

DATA LINKS

CENSUS 2001 UPDATE

In the past few months, Statistics Canada has released most of the final Census files for the Profile and Topic-Based Tabulation series. Cumulative Profiles in B2020 format are now available in two series. Canada-wide cumulative profiles include all profile variables in files by geographic level (eg. Census Subdivisions) for all of Canada. Provincial Cumulative Profiles provide all profile variables in separate provincial files by geographic level within each province. The provincial files are much smaller than the Canada-wide files and researchers focusing on a particular province will find them easier to use.

The Census 2001 Analysis Series of reports have also been released based on the 21 topic areas. Fourteen reports have been released to date. View <http://tdr.uoguelph.ca/DATA/CENSUS/2001/DOCS/ANALYSIS/cenb01anal.html> for the reports in PDF format.

Fall 2004 remains the release date for the Public Use Microdata Files (PUMFs). A new list of cancelled and amalgamated products was released on March 4, 2004. See Cancellation Table for details.

CONTENTS

CENSUS 2001 UPDATE.....	1
NESSTAR PILOT PROJECT.....	1
CANADIAN INTERNATIONAL TRADE DATABASE.....	2
HOUSEHOLD INTERNET USE SURVEY.....	3
TDR NEW ACQUISITIONS.....	3
REMOTE DATA ACCESS - Pt. 2.....	3
NEWS	4
CENSUS 2006 TESTING	4

Editor: Carol Perry
Email: carolp@uoguelph.ca

*Data Links is published by TDR every
Fall & Winter to inform researchers of the
latest news on our products and services*

*For previous issues see:
[http://tdr.uoguelph.ca/NEWS/
mainnews.htm#dl](http://tdr.uoguelph.ca/NEWS/mainnews.htm#dl)*

NESSTAR PILOT PROJECT

Tri-University Data Resources (TDR) will embark on a pilot project this summer using NESSTAR software as a new web-retrieval interface. The major difference between the TDR web-retrieval system and the NESSTAR product lies in the advanced search capability available in NESSTAR. Another asset is the enhanced user-friendly interface. Variables can be searched for and accessed in a clear, easy-to-read format which will be more intuitive for users new to electronic data products.

Simple or complex searches are possible using the NESSTAR interface. Searches are available at the dataset level or at the variable level within a particular dataset. Advanced search options allow targeting a search by field options such as author, title, study description or variable name. Standard search operators apply. The results screen brings the user directly to the variable in question and its associated data. Subsetting of data and basic tabulations can be performed directly from the results screen. Results can then be downloaded to a variety of statistical software packages for further manipulation.

There are several components to the NESSTAR suite, including the publishing software, a cube-builder, a web server and a web engine. The publisher is the authoring tool. The cube-builder allows for drill-up and drill-down displaying of cubed data. The web server and web engine combine to provide a web interface for the user to search and retrieve data.

Continued on page 2

Using templates, which can be customized by staff, the publishing software applies xml (Extensible Markup Language) coding in order to make all aspects of a data set searchable. It does this by using DDI (Data Documentation Initiative) specifications in tagging each element of a dataset. The result is a document known as a codebook. A DDI codebook contains all of the information necessary to produce several different types of output, including, for example, a traditional codebook, a bibliographic record, or SAS/SPSS/Stata data definition statements. The codebook can be created from scratch within the publisher, or pre-tagged DDI compliant codebooks can be imported directly into the publisher.

The codebook is published to the web server and web engine, which allows searching and retrieving of data in a similar fashion to our home-grown data retrieval system. Over the past year, TDR staff have created partial codebooks using DDI specifications in preparation for choosing a new web interface. These codebooks can now be imported directly into the NESSTAR publishing software, new elements added and published to the web server.

For more details on the Data Documentation Initiative see <http://www.icpsr.umich.edu/DDI/>.

NESSTAR technology is now employed by such institutions as data archives in the United Kingdom, Denmark, Norway, and Health Canada.

Statistics Canada has just announced an agreement to use NESSTAR products. For more information see: <http://www.nesstamericas.com/news/>

Carol Perry - University of Guelph

CANADIAN INTERNATIONAL TRADE DATABASE

The Canadian International Trade database is collected and published by Statistics Canada International Trade Division. Coverage includes import and export figures dating from 1988. The database provides the value and quantity of each commodity traded. Value is presented in Canadian dollars. The quantity is expressed in the respective units of measure.

Commodities are presented by Standard International Trade Classification Revision 3 at the 2-digit, 4-digit, 6-

digit and 8-digit codes for exports, and 2-digit, 4-digit, 6-digit, 8-digit and 10-digit codes for imports. The files are housed and available on the TDR website.

To facilitate data retrieval, the presentation of the database on the TDR website has been modified. Users now have the option of selecting the level of detail of the traded commodity first. Examples as follows:

EXPORTS		
2-Digit	01	Live Animals
4-Digit	0102	Live bovine animals
6-Digit	010210	Bovine, live pure-bred, breeding
8-Digit	01021010	Bovine, live pure-bred, breeding, dairy
IMPORTS		
2-Digit	01	Live Animals
4-Digit	0102	Live bovine animals
6-Digit	010210	Bovine, live pure-bred, breeding
8-Digit	01021010	Bovine, live pure-bred, breeding, dairy
10-Digit	0120100090	Bovine, live pure-bred, breeding, except dairy

All commodities are available in each web retrieval dataset, therefore, users no longer need to run separate retrievals for commodities from different commodity sections. However, due to the size of the database, the files have now been divided by years as follows: exports: 1988-1997, 1998-2002 and imports: 1988-1992, 1993-1997, and 1998-2002.

Units of measure have also been included in the new presentation. These are only available for 6-digit, 8-digit and 10-digit commodities. There are two variables available for retrieval, the unit of measure, which is an abbreviation and the unit of measure description, which is a full name and description for the unit of measure. To obtain these with your retrieval, please select the Unit of Measure and Unit of Measure description variables from the Choose Variables box.

For more information please contact drc@uoguelph.ca

Michelle Edwards - University of Guelph

HOUSEHOLD INTERNET USE SURVEY

The most recent Household Internet Use Survey (HIUS), conducted in January 2003, has been released through the Data Liberation Initiative (DLI) to member Canadian post-secondary institutions and is now available from the TDR site.

The HIUS, first produced for the 1997 reference year, is administered annually to a sub-sample of Labour Force Survey (LFS) households using telephone interviews. It collects detailed data regarding the Internet activities of Canadian households, specifically how they use the Internet, from where they use it, how often they use it, and why they use (or do not use) it. The 2002 survey also duplicated the electronic commerce module that was introduced in 1999.

Main objectives of the HIUS include:

- To gain an increased understanding of Canadian household Internet use;
- To identify the types of Internet services used at home and to determine the demand for these services;
- To track purchases of goods and services over the Internet from home; and
- To determine the concerns of Canadians regarding security and privacy issues in Internet use.

Analysis of the data resulting from the Survey provides insight into the relationships amongst usage, location of use, and

demographic factors, such as household income. Some of the results of the 2002 Survey indicate that:

- Growth in Internet use amongst Canadians has levelled off;
- Nearly 8.4 million households, or 69%, have at least one member who has used the Internet at some point in their lives;
- There is continued growth in Internet connections by cable from home;
- Use of the Internet correlates with household income; and
- Canadians continue to use the Internet from home largely for e-mail and browsing, although increasingly, people are using it to obtain health and government information, to make travel arrangements, and for specialized uses such as electronic banking.

All annual files for HIUS, from 1997 to 2003, can be accessed via the TDR site under the Communications database group. Documentation for the Survey can be accessed through the readme file for each survey year.

For further information, please contact your university's data representatives listed on page 4 of this newsletter.

Sandra Keys - University of Waterloo

TDR NEW ACQUISITIONS

Canada's Retirement Income Programs

- 1991 - 2001 - Excel files

Canadian Tobacco Use Monitoring Survey - 2003

- Cycle 1

Census of Population - 2001

- Cumulative Profile Series - B2020 files

Kitchener-Waterloo Metropolitan Area Study

- 1998

Postal Code Conversion File

- Census 2001 - June/03 file

Postal Code Federal Riding File

- Census 2001 - June/03 file

REMOTE DATA ACCESS– Using the National Longitudinal Survey of Children and Youth (NLSCY)

Remote data access (RDA) is an alternative to Research Data Centres (RDCs) for accessing Master Survey files. As outlined in the last issue of Data Links, the National Population Health Survey (NPHS), the Canadian Community Health Survey (CCHS), and the National Longitudinal Survey of Children and Youth (NLSCY) are available by RDA, although some have an associated cost.

To access NLSCY Master Files through RDA, the researcher must first submit a two to three paragraph proposal outlining the purpose of his/her research. Depending on the number of requests received by the authoring division, a response can be expected within three to four workdays. Our experience to date has been a maximum response time of three workdays.

Continued on page 4

Once approval has been granted, the researcher must construct a SAS or SPSS program to access and analyze the data. The synthetic files housed on the TDR website should be used to develop and test the SAS or SPSS program. Items that must be incorporated in the program are Master data filenames and the directory used to access the master files. Master data filenames are supplied with the synthetic data. For example, the NLSCY – Cycle 1 – Primary Master Data filename would be NLSCY_94_C1_PR_MAS.TXT, as listed in the Synthetic data directory. The directory structure is currently unknown while the authoring division is implementing directory changes. The directories will be added to the researcher's program by the NLSCY project team until the new directories are in place, at which time the researcher will be responsible for including these in their program. Once the changes are complete at the NLSCY division, we will be notified and the directory structure will be archived on TDR for future researcher use.

With data analysis, a researcher may need to create temporary datasets. The NLSCY division has made provisions for this to occur. Temporary datasets must be appropriately named and placed: as illustrated in the following examples: K:_Researchers & Dissemination_Tabulations and Data Requests\RDA Runs\Michelle Edwards\test1.

Before submitting a program to the NLSCY project team, please ensure that it runs without errors against the synthetic data files. The NLSCY project team will NOT alter your program to fix any programming errors. Also be aware that the output generated by your program will be vetted by the team to ensure the confidentiality of individual responses. Any output or table where confidentiality is not ensured will be omitted. An example of this would be a table where the number of observations is less than or equal to five.

For further information on accessing the synthetic files or preparing for an RDA submission, please contact your local data service providers.

Michelle Edwards - University of Guelph

NEWS

The first week of April brings with it the annual Data Liberation Initiative (DLI) training week sponsored by Statistics Canada. These workshops give university data providers an opportunity to gain hands on experience with new Statistics Canada data products. It also provides the partners in this initiative a forum to discuss ways of making data more accessible to students and researchers in the university community.

CENSUS 2006 TESTING – Summer Job Opportunities

Planning for the 2006 Census is now underway. Public consultations have been completed and now teams are being assembled to provide sample testing in preparation for the coming census day of May 16, 2006.

Job opportunities are available beginning in April for individuals to assist in the testing process. Successful applicants will join teams to test the questionnaires, as well as, distribution and processing of the census. See <http://www12.statcan.ca/english/census06/jobs/jobs2004.cfm> for details about the opportunities in your area.

Regions covered by this testing include the Atlantic region, Quebec and the Prairies.

YOUR CONTACTS FOR DATA SERVICE AT OUR THREE LOCATIONS

UNIVERSITY OF GUELPH
Michelle Edwards
edwardsm@uoguelph.ca
519-824-4120 x54539

UNIVERSITY OF WATERLOO
Sandra Keys
skeys@uwaterloo.ca
519-888-4567 x2279

Shabiran Rahman
srahman@library.uwaterloo.ca
519-888-4567 x2882

WILFRID LAURIER UNIVERSITY
Helene LeBlanc
hleblanc@wlu.ca
519-884-0710 x3743