Smart Meter Data as a Source for Official Statistics

Keywords:
Electricity hub, register, linking, energy statistics

1 Introduction

In recent years, electricity smart meters have generated a lot of interest from producers of official statistics. Work on smart meter data has been reported for example by Carroll et al. [1], UNECE [2], Gask and Williams [3], Virgilito [4], and Williams and Gask [5]. The smart meters read electricity consumption and production from a distance and record data at short intervals. Large amounts of data, not only measurements but also attributes of metering points, customers, and providers, are collected in a common database (a hub). The data have an obvious attraction for official statistics producers. A Swedish data hub is under development and is planned for the end of 2020. The hub will contain data for more than 5 million metering points.
Since a few years, Statistics Sweden is recognized as a stakeholder. In additon, Stat Sweden participated in the recent project ESSnet Big Data [6]. In our paper, we describe the work we have done so far to prepare for the electricity hub as a possible source for relevant statistics on a number of topics, such as electricity use, environment and buildings. During preparation for the hub, we have access to a small set of test data.

2 Methods

2.1 Collaboration and project

To prepare for the hub as a source for official statistics, close cooperation with the parties involved is crucial. The hub is an administrative database, and its main purpose is to facilitate a supplier centric market model. Its potential for official statistics production is a bonus. Many issues have to be considered and solved before we can be sure how useful the data from the hub will be, or even if they will be useful at all. There are issues of quality, content, linking and modelling, as well as legal issues and questions regarding security, architecture, and it. In the Swedish case, there are several agencies involved. Svenska Kraftnät and the Energy Market Inspectorate are responsible for the hub, while Stat Sweden produces official statistics on commission for the Swedish Energy Agency. A common project involving all parties started in the fall of 2018, and will continue until May 2019.

2.2 Data quality

The final phase of collecting the needs of Stat Sweden is an important part of the project. Some major quality issues under discussion are summarized below:

- It is crucial that addresses are correct. The addresses will be of key value for linking to for example the Swedish Business Register (BR) since local unit identification is not available. The addresses are also important since they connect buildings, properties and households in the registers at Statistics Sweden. More

than 300 grid operators and electricity suppliers will upload their data in the hub. When data are migrated, the addresses will be checked against the register at the Land Surveying Authority. However, there will not be any continuous checks as new addresses are added to the hub.

- The quality of organization numbers must be carefully checked.
- An attribute for type of metering point will be added to the hub. The purpose of the attribute is twofold. It will keep track of "multiple use", i e if a metering point measures the electricity use of more than one user (apartment buildings, gallerias, etc). It will also give a detailed classification of buildings and property, and further classify object without an address, for example traffic lights and charging devices for electric cars.

2.3 Available data

We have access to a restricted set of test data, received from the systems of two main grid operators in Sweden. The data comprise two municipalities (Täby and Högsby) and cover two years, 2015 and 2016. Data contain monthly aggregates of consumption and production, and include both businesses and households. In the future Stat Sweden will be free to require any level of detail. The choice of detail depends on the statistics we wish to produce, but also on our ability to handle large amounts of data. The units are metering points. Each point has a unique identification number. If the metering point pertains to a business, it has an organization number (business level). Each metering point has an address. For the Täby data, we have coordinates on the metering points.
If a metering point does not have an organization number, we assume that it belongs to a household. The Täby data further contain apartment identification, so that the data can be divided on households in apartments and households in detached family houses. Businesses or households may produce their own electricity. The surplus is transferred into the grid and measured at the metering point. Thus, the production variable is a measure of surplus production, but there is no measurement of the amount of own produced electricity used by the business or household. For the Täby data, we also have information on type of production (solar cells, etc). The net production of solar energy by households in Täby was 52230 kWh.
Even though it is not real hub data we use the test data for two purposes:

- To identify future possibilities for improved or new statistics and investigate the possibilities to produce them. This includes for example how to link data to the registers kept at Statistics Sweden.
- To investigate data quality and gain insights into the content and structure of the data.

2.4 Outputs of interest

2.4.1 Data at local unit level

Today data for the survey Yearly energy statistics are collected by a web questionnaire, which is difficult for the respondents (grid operators) to answer and does not calibrate with their systems. Utilising hub data to estimate electricity use on NACE would lower the response burden, improve quality, and save time. It would also allow for estimates on new breakdowns and subgroups of interest.

With address information and coordinates, it will also be possible to link the data to the BR and the Swedish Population Register, as well as subject matter registers. This opens possibilities for statistics on a number of issues, such as consumption and production on relevant building characteristics, characteristics of household production of energy, etc. The Population register includes an apartment key for each individual, which makes it possible to distinguish households and facilitates matching of metering points to households. However, buildings with rented apartments usually only have one metering point for the whole building which makes it impossible to measure the energy used by the households, unless additional information is available. Similarly, local units without metering points (for example businesses renting space in a building) can not be distinguished. We forsee that modelling approaches have to be employed.

2.5 Linking to the Business Register

In order to classify energy consumption on NACE and test if relevant statistics can be produced, we have tested to link the metering points with businesses at local unit level using the BR. The register contains a lot of information, including NACE code for local units and unique identification of businesses and local units. Information in the test data useful for linking is organization number of the businesses for which energy consumption is measured, the street name and number of the metering point, and coordinates of the metering point.
A local unit may have several metering points, all of them having the same organization number, but different addresses. Matching on organization number alone revealed that some organization numbers could not be found in the BR. This probably indicates that they are invalid in the test data since it is unlikely that businesses are missing in BR. The organization number is not a crucial piece of information for the grid operators, and there is reason to doubt the quality.
The Täby data for 2016 contain coordinates of the metering point, measured with centimeter precision. The BR have coordinates for the addresses of local units, measured with precision of one meter. We tested whether these coordinates could be used to match metering points to local units.
The distances between all metering point and all addresses of local units in Täby were calculated. Distance is measured by the Euclidean distance,

$$d = \sqrt{(x_m - x_e)^2 + (y_m - y_e)^2},$$

where (x_m, y_m) are the coordinates for a metering point and (x_e, y_e) the coordinates for a local unit address.
We assume that a metering point measures the consumption of the local unit closest to the metering point, and that the nearest local unit must be within a certain distance. Thus, the local unit closest to the metering point is chosen, and its NACE code will be used to classify the consumption and the production at the metering point. The metering points belonging to the municipality of Täby were not included. They are identified by the organization number of the municipality. For each metering point, the distances to all local units in Täby were measured.

3 Preliminary results

We assume (on advice from the grid operators) that a metering point should be within a distance of 100 meters from the local unit it belongs to. There are 216 metering points that are too far away from any local unit and cannot be classified when this

rule is applied. Of those, only 44 are more than 200 meters from the nearest local unit. More than 67 percent of the metering points are within 50 meters from the nearest local unit. A local unit may have more than one metering point.
For the metering points that are within a distance of 100 meters from a a local unit, the adress for the metering point and the local unit are compared. If the addresses match, the NACE of the local unit is assigned to the metering point, otherwise it is not. In this way the algorithm classifies 2187 metering points and fails to classify 987.

4 Conclusions

Using the hub data and the BR it is possible to produce the Yearly Energy statistics. The linking method has to be further developed and the quality has to be improved, but our initial tests are promising. In these tests the addresses have been used but the quality of them are varying. In the final hub they will be of better quality.
In the real hub the organization numbers will be of better quality and can be used in the linking process, checking if the assigned local unit has the same organization number as the metering point. In our tests, a check revealed that large businesses are missing in the Täby data. Täby has a large shopping mall with 50 metering points that are difficult to match. For example, a bank office in the shopping mall has unreasonably large consumption. It is probably close to one of the metering points in the mall, but it is not the (only) local unit using the electricity measured at the point. Thus, large businesses cannot be distinguished because they are included in the consumption of the shopping mall. The minimum distance method clearly has many flaws, but could possibly be refined using more sophisticated tools for analyzing geo data.

References

[1] P. Carroll, T. Murphy, M. Hanley, D. Dempsey, and J. Dunne. Household classification using smart meter data. *Journal of Official Statistics*, 34, 2018.

[2] UNECE. *Experiment report: Canadian Smart Meter Data*, 2015.

[3] K. Gask and S. Williams. Analysing low electricity consumption using decc data. Technical report, Office for National Statistics, 2015.

[4] A. Virgilito. *Computing Energy Consumption from Smart Meters Data*, 2015.

[5] S. Williams and K. Gask. Modelling sample data from smart-type electricity meters to assess potential within official statistics. Technical report, Office for National Statistics, 2015.

[6] ESSnet Big Data. *Work Package 3, Smart meters*, 2016-2018.