

Foundation Working Group Meeting #3

FWG Members Experiences with De-identifying
Information for Disclosure to Third Parties

June 17, 2015

Halton Hills Hydro

- Halton Hills Hydro has not provided energy consumption at the individual customer or location level.
- We regularly provide consumption aggregated by billing class to meet our various reporting requirements with the OEB and IESO but not at a more detailed level.
- For our CDM Framework, we did conduct a load study of commercial/industrial customers by NAICS code but again, that information was aggregated by NAICS code. Internally, we separated the data into two geographical areas, but no specific customer or location information was identified.

Hydro Ottawa Experience

Hydro Ottawa used a technique to de-identify accounts that participated in the TOU study completed by the OPA in 2013.

- ***What de-identification technique(s) was(were) used?*** Hydro Ottawa created a 10 character long “Study ID” for each account that was part of the TOU study. This was a unique ID comprised of 6 digits from the end of the USDPID, and 4 digits from the end of the BADGE_NBR. We felt this method was pretty safe because the USDPID is an identifier that is non-public facing. Essentially, only Hydro Ottawa employees and maybe MDMR employees could identify a customer from this ID, if they really put some time into it.
- ***How was each technique used or applied? How difficult was each technique to use or apply?*** This was not a difficult technique to apply.
- ***How was de-identified data used or analyzed? What were the limitations on the use or analysis of de-identified data?*** Data was used to determine the impact of TOU savings for CDM targets. Not aware of any limitations.
- ***How and under what conditions was de-identified data disclosed?*** The de-identified data method was not disclosed
- ***Lessons learned*** This was an effective way to de-identify the data. We would use this technique again, however we may choose to use a different combination of ID’s depending on the use of the data- perhaps an ID that is static unlike the badge number.

NRCan Experience

The Tract and Neighbourhood Data Modelling (TaNDM) project

- ***What de-identification technique(s) was(were) used?***
 - De-personalization
 - Aggregation to privacy thresholds
- ***How was each technique used or applied? How difficult was each technique to use or apply?*** Building attributes held by BC Assessment (equivalent to MPAC) identified as required by data users (in this case, municipal energy planning and their collaborators) were reviewed through a Privacy Impact Assessment. One attribute, number of bedrooms, was deemed to be personal information and was removed. Addresses were maintained.

Data was aggregated to census tract and municipal levels both by building type and level of geography to privacy thresholds. Initially, thresholds were established for both residential and commercial at “no less than 5”. After review, these were changed to no less than 20 residential accounts and 3 commercial accounts. This is one aspect of aggregation that should be discussed, tested and verified to mitigate risk.

This work was done based on the privacy thresholds and a standard building category matrix describing buildings at the sector, category, sub-category and BCA manual class code levels. (Please note that data was not aggregated by postal code. Energy data aggregation by postal code is not recommended. Sources of error include inability to reliably associate postal code with building type and boundary issues with postal codes that vary in type and spatial extent and may cross jurisdictional boundaries.)

Python scripts were used to aggregate the data.

NRCan Experience (cont'd)

The Tract and Neighbourhood Data Modelling (TaNDM) project (cont'd)

- ***How was de-identified data used or analyzed? What were the limitations on the use or analysis of de-identified data?*** Summaries of energy use and GHG emissions were developed for pilot municipalities by BC Hydro and Fortis BC. Uses included for inventory purposes for community energy plans, verification of building energy usage for various building types against modelled or estimated values. Also for energy mapping exercises where aggregated average values were applied to buildings of the same type by floor area to come up with perhaps a more accurate energy map than could be obtained through modelled data alone. Additionally, BC Hydro made use of the improved information on the building stock for it's study of MURBs and multis, load analysis and forecasting etc....

NRCan Experience (cont'd)

The Tract and Neighbourhood Data Modelling (TaNDM) project (cont'd)

- **How and under what conditions was de-identified data disclosed?** BCA standard building report was made available through BC Assessment to municipalities. This same report was made available to third parties exclusively for energy and climate planning purposes (i.e. not profit seeking or other activities that would harass the public) through the BC Ministry of Environment, Climate Change secretariat. Any third party seeking this data is required to sign a data sharing agreement.

Community Energy and Emissions Inventories developed according to the TaNDM method were made available to pilot municipalities for planning purposes and selected researchers upon request.

- **Lessons learned** Data integration (matching parcel, to building to meter and associated energy use) is time consuming. Ideally, it should be done once and maintained in a central, authoritative and secure manner.

Thresholds for commercially sensitive data may be different for different companies. For instance, BC Hydro a provincial crown corporation established the 3 and 20 thresholds. Fortis BC a publicly traded company, abided by these thresholds for the purpose of the pilot project but then declined to do work for additional municipalities citing commercial sensitivities.