

Data Purchase and Access Working Group

Working Group meeting – September 26, 2011 – 1:30pm EDT

In attendance:

Louisa Wong (Hamilton), Ian Faris (CRD), Ken Murdoch (Winnipeg), Isabelle Lépine (Montréal), Irena Pozgaj (Simcoe), Karin Kronstal (ACR), Michael Ditor (ACR)

ACTION ITEMS:

- Michael Ditor Provide a comprehensive list of potential data sets, with prices and other relevant information – see the file ‘Community Data Catalogue 2012-2017.xls’ (click the links in the ‘cost request summary’ tab to view product details).
- Karin Kronstal Solicit feedback from the Working Group regarding the upcoming data purchase/consortium engagement survey.
- Ian Faris Re-send the Custom Geographies list and solicit Leads to complete their list
- Louisa Wong Compile and distribute TransUnion demonstration files, sectioned by geographic region.
- WG members Review the variables being considered for the next data purchase. Of particular interest are the General Social Survey and Households and the Environment Survey to see if the lists of variables can be pared down. These variables can be viewed in the file ‘Community Data Catalogue 2012-2017.xls’.

Added to the agenda:

Field comments regarding the 2008-2012 agreement amendment that was sent out to Leads the previous week.

1. Review of potential data sets

Questions about strategy for next data purchase:

Will members have to purchase a total package or be able to choose products of interest? ...or some variation (e.g., UPP vs. total package)?

Are the QOLRS data included in the data package? **YES, THE QOLRS DATA SETS ARE BEING PRICED.**

Regarding the issue of 'a la carte' data offers: logistics are difficult, but if it's not done, then some consortia pay for a lot of data they can't use. However, this is already the case for smaller geographies, but the benefits seem worth it (Irena).

ACTION ITEM: Provide a comprehensive list of potential data sets, with prices and other relevant information.

Comment on need for the data costs: It's hard to have conversations with local members about the next round without a price.

Request regarding the LFS data: Part-time work, reason for part-time work (Louisa)

2. State of current requests for quotes

The file 'Community Data Catalogue 2012-2017.xls' contains a list of data sources that are being pursued. It should be noted that almost ALL cost estimates come with the qualifier that the future prices have not yet been set. It is very likely that many prices will change, however it is unknown whether this change will come about before the next agreement is drawn up. Estimates based on cost recovery of time spent by STC analysts will likely not change greatly.

3. Data purchase survey/data wish lists from consortia

Karin Kronstal to work on data purchase survey. The survey will be fairly basic and consist of two main requests :

- (1) Are you interested in these new data sets?
- (2) Given the current list of consortia, are you aware of any potential consortia members? Please provide contact information.

The Consortium is looking to expand to ~35 consortia. It needs to be indicated either in the survey or in future communications what the advantage is of increased membership?

Need to indicate what is meant by a consortium, as there are some members whose Consortium wish to represent the entire province.

Karin will solicit feedback from the DP&A WG regarding this survey.

4. Trial of new data products

Discussions have been ongoing with several data suppliers about receiving a demonstration version of their product. These products include the following:

Product: Permanent Residents Rounded Data Cube Facts and Figures
Source: Citizenship and Immigration Canada
Geography: Canada, Province, Economic Region, Census Division, Census Metropolitan Area, Census Subdivision, Destination
Time Series: Annual, 2000 to 2009
Description: Information on immigrants landed from 2000 to 2009 for all of Canada, including several socio-demographic variables, including age, country of birth, education level, marital status, skill level, years of schooling, among others.

Product: Credit Report Characteristic Data
Source: TransUnion Canada Ltd.
Geography: 6-digit postal code level for all of Canada
Time Series: Q4-2011 and Q1-2008
Description: TransUnion Risk Score, TransUnion Bankruptcy Score, Total Balance of Loans

Product: Envision software and various data products
Source: Environics Analytics
Geography: Canada, Province, Economic Region, Census Division, Census Metropolitan Area, Census Subdivision, Dissemination Area
Description: Data reporting and mapping tool that integrates multiple data sets into user-friendly formats

Product: TrendPoint
Source: Centract Limited
Geography: Canada, Province, Census Metropolitan Area, Census Division, Census Subdivision, Census Tract, Custom Geographies
Time Series: Quarterly Since 2004
Description: Real estate sales and appraisal data and property characteristics aggregated from across Canada.

TransUnion data files:

These files have been made available to the Working Group members via the website. Louisa Wong has compiled the data and can section the files by geographic region.

The other demonstration products will be made available once we receive them.

Environics Analytics:

How do these data compare with the demographic updates that Statistics Canada provides?

Need to clarify what to do with members who are already with Environics Analytics.

What data are used by Environics Analytics to run their data? Most likely they are rounded data, and not the original data. Leery about getting data second-hand (Ian).

Citizenship and Immigration Canada Rounded Data Cubes:

Technical discussions regarding how to make the CIC Data Cube available via the Consortium website are to take place at the beginning of October.

5. Review data sets still to be delivered before end of current contract

A list of data sets that are still to be delivered from the last agreement can be found on the following page of the Consortium web site:

<http://communitydata-donneescommunautaires.ca/node/7541>

6. Custom geographies

Ian has sent out the list of Custom Geographies to the Leads again, requesting that the gaps in the file be filled and returned to him over the next two weeks.

7. Purchase of June/Dec 2011 CBP and/or the 2011 Census Cartographic Boundary Files before the end of the current contract

Two data gaps in the last agreement have been identified. The following data files are not included in the last contract, but will be released before the next agreement is in place:

- June and December 2011 Canadian Business Patterns – released September 2011.
- 2011 Census Cartographic Boundary Files – to be released November 29, 2011.

Proposal (ACR): Survey the willingness of the Leads at the next meeting (October 20, 2011) to purchase these files before the next agreement is in place and add the cost of the files to the requested \$5000 amendment. Statistics Canada is unable to provide a cost estimate for the Boundary Files at this time, however has indicated a price of **\$5820** for the Canadian Business Patterns data down to the Census Tract level:

June 2011 Establishment Counts for Canada, Provinces and territories, Census Metropolitan Area/Census Agglomeration (CMA/CA), Census Divisions (CD) and Census Subdivisions (CSD) by 2, 3, 4 and 6-digit NAICS and employment size ranges *comes to:* **\$1,500 (\$2,500 – 40% disc.)**

December 2011 Establishment Counts for Canada, Provinces and territories, Census Metropolitan Area/Census Agglomeration (CMA/CA), Census Divisions (CD) and Census Subdivisions (CSD) by 2, 3, 4 and 6-digit NAICS and employment size ranges *comes to:* **\$1,500 (\$2,500 – 40% disc.)**

June 2011 Establishment Counts for all Census Tracts in Canada/Select Census Tracts in list below, by 6-digit NAICS and employment size ranges **\$1,410 (\$2,350 – 40% disc.)**

December 2011 Establishment Counts for all Census Tracts in Canada/Select Census Tracts in list below, by 6-digit NAICS and employment size ranges **\$1,410 (\$2,350 – 40% disc.)**

In the last agreement, licensing fees for these data were about 3.5 times the data fees. However, Statistics Canada is currently reviewing its licensing fees and can not provide a cost estimate this part of the purchase. Assuming no change to the licensing fees, the cost for purchasing the June and December 2011 CBP data would be \$26,190 – split between 20 consortia ≈ \$1310.

8. Other business

What are the expectations for the Working Group?

IAN FARIS: working on custom geographies

Other members of the Working Group will review some of the potential data sets with the goal of assessing whether all of the selected variables are of interest and if any of these can be eliminated in order to bring down the cost. The data sets were this would be most useful are:

- General Social Survey
- Households and the Environment Survey

The variables to be reviewed can be found in the file 'Community Data Catalogue 2012-2017.xls'. Links to the appropriate survey can be found in the 'cost request summary' tab.

Karin Kronstal (ACR) to solicit feedback from the Working Group regarding the upcoming survey.

9. Next meeting

November 3rd @ 1:30pm