



Data Links

TriUniversity Data Resources

<http://tdr.tug-libraries.on.ca/>

Fall 2007

Vol 10 Issue 2

Contents:

ODESI – A Journey in Data Discovery.....	1
Census 2006.....	1
News.....	2
New CANSIM II interface through University of Toronto's CHASS.....	2
TDR New Acquisitions.....	3
IASSIST Conference Notes.....	3
National DLI Training Day.....	4
Announcement.....	4

Your Contacts For Data Service At Our Three Locations:

University of Guelph

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

University of Waterloo

LibEDS@library.uwaterloo.ca

Wilfrid Laurier University

Hélène LeBlanc
hleblanc@wlu.ca
519 884-0710 x3743

Editor: Carol Perry
carolp@uoguelph.ca

Design/Layout: Lisbeth Sider

Data Links is published by TDR every Fall and Winter to inform researchers of our products and services.

For previous issues see:
<http://tdr.tug-libraries.on.ca/NEWS/mainnews.htm#dl>

ODESI – A Journey in Data Discovery

by Michelle Edwards, University of Guelph

The Ontario Data Documentation, Extraction Service and Infrastructure Initiative (ODESI), a jointly funded project between the Ontario Council of University Libraries (OCUL) and OntarioBuys, will provide university researchers with unprecedented access to a significant number of datasets in a web-based data extraction system which will be delivered through the highly successful Scholars Portal model.

The project will target Statistics Canada datasets, datafiles from Gallup Canada and other polling companies, public-domain files such as the Canadian National Election Surveys and selected files from the Inter-University Consortium for Political and Social Research (ICPSR). The files will be marked-up using DDI, an international, XML-based metadata tagging system which allows data resource discovery, distributed access, extraction and analysis.

This project will expose undergraduates to the research enterprise at an early stage in their careers and will help in developing their numeracy skills. Researchers working with opinion polls will be able to search across hundreds of datasets and collections to see trends and historical patterns over time.

The University of Guelph and Carleton University are the lead institutions involved with the ODESI project. This summer a Best Practices Document outlining the DDI tagset recommended for use, along with detailed explanations and several examples was created with the help of Queen's University. The document is available upon request and can be viewed on the ODESI wiki at

<http://odesi.uoguelph.ca/wiki>. Comments and suggestions are welcome.

Co-op students have been hired at both institutions to work on the project. At Guelph, the students are retrofitting and updating datafiles currently residing on the TDR-Nesstar website. Files will conform to the Best Practices Document and will include question text, question universe and any interviewer instructions. Throughout the process, updated files will be made available on the TDR-Nesstar service and will be accessible through the TRELIS system. Files located in the Health directory are the first to be completed and are now available on the TDR-Nesstar service.

At Carleton, Co-op students are creating metadata for the Gallup poll data dating back to 1945. These rich and interesting files will then be made available to all OCUL institutions through this project increasing the data resources currently available to us.

On October 1, 2007, Paula Hurtubise, formerly of Statistics Canada, began her tenure as Project Manager for the ODESI project. "Paula comes from the Geography Division of Statistics Canada, where she was the Chief, Client Liaison and Dissemination. Paula has an expertise in dealing with clients and the research community and a deep understanding that will directly support the goals of the ODESI project." (Press Release October 5, 2007)

For more information on the project please visit the ODESI wiki at <http://odesi.uoguelph.ca/wiki>.

Census 2006

by Michelle Edwards, University of Guelph

Another Census under our belts and new data seems to be arriving at a regular pace. Profiles of Age and Sex and Marital Status, Common-law Status, Families, Dwellings and Households are already available on the TDR-Nesstar site with Age and Sex, Family and Households, Housing

and Shelter Costs, and Marital Status Topic-Based Tabulations (TBTs) available as well.

Whenever a Census is released we often grapple with the restricted access question. What files are available for public use? Which ones are only

continued >

News:

Staffing Changes at U of G

It has been a busy year at Guelph with numerous staff changes in the Data Resource Centre (DRC). Earlier this year, Jack Cooper and Steven Gu, analysts with the CCS (Computing and Communications Services) component of the DRC left the university to pursue other interests. We wish them well in their new endeavours. More recently, Carol Perry was appointed academic liaison librarian for Political Science and Government Publications, moving her out of the day-to-day operations of the Centre. Jenny Marvin, GIS librarian, has just returned to us from family leave. Welcome back Jenny. Michelle Edwards has been seconded ½ time for one year to the position of Developer with the ODESI project (see article on page 1). She has also been elected the 2007 President of CAPDU, the Canadian Association of Public Data Users.

Data Day 2008

Mark your calendars for January 2008. The Southwestern Ontario Research Data Centre (SWORDC) and its partner institutions will once again be sponsoring two days of presentations on the research use of data from Statistics Canada master files. Wilfrid Laurier University and the University of Guelph will be hosting the series. Dates will be announced in the coming weeks.

Census 2006 (Continued)

by Michelle Edwards, University of Guelph

available through the Data Liberation Initiative (DLI)? Why does it appear that all the files are available on the Statistics Canada website? What geospatial data are available and where?

Statistics Canada has developed an access policy that helps us determine the level of access to products. Restrictions are imposed at specific levels of geography, therefore, the smaller the geographic area, the less information is publicly available. Standard products for Census 2006 include: Population and Dwelling Counts, Highlight tables, Analysis Series, Topic-based Tabulations (TBT) and Profiles.

Population and Dwelling Counts are available to the general public down to the Census Division/ Census Subdivision (CD/CSD), Urban Area (UA) and Designated Place (DP) level. Our users can access population and dwelling counts to a lower level of geography, for instance, Dissemination Area (DA) and/or Forward Sortation Area (FSA), through the GeoSuite2006 program on compact disk. This product is not freely available to the general public.

Highlight Tables, Analysis Series, TBTs and Profiles are freely available on the Statistics Canada website to the public at a Province/Territory level, Federal Electoral District (FEDs -2003

Representation Order) and Census Metropolitan Area or Census Agglomeration (CMA/CA) level. In some cases, selected Census Division (counties) or Census Subdivisions (municipalities), may also be available for the TBTs. All smaller geographic levels are only available to key stakeholders and partners. As a member of the DLI, we have access to the smaller geographic detail through Statistics Canada's website and through our local data repository, the TDR-Nesstar service.

Geographic files follow a similar level of access. All Reference and Thematic Maps are freely available on the Statistics Canada website in PDF format. Access to digital files is restricted by the level of geography. Canada, Provinces/Territories, CD, CSD, Economic Regions (ER), CMA/CA and Census Consolidated Subdivision (CCS) levels of geography are available at no cost to the general public on the Statistics Canada website. Lower levels of geography are available to the public for a fee and are available to us through the DLI.

Enjoy the new Census data. If you have any access questions or questions regarding how the Census data has been mounted onto the TDR-Nesstar website, please contact: drc@listserv.uoguelph.ca.

New CANSIM II interface through University of Toronto's CHASS...

by Hélène LeBlanc, Wilfrid Laurier University

For free access to Statistics Canada's time series database, CANSIM, the universities of Waterloo, Guelph and Laurier all subscribe to the University of Toronto's CHASS (Computing in the Humanities & Social Sciences) online interface to the data, found at: <http://dc2.chass.utoronto.ca/cansim2/English/index.html>. Historically, though, the look and feel of the CHASS interface, especially the design and function of its search page, has been significantly different from Statistics Canada's own online (fee-for-data) rendition which can be seen at: <http://cansim2.statcan.ca/>

Consequently, users familiar with the CHASS CANSIM II interface will be very happy to learn

that as of February, 2007, CHASS has created a new way to interact with CANSIM II, their **CANSIM Multidimensional @ CHASS**, built to look and work similarly to the Statistics Canada's web site version.

Perhaps the most striking difference between the two CHASS CANSIM II versions is that you can now create tables containing multiple series. Searching for needed series to populate these multidimensional tables can be done in a variety of ways. The traditional **text search** is still available, but is greatly enhanced, allowing, among other things, phrase searching and the stipulation of the frequency of the needed time

continued >

TDR New Acquisitions

Canadian Internet Use Survey 2005

Canadian Tobacco Use Monitoring Survey, 2006
– *Annual & Cycle 2 files*

Census of Canada 2006
– *Assorted Profile and Topic-based Tabulation Files – B2020 files*

Census of Agriculture 2006
– *Initial release*
– *B2020 files*

Elementary-Secondary Education Statistics Project (ESESP) 2004-2005
– *Excel files*

Financial Information of Universities and Colleges Survey (FIUC) - *Excel files*

GeoSuite 2006

Provincial Economic Accounts 1961-2006
– *Excel files*

Survey of Financial Security - 2005

Survey of Household Spending - 2005
– *Excel files*

Survey of Earned Doctorates 2003-2005
– *Excel files*

New CANSIM II interface through University of Toronto's CHASS... (Continued)

by H el ene LeBlanc, Wilfrid Laurier University

series, such as monthly, quarterly or annually. (Users of Stats Canada's E-Stat to access their CANSIM data will see that one of the drawbacks of CHASS's CANSIM II products still remains that you cannot manipulate your results, such as creating semi-annual time series from monthly data points.)

New ways to search for your data include **browsing tables by keywords**, which is a page displaying predefined keywords; the larger the font used for the keyword, the more CANSIM

tables are associated with that keyword. If you **browse tables by surveys**, you can locate all CANSIM tables associated with a particular Statistics Canada survey.

You can also search for a variety of series to **build a multidimensional table**, by selecting individual series from pre-arranged listings called "dimensions". Selected series are entered into the "series cart" from which they can then be chosen to create desired tables.

IASSIST Conference Notes

by Sandra Keys, University of Waterloo

This year's International Association for Social Science Information Service and Technology (IASSIST) Conference, with the theme of Building Global Knowledge Communities with Open Data, was held at McGill University, Montreal, from May 15-18. As always, it brought together a wide variety of people, including data providers, social scientists, and information professionals from all over the world and notably, a good-sized contingent from Africa! Workshops, sessions, roundtables and working groups, and social events provided ample time to network and learn about data initiatives and experiences worldwide.

Plenary sessions dealt with the topics of using data from national surveys, building health information services, and statistical literacy initiatives at Statistics Finland and Statistics Canada. Numerous concurrent sessions were available on a myriad of topics to satisfy both novice and savvy data professionals. Topics included issues with open data, archiving, preserving, and managing data; data access; national data archives; the DDI (Data

Documentation Initiative); and statistical literacy. Many of the presentations, including webcasts of some of the sessions, are publicly available at the conference website: <http://www.edrs.mcgill.ca/IASSIST2007/program.htm>. It is always interesting to note that many of the issues, which we may think are only our own, are actually global in nature and we have much to learn from each other so that we can improve data collection, access, dissemination, and preservation.

From my perspective as a librarian, the sessions focusing on statistical literacy and data access were of primary interest and thus, it was those which I generally attended. Nevertheless, it is always good to receive a broader knowledge of the data world by attending sessions on topics to which I would otherwise not be privy and attending conferences such as IASSIST allows me to do just that. I encourage others to attend IASSIST next year, May 27-30, at Stanford University, where the theme will be *Technology of Data: Collection, Communication, Access and Preservation*.

National DLI Training Day

by Sandra Keys, University of Waterloo, and H el ene LeBlanc, Wilfrid Laurier University

The National Data Liberation Initiative (DLI) Training Day, held every four years when the International Association for Social Science Information Service and Technology (IASSIST) conference is in Canada, is a wonderful opportunity for meeting with our DLI colleagues across the country to share experiences and learn from each other, as well as to hear about the activities of the DLI. This year's National Training Day featured a variety of presentations from both Statistics Canada and our colleagues at other post-secondary institutions which both informed and challenged.

The day began with a Welcome Address and Update from Michel S equin, Chief of the Data Liberation Initiative, and his staff. The main session of the morning was a presentation by Vicki Crompton, Director General of the Communications and Information Services Branch of Statistics Canada, and Mary Townsend of the Education Outreach program entitled "Moving from Data Liberation to Data Literacy." They emphasized the importance of "using correct information and using it correctly" and focused on two initiatives, The Media Room (http://www42.statcan.ca/smr_r000_e.htm) and the Education Outreach Program's release of the Teacher's Guide

to Data Discovery (<http://www.statcan.ca/english/freepub/12-593-XIE/12-593-XIE2007001.htm>).

This guide supports elementary and secondary teachers in helping students develop basic statistical skills, such as creating tables and graphs and calculating measures of central tendency. This session was followed by several product reviews focusing on geography, small area data, and health.

The afternoon consisted of two workshops, one highlighting a new DLI representative toolkit to assist reps in making the best and most efficient use of the DLI program and the other expanding on the area of education statistics. In this latter session, a representative from Statistics Canada's Culture, Tourism and the Centre for Education Statistics spoke about the Centre and the surveys and products overseen by them. The Centre focuses on statistics from primary and secondary schools, postsecondary and adult training, household surveys, and the assessment of gained competency levels. Financial data on higher education, such as tuition and cost of living for students, is available at a fairly good level of detail, with some statistics available for free, others available through custom tabulations.

Announcement

Final issue of Data Links

This issue will mark the final production of the Data Links newsletter. We would like to thank all the contributors to our newsletter over the past 10 years. We would also like to extend our thanks to our readers for their interest and support. As editor it has been my pleasure to have been part of this endeavour.

Please monitor the news section of the TDR website or your local data centre websites for ongoing updates and news on advances in data dissemination through the TriUniversity Data Resources. Farewell to our readers.

Carol Perry, Editor, University of Guelph

TDR-Operational Management Group

Doug Horne, University of Guelph – Chair
Michelle Edwards, University of Guelph
H el ene LeBlanc, Wilfrid Laurier University
Richard Pinnell, University of Waterloo