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
The core metadata is captured using Nestar Publisher which results in two DDI files per survey (English and French), along with two RDF files describing external resources. New fixed ASCII data files are also exported for storage in the data repository, along with external resource files.

The Nestar DDI and RDF files need to be validated to ensure consistency. They are then merged into a single multilingual DDI 2.5 file. English is used for non-translatable elements.

The DDI 2.5 file is upgraded to a DDI 3.1 suitable for ingestion by the Metadata Editor.

Using the Metadata Editor, the coder/manager checks the survey out of the repository and then adds ("imports") the DDI 3.1 describing the files, variables, and resources to the base IMDB study description. The metadata can also be enhanced and updated as needed. Once completed, the survey is checked back in the repository.

Through the Dataset Builder, researchers can access metadata for surveys to which they have been granted access and manage virtual datasets. This generates scripts to produce new data from ASCII using their preferred stat package.

Note that the Data Repository is no longer coupled to the Metadata Repository. Finding and removing the files is the responsibility of the user.
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
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
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 organizations; 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.
Dataset Builder: Variable Basket
Dataset Builder: Creating New Dataset
Dataset Builder: Generating Scripts
Dataset Builder: Generating Codebooks