

Data Purchase and Access Working Group

August 22, 2017

Adobe Connect: <https://cdp.adobeconnect.com/theboardroom/>

Teleconference: 1.888.271.3643

Attendance

- Amanda Richards (Peel)
- Auburn Larose (WDG)
- Christina Maes Nino
- Cheryl Hitchen (Kingston)
- Heath (Toronto)
- Isabelle Lepine (Montreal)
- Jasmine Ing (Calgary)
- Kelvin Ndoro (Nova Scotia)
- Natalie Hui (York)
- Evan Nemeth (Nova Scotia)
- Michael Ditor (CCSD-CDP)
- Julie Lam (CCSD-CDP)

Agenda

- Recent data release
- Target group profiles
- Custom geographies
- Equifax data

Action items

- For each census release, make note of what variables are not present. Look at the profiles with each release and see what was repeated in 2016 and what was left out. Flag any gap items and let the group know for priorities. **-CDP team**
- Clarify difference between 2011 and 2016 profiles with Statistics Canada representative. **-CDP team**
- Review the age and sex tables at custom geos when it is released and identify low population custom geographies. **-CDP team**
- Look at smaller custom geos and determine which ones are lower priority – will have the leads flag high and low custom geographies. **-Leads**
- Set a meeting in 5 weeks, near end of Sept and continue with 2:30 pm ET. **-CDP team**

Recent data release

The team will try to finalize the priority document with requested changes and get the orders done before Sept 13. The team will need definitive purchase lists – such as the top 30-35 tables and order what's available.

There are some requests for the census, family structure, and tables on age of children, but priority for these tables are not particularly high. It may be best to get the top 35 and the rest as they are released, the team should not order medium priority tables yet. On Nov 29 when everything is released, the team can start to get pricing and requests for other orders. The team wants to ensure no high priority tables are missed because of medium priority tables.

Heath: Do we have any information on non-released data, particularly the activity limitations tables?

Activity limitations are released on Oct 25th, we can place order for those tables. That's all information we have regarding activity limitations.

Heath: Is that now considered high priority? Eg. activity limitation and place of work

We do place of work profiles, we will look at that Nov 29. There is also a place of work dimension.

Jasmine Ing - City of Calgary: My question is about if we find that items we were wanting turn out to not be in the census profile. As an example, we were planning on getting our information about seniors living alone for custom from the census profile, but this information was not included in the census profile that has been released about households and living arrangements.

Jasmine Ing - City of Calgary: so would we need to submit an order?

Jasmine Ing - City of Calgary: We hadn't taken that into account when we prioritized files

Jasmine Ing: don't want to order things separately. Need a process for filling in the gap.

Heath (Toronto): yes, we'll need to look at what else gets dropped from the general profiles.

Heath (Toronto): and other stuff e.g. age by single year is already available down to DA level and we only need for custom geos

This is a good flag to raise, we need to know what's in the profile.

Jasmine: sometimes these items are released in standard as additional. Some of the info is included, but we do order at custom? So information

Jasmine Ing - City of Calgary: Is it possible for the CDP team to review the differences between the variables in the 2011 and 2016 profiles

Jasmine Ing - City of Calgary: and then to check if those items are covered by another table that we would be ordering?

Jasmine Ing - City of Calgary: and if so, if that table will be ordered at custom geos?

Heath (Toronto): I suspect this may be too much work for just the CDP team. Maybe a small sub-working group to investigate and advise the broader group?

Jasmine Ing - City of Calgary: Potentially, Heath. I think the thing I am trying to avoid is multiple people reviewing the same table for the same information

Jasmine Ing - City of Calgary: and then also having to go through the CDP catalogue and try to do the matching process. It would not be efficient for multiple people to do that same work

Heath (Toronto): yes, agreed., we def. need to coordinate

There are a lot of gap items without rank because of downloads. The CDP team and this working group will need to identify priority items among the gap items.

Action: For each census release, make note of what variables are not present. Look at the profiles with each release and see what was repeated in 2016 and what was left out. Flag any gap items and let the group know for priorities.

This meeting can be long if we discuss individual needs. We welcome anyone who joins the meeting and listens in, and appreciate those who do work on the files and give feedback. Please keep email discussions going. The CDP team will take on the task of making sure there are no redundancies and catching gap items. We will enlist the help of this group if we find the task overwhelming.

Jasmine Ing - City of Calgary: As well, we face the issue of the TGPs. We will have assumed these items would be in the TGPs, and if they are not, we may need supplemental tables to fill the gaps.

Yes, this is the same issue in profiles and target group profiles. No definitive answers yet but we understand this can raise some potential problems. This will be talked about and get back to the working group when we get more information from StatCan.

Jasmine Ing - City of Calgary: just flagging that the census profile that is the base for the TGPs

Heath (Toronto): it would be good to know if this will still be the case... do we know what Stat Can plans to put in their Aboriginal profiles? Will this still be built on the census profile?

Heath (Toronto): The Aboriginal profile is the only "target group" profile that StatCan does as standard.

Action: The TGP should be same as regular as the census. Clarify whether there is a difference between target group profiles and census profiles with Statistics Canada representative.

Target group profiles

We have talked about making the profiles into one large table and have the TGP variables be in a separate dimension. We will have a discussion with StatCan next week. They will likely say one significant problem is the size of the tables if we want other variables thrown in. Beyond 20/20 has file size limits.

We may want to turn the TGP into a robust custom table that has a lot of the profile variables and other dimensions that we can cut this profile dimension with.

Cheryl: For our purpose there is lots of data in the TGP that we don't use... for example all of the language data which may be relevant for some profiles and not others.

To some members this is important for them. However, we may have to cut out citizenship. That's an example where we can't include all the variables because of the sheer size of data. We either need to change something or order a bunch of different TGPs. We will have more information after talking with StatCan and hopefully will help the working group come up with requests.

Another issue is the definition of target groups that would allow families and households data to be included in the TGPs. We will need to split out individual profile variables from household profile variables. We did this when we did CPP tables.

Jasmine Ing - City of Calgary: For what it's worth, I think the standard TGP format is very useful because it provides a standardized format and set of variables for the user. For medium and power users, it lets us know very quickly what information is available without needing to look through each table.

Jasmine Ing - City of Calgary: Our suggestion was to have a household level TGP and an individual level TGP as separate tables

Heath (Toronto): it would be helpful to know the cost of that, because that sounds expensive!

Auburn Larose: I would second Jasmine's suggestion

Natalie Hui (York Region): Agree with Jasmine's suggestion

This conversation on TGP does sound expensive. They rank high on their own. This is clearly a priority. We can proceed to find out how much it costs. It may turn out to be something that does become a big-ticket item, but may be worthwhile in the end. We will give StatCan our ideal situation and have them give us the breakdown.

Custom geographies

Some of these tables... don't want to supplant high priority tables.

Kelvin: What's the limit of tables that can be acquired at custom geographies?

Don't have an answer for that right now. We set aside a certain amount of the budget for custom geos, a quarter of the budget. It is currently within the budget. We haven't gotten any prices back for geocoding, so there are still a lot of unknowns. On Nov 29, we would go ahead and order the main profile and custom geos, some of the top TGP at custom geographies.

Evan: the custom tables, custom geo, is that outside of typical sub division?

Yes, members can submit custom geo that we submit to StatCan, they geocode the geographies (you give is shape file or concordance file) and then we order a subset of tables at those custom geographies.

Evan: in my case, I requested through Kelvin. So they would be groups of census subdivision? The data we are looking for, we have province wide data but not regional level. Including the groups of census subdivisions.

Yes, that is right, whatever custom geo you submit.

Evan: submitting suggestions by email?

We welcome any suggestions, please send an email!

Heath (Toronto): is there any thought to segmenting purchasing by the size of the custom geographies?

Heath (Toronto): some deep crosstab tables may not be worth acquiring for small custom geos

Natalie Hui (York Region): Heath's suggestion is worth considering

To elaborate, smaller custom geo may suffer a lot of suppression if the table has a lot of cross tabs. We can ask StatCan to review some of the geographies. It is by the population size, so we don't have that information yet.

Kelvin: size might need to be relative to the province

Heath (Toronto): sure, but if we buy single year of age data for custom geos

The first table that we can order, age by single year and sex by custom geos, has been released already. The production system for tables at custom geo will be ready soon.

Heath (Toronto): that's already available, yes? you can already get DA level data for free from StatCan

It's the one custom table that has high priority and is already available. If the production system is up and running, then we can proceed with batch 1.

Heath (Toronto): no, I'm thinking more of Custom and CPP tables.

Jasmine Ing - City of Calgary: Another idea might be higher and lower priority custom geos. For example, our communities are very high priority for us, but some of our other customs are useful in many cases but not all.

There is a concern that if we cut out small population custom geos from certain deep cross-tab tables, that important data from simple cross-tabs out of those same tables would not be available at custom geographies.

Heath (Toronto): sure, but we could buy less complicated tables at those geos.

Action: Review the age and sex tables at custom geos when it is released and identify low population / low priority custom geographies.

We are having all custom geos geocoded for taxfiler tables. That will run into some suppression, overlap and confidentiality issues. With multiple custom geos that cover the same area, it may not be possible to have all custom geos geocoded, as someone could potentially get around suppression rules by comparing two custom geos that cover the same area. We will have conversations about this with some people, ie which custom geos take priority.

Jasmine Ing - City of Calgary: some of our important custom geos (e.g. Communities) have polygons with large and small populations. We are willing to accept that not all polygons get information when doing a deep crosstab so that we can get information for most of the communities. So just because a geo family has some small population polygons doesn't mean it's not a priority for us. (Some of the communities have a population of only a few hundred while others have a population of 20,000+. We don't want to lose the information for the majority of communities that are large because a few are small)

We may need to have that conversation here – or we can start by asking some of the leads who submitted custom what their priority is. This question will come up again.

Heath (Toronto): makes sense - but let's do the analysis first. we may find there are no real savings to be found. I just think it's worth considering.

Action: Look at smaller custom geos and determine which ones are lower priority – will have the leads flag high and low custom geographies

Kelvin: I agree with Leads submitting priority geos

Equifax data

The methodology for calculating NMCD needs to be confirmed. The original calculation provided by the Equifax client manager was incorrect. We apparently have been given the correct calculation now but the total debt for all Canadians still doesn't agree with Equifax press releases.

It should also be noted that the data we have can only calculate average debt owed by any Canadians with a credit file, that may or may not have debt. Equifax press releases provide the average across all Canadians who have debt. The total debt should still be the same, however the numbers do not agree. We will continue discussions with Equifax on this issue.

Heath (Toronto): good luck with the negotiations, mike! :)

Other

Comments on how these working group meetings are held – sometimes issues with Adobe Connect or the bridge – please let us know. If you have suggestions on how the meetings can be run better, it is very welcome!

Is it worth having these meetings with increase frequency? Will meet again near end of Sept or beginning of Oct. Will update by email as well.

Jasmine Ing - City of Calgary: that's fine with me if there are things to discuss.

Heath (Toronto): yes, I think it's a good idea. would like to hear feedback from StatCan from when you meet them.

Action: Set a meeting in 5 weeks, near end of Sept and continue with 2:30 pm ET.

Meeting finished at 2:40pm.

Extra

Kelvin: Hi Mike. Is it possible to resume ordering Labour Force Survey data at CSD level? That'll be great as it will be really helpful for folks outside HRM. we only have 1 CMA and 4 CAs in NS

Michael Ditor: I'm guessing probably not, I'll ask again. did we go over why we couldn't in the first place? It was hard to obtain LFS data at CSD geos when they were 'available'. The subject matter division was not that willing to release the data, but we had an order 'grandfathered in' from another project. after the NHS debacle, there was no gold standard with which to calculate rates/counts at CSD level with reasonable levels of confidence, so all CSD orders were cancelled.

Kelvin: that's understandable.

Michael Ditor: perhaps that will change now that the mandatory long form is back, I'll ask again

Kelvin: is it possible to maybe get LFS data at custom geos similar in size to CMAs?

Michael Ditor: maybe - that's something we haven't looked into. definitely worth asking about

Kelvin: ok thanks. Just keep us posted

Michael Ditor: will do!