**Electricity data hub**

*– a new data source in statistics*

Thomas Aanensen, Statistics Norway, Thomas.aanensen@ssb.no

**Abstract**

*An electricity market data hub was launched in Norway in the first quarter of 2019 and provides an exciting new and extensive data source which can be used for many different statistical purposes. Other Nordic countries like Sweden and Finland are in the processes of developing similar electricity market hubs and in Denmark and Estonia these types of hubs are already operational. The electricity data hub is a centralized information exchange system for the electricity retail market which contains data on consumption and production from remotely (and traditional) readable electricity metering points. The information stored in the data hub will primarily be used by electricity suppliers and distribution network companies serving electricity consumers more efficiently, however this very rich data source can also be a very useful source of information for us statisticians.*

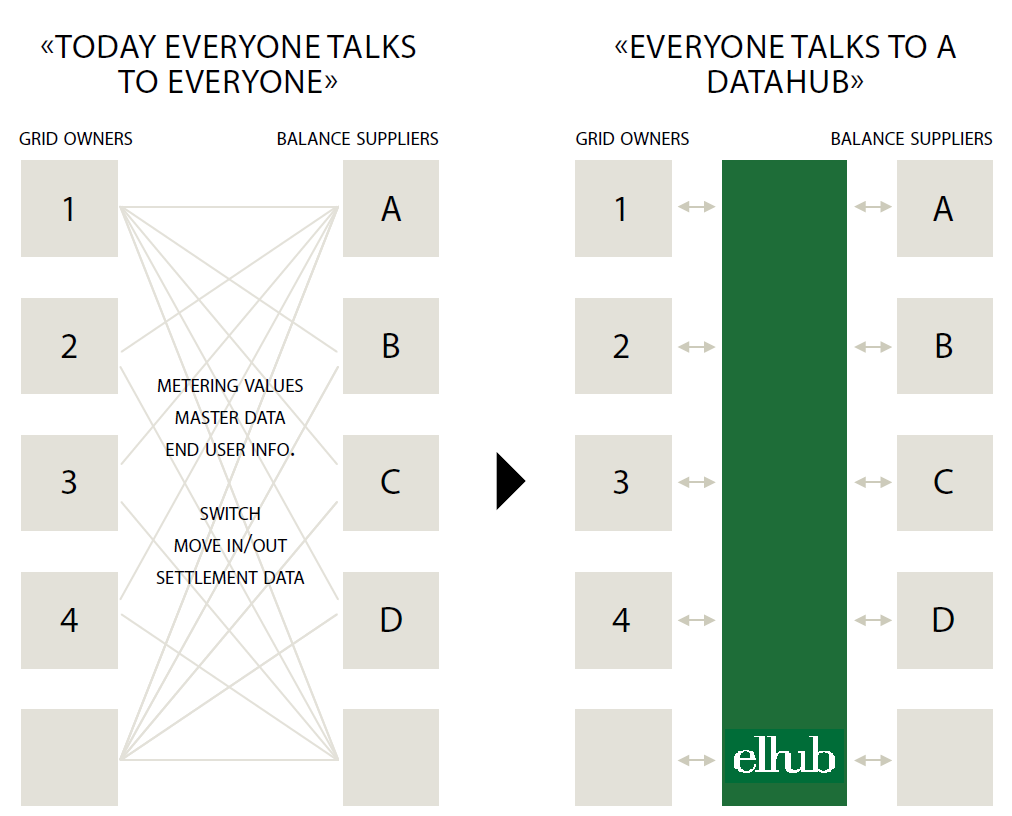
*In the paper we want to share our experiences from the process of getting hold of the data from the hub and discuss how this new and rich data source can be used to produce better and more timely electricity statistics and briefly look at other statistical purposes as for instance improving the quality of our statistical registers, identifying empty dwellings and use of the data for research purposes. We will also discuss some of the limitations in the data from the hub and how we plan to approach the regulating authority in the Norwegian electricity market with recommendations for improving the quality of the data and adding additional information to make the electricity hub more applicable for statistical purposes.*

**Keywords:** data hub electricity big data

**1. Introduction**

Installation of automatic smart electricity meters and the development of electricity data hubs to process and distribute market data more efficiently have increased the digitalization in the Nordic electricity markets. This new data source provides an exciting new source for official statistics. In Norway the electricity data hub went live in February 2019. Elhub, a subsidiary of Statnett[[1]](#footnote-1), is responsible for the project in Norway and the work to complete the data hub has been in progress for several years.

The electricity data hub[[2]](#footnote-2) is a centralized information exchange system for the electricity retail market which contains data on consumption and production of electricity and different market processes from all remotely (and traditional) readable electricity metering points in Norway. This adds up to around 3 million metering points. The information stored in the data hub will primarily be used by electricity suppliers and distribution network companies serving electricity consumers more efficiently. Earlier data on consumption and production of electricity, together with information about market processes, were handled by all the network companies and electricity suppliers. In the electricity data hub all these processes are centralized. Also, the hub utilizes the technological potential of the Smart Metering System in relation to smart solutions and services and it allows the consumer to give third-party service providers access to information from their electricity meters. This enables companies to develop technical solutions like mobile applications for monitoring of consumption and create systems to optimize electricity use in households.



Source: Elhub.no

**2. Getting data from the hub**

The electricity hub is an extensive new rich data sources on the electricity market which makes it a very useful source of information for statistical purposes. Early in the process of building the data hub Statistics Norway contacted the Elhub project, a subsidiary of Statnett, to ensure that Statistics Norway would receive the data relevant for statistical purposes. The statistical act in Norway[[3]](#footnote-3) states that Statistics Norway has the right to use administrative data systems in the state administration as the basis for official statistics. When new administrative data systems are established, as in the case of the Elhub system, Statistics Norway may put forward proposals on how the system should be designed to safeguard consideration for statistics. Early in the process we sent documents with suggestions on how to ensure that data from the electricity hub was suitable for producing official statistics. This work will continue also after go-live of the system in cooperation with Elhub and the Norwegian regulator of the electricity market NVE (the Norwegian Water Resources and Energy Directorate).

The electricity hub is a very complex system taking years to complete and designed to handle all the processing and calculation of the measured electricity meter data and market processes in the electricity market. Thus, understanding all aspects of the electricity hub, or collecting all the information, was not possible or desirable for us given our purpose to use the data in statistics. Naturally we only have limited resources to use on the data collection from the hub. It was therefore important to specify the data relevant to Statistics Norway early in the process and find a structure and format we could easily comprehend and adapt into our statistical systems at a low cost. After a series of meetings and close collaboration with developers from Elhub-team, we were able to identify variables relevant for statistical purposes and find a structure suitable for our needs. This work resulted in a formal specification of requirements that was sent to the Elhub-project in 2016, including information about formats, aggregation, type of variables, code lists, reporting frequency and so on. Also, Statistics Norway sent Statnett a formal letter with the disclosure requirements containing information about the legal authority, deadlines for data transfer, data transmission channel and the purpose of the data collection.

**2.1 Description of the data from Elhub**

Elhub contain hourly data on consumption and production on all metering points in Norway and the hourly data is sent from the distribution network companies to the Elhub once per day. In fact, the Elhub-system and smart meters are built to handle measurement every 15 minutes if it is decided to implement that frequency later. Since the Distribution System Operators (DSOs) are obliged to have installed smart meters at all sites by January 2019, close to all metering points in the Elhub system are now measured and sent automatically to the hub.

The dataset Statistics Norway receives from Elhub is aggregated up from hourly to monthly data before it is sent from Elhub. The reason for this aggregation is that we, as of today, don’t need hourly readings for producing official electricity statistics. However, if later on there is a need for data on an hourly interval, the dataset and data integration system is structured so that it will be relatively easy to adapt. It is possible that hourly data, for a limited time interval, might be relevant for research purposes or other statistical purposes in Statistics Norway at a later stage.

Statistics Norway receives the data from Elhub every month in a file with a comma-separated value format. The file contains about 50 variables arranged in columns and about 3 million rows corresponding to the number of metering points in Norway. All metering points have a unique identification number. The variables contain many different information elements about the metering points. Most importantly we get the measured production and consumption of electricity in kWh from the metering point. Other variables help us identify whether it is a consumption unit (enterprise, household, cabin, streetlight, etc.) or a production unit (hydro, wind, thermal, solar, etc.). Importantly each metering point have an end-user-id which is a national identity number in the case of a household or an organization number if the metering point refers to an enterprise. We also get other useful information about the metering point like the installed capacity (production units), ordinary or flexible[[4]](#footnote-4) consumption and whether the metering point was automatically or manually read.

There are multiple variables that contain information about the location of a metering point. This includes metering point address, municipality code and GPS location of the metering point. Currently only a third of the metering points have data on GPS location and this is voluntary information for the grid operators to report. Also, not all address variables in all metering points contains data. This will be discussed more in chapter 6. Other types of address information include end-user postal address and invoice address.

**2.2 Safe transfer of data**

Statistics Norway receives the dataset from Elhub six days after the reference month. The dataset is generated by a separate program in Elhub system designed to extract the relevant data from the hub and aggregate it up from hourly to monthly data. Due to the size of the dataset and the large amount of sensitive data, the data file is automatically transferred in encrypted format using a SFTP-solution. Currently Statnett is evaluating when they will start using a file transfer system called ECM (Energy Communication System[[5]](#footnote-5)) specifically designed for the energy market instead of the existing SFTP-solution. This will be used as file transfer system in the future. Once the file is transferred to statistics Norway, we store the data on a secure location and use a data integration platform called ODI (Oracle Data Integrator) to structure the data so that it is ready for further processing in our statistical systems.

**3. Elhub data in the electricity statistics - first use of the data**

The first use of the Elhub data was in the *Monthly electricity statistics*[[6]](#footnote-6) which was published in July 2019. The project of developing the new technical solution for compiling and disseminating the monthly electricity statistics, using Elhub data as source, started well in advance of Statistics Norway actually receiving data from the Elhub. Detailed specification of the datasets from the Elhub team and several rounds of test data made it possible to start the project of building the system well in advance of receiving the real datasets. This early start of the project ensured that we were able to use the new data source just a few months after the Elhub go-live date and we could start to assess the quality of the data and reveal problems at an early stage of the process which was very useful with this kind of new data source.

In the new production system for the *Monthly electricity statistics*, data for electricity production and consumptions is obtained from the Elhub data. We use information on the type of production to identify whether the electricity production from the metering point is hydro-, thermal- or wind power. In the statistic, electricity consumption is divided into four categories:

* *Power intensive manufacturing* (including sub-categories),
* *Extraction of crude oil and natural gas*
* *Pump storage use*
* Residual electricity consumption (all consumption except the groups mentioned above)

Earlier electricity consumption in *power intensive manufacturing*, *extraction of crude oil and natural gas and pump storage* was collected using surveys on a monthly basis. Now we use information from the Elhub dataset and aggregate metering points into consumer groups using information about the organization number of the metering points together with information from our business register.

The statistical production system is for the most part automatic and we use a combination of programs like ODI (Oracle Data integrator), SAS (Statistical Analysis Software), FAME (time series database) and editing tools to process and edit the data and create tables for dissemination of the results. The same work flow and programs are used each month which makes the process time-saving. Since the dataset is large and the programs perform some processing of the data, running the main programs in the initial phase takes a few hours.

There are several benefits from using the Elhub data in the Monthly electricity statistics. Timeliness can be increased significantly because we receive data only 6 days after the end of the reference period. This is much earlier compared to when data was collected in surveys. It also reduces the response burden for businesses since we no longer need to send surveys to collect information about electricity consumption on a monthly basis.

Another benefit is that Elhub data makes it possible to have a much more detailed division of consumer groups. This is something we will look into later this year and hopefully we can publish more detailed information on electricity consumption, on a monthly frequency, in the near future. It is also possible that the Annual electricity statistics, which use a different data source and is much more detailed, can be compiled from Elhub data in the future. That will make the production process of monthly and annual electricity statistics similar, more streamlined and standardized. This depends, among other things, on the comparability of the metering points to statistical units and to other registers in Statistics Norway. As with other administrative data, incomplete or noisy data and ambiguous linking to statistical units can pose challenges in the statistical production process.

It is anticipated that the Elhub data will be used in other energy statistics such as *Energy use in the manufacturing sector,* were it might be possible to omit questions about electricity use,and *Energy consumption in households* whichtoday collects electricity data from a sample of Grid owners. In the future, information about electricity use in households can be obtained from the Elhub data for the whole population, reducing response burden for the Grid Owners and increasing quality in the statistic. Also, it might be possible to publish electricity use in households by household size, type of dwelling, place of residence and so on, based information from Elhub and statistical registers in Statistics Norway.

**4. Other statistical purposes for the data**

The large and detailed dataset from Elhub has potentially broad applications in statistical production in Statistics Norway. Now that the data hub is launched, and we receive data on a regular basis, we can start to explore the application of the data in different areas in Statistics Norway. Because of the large potential of this data source and the possibility to link the data to other registers, it will take time before we know all the applications of this new data source.

Several projects in Statistics Norway are working on the potential use of the Elhub-data. One of the projects is Essnet Big Data 2 Work package D on smart energy, where Statistics Norway will look at the possibility of producing statistics on vacant dwellings. Another project is a European Commission grant where the main purpose is to study the potential role of Elhub for population management of the Register of Addresses and Buildings.

**5. Elhub data in research**

Data from Elhub has the potential of being a very useful data source for research purposes. The large amount of high frequency data, potentially combined with other registers in Statistics Norway, opens for a variety of different research possibilities. Elhub is a completely new data source and we do not have much experience regarding research possibilities yet. In a few years’ time we will hopefully be able to tell more about different applications of Elhub data in research-projects.

Statistics Norway has for many years published studies and research projects on household energy- and electricity consumption. Now, as micro data on all households in Norway is available for research purposes, these projects can most likely benefit from this more reliable data source which comprehend information about the whole population. It will also reduce the cost of collecting the data both for Statistics Norway and the grid companies, which provided data in earlier studies.

Earlier, only dataset for a whole year was collected and the collection was infrequently. Now that we have monthly datasets from Elhub it is possible to study the effects of climate and temperature on energy and electricity consumption over time and how electricity is used within households. The monthly data, and perhaps later on also hourly data, make it possible to use many different time series and panel data estimation methods. The quality of such analysis will crucially depend on the possibility to link data from Elhub to information from other registers and survey data in Statistics Norway like household size, type of building, household income and so on.

**6. Limitations in the data and possible solutions**

The usefulness of the information from Elhub data in official statistics depends crucially on the compatibility of information about the metering points to statistical units used by the Statistics Norway and the quality of the variables reported by the network and electricity companies in the Elhub-system. To improve the quality of the Elhub data it is important for us to continue the close cooperation with the Norwegian regulator of the electricity market (NVE), the owner of the Elhub project (Statnett) and the Elhub organization to make sure the Elhub data is applicable for statistical purposes so that the full potential of the data can be used.

To address these issues Statistics Norway has established a team that are looking into the quality of the Elhub data and how well the data is suitable for statistical purposes. The work of the team will result in a report to NVE and Statnett later this year. The report will address issues like the completeness of the data we receive, the quality of the variables and if variables are correctly defined and reported according to national standards. Also, the team will suggest new and additional variables if they think it is needed to ensure that data can be used efficiently for statistical purposes. It is important to work closely with the electricity regulator and data owner since they have the legally authority to instruct the network and electricity companies on how they report data and, if relevant, demand Elhub to make the necessary changes in their data systems.

We are just in the initial phase of analyzing the data form the hub, thus it is too early to draw any conclusions on data quality for extensive new statistical use now. Some of the challenges we have identified in the first face of using the data is that not all variables we receive are mandatory to report for the network companies. For instance, some of the variables relating to addresses and location of metering points, like GPS location of the metering point, is not mandatory to report and only a limited number of the network companies report this information. Another concern is that some of the variables relating to address information seem do deviate from the national standards for geographical information and localization, making it challenging to separate metering points in buildings with multiple electricity meters.

An important variable for statistical use is the organizational number of the establishment, which is the locally delimited functional unit that is situated in a single location and in which only a single productive activity is carried out. At present the data on metering points from Elhub does not include information about the organizational number of the establishment (local kind of activity unit). It only contains the organization number of the enterprise (legal unit). This will pose challenges in the statistical production since a proportion of the enterprises consists of more than one establishment.

Despite the concerns raised in this initial face about possible challenges in the data from the electricity hub, much of the data looks promising and has the potential of being very useful for many different statistical purposes. We will know more about the possibilities and challenges in the data once we get a closer look at the data material from the hub. Hopefully some of the challenges can be solved in cooperation with NVE, Statnett and Elhub.

**7. Electricity data hubs in the Nordic countries**

The transitions towards data hubs in the electricity market are also in progress in the other Nordic countries. As in Norway, the energy market regulators in the Nordic countries have asked their transmission system operator (TSO) to build data hubs for their respective electricity market. The Nordic data hubs have many similarities. The main common feature is that they are all designed to handle metering values and several of the market processes in the electricity market in an effective way where electricity suppliers and grid owners get information from the hub instead of a situation where “everybody talks to everybody” (NordReg, 2018).

Denmark was the first Nordic country to launch their electricity data hub back in March 2013 and later, in 2016, updated it to allow for a supplier-centric model[[7]](#footnote-7). In Finland they are progressing with their data hub plans. In June 2018 they announced the company CGI Finland as the system supplier of the data hub system and they estimate that the hub will be ready to launch in April 2021. In Sweden Svenska kraftnät has been assigned the task to develop and operate the data hub. They anticipate that a data hub for electricity market will be available at the earliest towards the end of 2020 and deployed the second half of 2021. On their web page they inform that the project will give an updated time schedule in second quarter 2019.

**8. References**

NordReg (2018), Implementation of data hubs in the Nordic countries. Available at: <http://www.nordicenergyregulators.org/wp-content/uploads/2018/08/NordREG-Status-report-on-datahubs-June-2018.pdf> (Accessed: 14 May 2019)

1. Statnett is the main transmission system operator (TSO) in Norway [↑](#footnote-ref-1)
2. For a short visual presentation of the Elhub project, please take a look at <https://www.youtube.com/watch?v=uYVo1MHBFvc> [↑](#footnote-ref-2)
3. <https://www.ssb.no/en/omssb/styringsdokumenter/lover-og-prinsipper/the-statistics-act-of-1989> [↑](#footnote-ref-3)
4. For consumers with flexible electricity contracts electricity supply can be stopped for a defined period of time in the event of limited net capacity. [↑](#footnote-ref-4)
5. For more information about the ECP-platform, please take a look at this web page <https://unicorn.com/files/ecp-energy-communication-platform-n03.pdf> [↑](#footnote-ref-5)
6. <https://www.ssb.no/en/energi-og-industri/statistikker/elektrisitet> [↑](#footnote-ref-6)
7. A supplier-centric model implies that electricity consumers only need to have one electricity trading company that will invoice both distribution and consumption of electricity. [↑](#footnote-ref-7)