and type 'SAS System Progr...'. The SPS Files group contains five files named 'My_dataset.physInst_1...' through '5...', all with a date modified of 29/03/2014 5:35 PM and type 'SPS File'. Below the SPS Files group, there are two more files: 'My_dataset.physInst_1...' with a date modified of 29/03/2014 5:34 PM and type 'DCT File', and 'My_dataset.physInst_1...' with a date modified of 29/03/2014 5:34 PM and type 'DO File'.

Name	Date modified	Type	Size
My_dataset.physInst_1...	29/03/2014 5:35 PM	SAS System Progr...	
My_dataset.physInst_2...	29/03/2014 5:35 PM	SAS System Progr...	
My_dataset.physInst_3...	29/03/2014 5:35 PM	SAS System Progr...	
My_dataset.physInst_4...	29/03/2014 5:35 PM	SAS System Progr...	
My_dataset.physInst_5...	29/03/2014 5:35 PM	SAS System Progr...	
My_dataset.physInst_1...	29/03/2014 5:35 PM	SPS File	
My_dataset.physInst_2...	29/03/2014 5:35 PM	SPS File	
My_dataset.physInst_3...	29/03/2014 5:35 PM	SPS File	
My_dataset.physInst_4...	29/03/2014 5:35 PM	SPS File	
My_dataset.physInst_5...	29/03/2014 5:35 PM	SPS File	
My_dataset.physInst_1...	29/03/2014 5:34 PM	DCT File	
My_dataset.physInst_1...	29/03/2014 5:34 PM	DO File	

Dataset Builder: Generating Codebooks

The screenshot displays the 'Report center' interface. On the left, there are buttons for 'Catalog Explorer' and 'Filter'. The main window is titled 'Report Selection' and contains a tree view with the following items:

- Survey Documentation
 - Codebook (HTML)
 - Record Layout (HTML)
 - Record Layout (PDF)
 - Classifications (HTML)

Below the tree view are buttons for 'Select All' and 'Clear Selection'. At the bottom of the window are navigation buttons: '< Back', 'Next >', and 'Finish'.

To the right, a window titled 'Output > Reports' shows a table of generated files:

Name	Date modified	Type
Classifications.html	29/03/2014 5:38 PM	HTML Document
RecordLayout.pdf	29/03/2014 5:38 PM	Adobe Acrobat D...
Codebook.html	29/03/2014 5:38 PM	HTML Document
RecordLayout.html	29/03/2014 5:38 PM	HTML Document