European Distribution System Operators

E.DSO Policy Brief on Open Data

December 2018

Contents

Contributors:

Coordinator: Ruud Berndsen **Secretary:** Henning Twickler

Participants:

Alliander	Bas de Vrind, Daan Rutten
EDP Distribuição	Luis Vale Cunha
Enedis	Jean-François Montagne
Fluvius	Sven van den Bosch, Noemi Loeman
Innogy	Torsten Knop

1. Introduction

DSOs hold many datasets which, when made publicly available, can help other stakeholders and market parties with e.g. better decision making, create new services and promote synergies between different sectors. Publishing data in the form of open data can unlock many types of benefits, amongst which contributions to the energy transition. On the other hand, not all data is suitable to make publicly available due to potential breaches of security or violations of privacy regulations. It is therefore important for DSOs to have a common understanding.

This E.DSO Policy Brief addresses the definition of open data as well as the benefits and trade-offs involved in deciding whether to make data openly available. To illustrate the possibilities of open data, actual uses cases are added as well as building blocks of an indicative open data framework.

2. What is open data?

E.DSO views open data as data which is made publicly available by the government, organizations, publicly owned as well as privately owned companies. Data is considered to be open if anyone can freely use, re-use and re-distribute it for any purpose, without restrictions. To be considered open, the data must be re-usable, meaning it can be downloaded in open formats and read by software, and users have a legal right to re-use the data.

Although open data has no legal definition, the Open Knowledge Foundation applies a similar interpretation.[1] In its Open Definition, the Foundation provides full details on the requirements for 'open' data and content. The definition[2] states that open data can be freely used, modified, and shared by anyone for any purpose.

In general, we note that open data has most of the following characteristics: (1) free of charge, (2) available 'as is', (3) duty free, (4) accessible without registration, (5) computer processable[3], (6) provided with metadata, (7) as complete and rough as possible, (8) current and (9) easy to find. Open data hereby differs from data-enabled services or products. In the latter, value is added to a dataset by, for example, providing tailor-made subsets of data, produce advice or app-development. In this case, pricing and revenue models could come into play. This means that in addition to an offer of open datasets, a layered model of (tailor-made) data services or products will arise.

[1] Open Knowledge is a worldwide non-profit network of people passionate about openness, using advocacy, technology and training to unlock information and enable people to work with it to create and share knowledge.
[2] Detailed information on the subject can be found at https://okfn.org/opendata/.
[3] As a growth model for an open data practice, reference can be made to the five star model for open data as proposed by Tim Berners-Lee (http://5stardata.info/en/).

E.DSO members are committed to pursuing an open data policy that aims to foster the development of innovative solutions, increase transparency and promote synergies and cooperation in the energy sector and beyond. DSOs have a lot of data at their disposal in the context of their legal tasks and obligations which can be extremely useful for numerous stakeholders and market parties. Broadly speaking, the data at the DSOs disposal can be divided into the following three categories:

1. Meter data; typically collected at customers' premises.
2. Grid data; includes information on the grid, its configuration and measurements, and can be either real-time, planned or historic.
3. Market data; refers to all types of exogenous data, such as market results or information on installations at customer premises necessary to offer services related to those assets.

In the tables below, examples of open datasets are listed. This listing is not necessarily exhaustive, though provides an indication of data that could be useful for different market parties.

1. METER DATA
1. Consumption data, can be aggregated per DSO, city or municipality or district, per type of renewable energy source.
2. Production data, can be aggregated per DSO, city or municipality or district, per type of renewable energy source.

2. GRID DATA
1. Requested installed power in grid studies, can be aggregated per DSO, city or municipality or district, per renewable energy source
2. Planned maintenance, can be aggregated by city or municipality, district or street.

3. MARKET DATA
1. Data on installed power, can aggregated per DSO, city or municipality or district, per renewable energy source.
2. Flexibility data (= data on location and type of flexibility source), can be aggregated per DSO, city or municipality or district.

3. Benefits of open data

Providing open data is seen as a high-potential facilitator of the energy transition. By unlocking data from its original owner and offering it publicly, re-use may enable system-relevant innovation. People, companies and organisations may use open data to launch new ventures, analyse patterns and trends, make data-driven decisions and solve complex problems. Large-scale sharing of data, therefore, has the potential to unlock economic value by enabling the development of new products, services and markets. E.DSO members view the benefits as follows.

Society

The wider availability of factual information enables better decision-making by local and national policy makers. Open data therefore has the potential to stimulate evidence-based policymaking and increase efficiency in public administrations. It can also be used for research purposes, for example by universities, and foster the participation of citizens in political and social life by providing valuable insights into their lives. This includes cross-border comparisons with other EU Member States.

Spark innovation and new services

By providing data free of charge for everyone to (re-)use, open data reduces potential market entry barriers for smaller enterprises and new entrants. It enables anyone to access and use the data provided and to combine it with other data sets or analytical tools. This way, third parties can offer new services to consumers and market parties, including the DSOs themselves. This has clear benefits not only for the parties directly involved, but for society as a whole. The more parties working on the same topic, the more options become available and the more efficient the overall system will become. Furthermore, open data also minimises the risk of excessive first-mover advantages and with-it potential lock-ins, thereby reducing the risk of the monopolisation of market segments.

Alliander use case: data product by a third party

Alliander publishes yearly aggregated household energy consumption data. Third parties use the open dataset to create data services or products. In this example, the Dutch Ministry of Economic Affairs and Climate has created a map, based on open datasets from various DSOs, that shows the average energy consumption per area code in the entire country. The darker the colour, the higher the consumption level.

Transparency

The distribution of electricity and gas is an essential public good and as such DSOs fulfil an important public role. Transparency provided through the sharing of open data enhances the service provided to society and enables analyses of DSOs performance, including benchmarking. Moreover, open data has an educational value as it increases the amount of factual information available to the wider public. This leads to a better understanding of energy distribution, an increasingly important segment of the energy system as a result of development that come with the energy transition.

4. Trade-offs

As operators of critical infrastructure, DSOs offer essential services to society and have access to large amounts of data. Before making data available to the general public, important considerations have to be addressed first. Once data is made publicly available, it's impossible to withdraw it. It is therefore essential to recognise sector-specific considerations in order to decide which data can be made publicly available and which not. This is of utmost importance, especially for operators of critical infrastructure and essential services such as DSOs, that are dealing with security- and customer-sensitive data. A horizontal approach to open data, as currently proposed in the Directive on the re-use of Public Sector Information (PSI), is not the right approach. Consequently, they should be exempted from a general obligation. Instead, it is suggested to take the following principles into consideration before publishing open data. Clear guidelines can help judge which data is suitable to become open data.

Protection of critical infrastructure and essential services

DSOs possess (data of) critical infrastructure and provide essential services for society. It should therefore be avoided that data made publically available can help malicious users to harm the infrastructure or disturb services. Security threats include, amongst others, acts of terrorism on the physical infrastructure, theft or hacks of IT systems.

Protection of privacy

The privacy of individuals has to be protected when disclosing data. When personal data is involved, anonymization or aggregation can be used to mitigate privacy concerns. In addition, care should be taken that personal data is still protected when different datasets are combined (so called mixed data sets), including by third parties. Privacy laws and legislation are to be respected at all times and the GDPR provides clear guidelines on how to deal with data sets that include personal data.

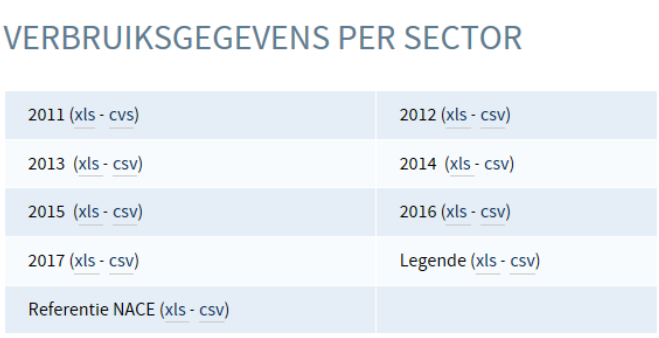
Avoidance of negative externalities

Although the purpose of open data is to provide benefits to society, be it economic, political or social, making data publicly available can also have negative unintended consequences. Any negative impact on DSO activities is to be avoided at all times, including the unwanted

disclosure of specific data on stakeholders or other infrastructure. At the same time, questions of liability must be addressed, e.g. through licensing, to prevent exposing open data providers to undue legal repercussions.

Right interpretation of data

When data is made publicly available, it's key that users interpret the data correctly. Misinterpretation can lead to wrong conclusions and undesired outcomes. Open data therefore requires adequate explanation of the meaning of data provided, for example through a legend. It is also important that open data providers should not be held accountable under any circumstances for misuse or misinterpretation of the data they make publicly available.



Fluvius use case: aggregated consumption data per sector of activity

Fluvius publishes on a yearly basis aggregated consumption data per sector of activity. Each open dataset is accompanied by a descriptive legend that explains to the user how to read and use the flat data list. Below is a random part of this open dataset.

A	B	C	D	E	F	G	H
Verbruiksjaar	Hoofdgemeente	Energie	Injectie/Afname	Sector	Subsector	Aantal toegangspunten	Benaderend Verbruik (kWh)
2017	AALST	Aardgas	Afname	Empty/onbekend	Empty/onbekend	762	22.330.208,66
2017	AALST	Aardgas	Afname	ENERGIESECTOR	elektriciteit- en warmte sector	12	3.166.095,01
2017	AALST	Aardgas	Afname	ENERGIESECTOR	Rest	3	481.159,55
2017	AALST	Aardgas	Afname	HUISHOUDENS	Rest	1	37.555,97
2017	AALST	Aardgas	Afname	Huishoudens	Huishoudelijk	26.303	383.667.402,70
2017	AALST	Aardgas	Afname	INDUSTRIE	Metaalverwerkende nijverheid	20	4.523.635,68
2017	AALST	Aardgas	Afname	INDUSTRIE	Papier en uitgeverijen	5	68.969,45
2017	AALST	Aardgas	Afname	INDUSTRIE	Rest	120	123.390.690,04
2017	AALST	Aardgas	Afname	INDUSTRIE	Voeding, dranken en tabak	33	2.360.416,91

Remuneration

Open data is made publicly available and anyone can freely use, re-use and re-distribute it for any purpose, without restrictions. The costs for providing open data must be reasonable and DSOs need to be able to recover costs through tariffs. This does not limit options to fulfil data requests that are costly, address specific sub sets or are tailored for specific stakeholders. Paid provisioning of data can be considered as a suitable alternative. However, as stated in the second section, this will be considered a data service and not open data.

5. Building blocks

To maximise the potential of open data, DSOs should aim to share data sets in the most accessible way possible. Without being exhaustive, what follows are some building blocks based on experiences so far, including two use cases.

EDPD use case: open data – Energy data

Early 2017, EDPD took a relevant step towards digitalization by creating an open platform for the provision of information associated with energy data. EDPD launched this platform with the utmost intention of promoting the involvement of society in the energy transition, namely by:

- creating a new platform to make energy data externally and freely accessible;
- making data re-usable for different purposes and promoting innovative approaches;
- giving external parties the possibility to create societal value based on existing data, acting in compliance with existing regulations (GDRP).

The energy data in the platform is loaded with 15-minute detailed historic information which goes back to 2014 and is divided in two main groups: generation and consumption, separated by voltage level. The data can be consulted with daily, monthly or yearly aggregation and the stats area of the platform provides an automate comparison with the homologous period. Furthermore, the platform has information regarding the temperature of each day, so that the user can perform a correlation analysis between energy data and temperature.

EDPD's open data platform was built on top of their own data handling systems and made available daily, with validated and certified information of the previous day.

As part of ongoing developments on this platform, EDPD is projecting:

- new functionalities (e.g. forecasting);
- some key functionalities that will further promote the advance use of data.

EDPD's open data platform is one of its answers to the needs of an ever-increasing digitalized society.

Data utilization (terms and conditions)

Once the decision is made to open and make data publicly available for re-use, it is important to specify the conditions of (re-)use of the data. This could include (non-exhaustive):

- obligation to mention data sources;
- data provider liability : restrictions, whatever may be the use of the data and its quality;
- obligation to publish new datasets built out of one or more open data sets;
- restrictions for commercial use / sales

These conditions may be secured through a license, of which several examples are:

- ODBL[4], inspired from the open source community, which obliges the user to re-publish (with conditions) any new dataset built from open data with an ODBL license;
- Creative commons[5], which terms of reuse conditions allow variants depending on the "openness" wished by the data source (attribution / modification / commercial use / share alike obligations);
- other licenses developed by countries which initiated an open government program.[6]

Platforms and infrastructure

(Data) platforms are an important part of the solution when we talk about 'how' to disclose the data. A data platform is typically characterized by a layered architecture, with different conceptual layers: a data layer, a data exchange layer and an application layer. Data platforms can gather several functionalities, e.g.:

- dataset* import;
- data storage (without data duplication)[7];
- datasets management (those who are not owners of the dataset must ask permission from the dataset owner if they wish to add, remove or change items to the dataset);
- data request routing to the right party that can deliver the data;
- data publication through standardized accesses and files;
- data conversion into standardized formats;
- data visualization;
- user management functionalities (e.g. subscriptions/notifications, ...).

Data platforms and infrastructure can take various forms and rely on Open Source as well as commercial software, for example:

- Platform as a service (provides a platform allowing customers to develop, run, and manage applications);
- Cloud based open data serve (runs at a remote facility such as a server farm or cloud)
- On premise architectures (is installed and runs on computers on the premises of the person or organization using the software)

Data model, data format, access mode

The easy re-use of open data relies on:

- the easy access of it through standardized interfaces (API REST, JSON, …) and data file formats (XML, CSV, …) when data is downloaded as a file;

[4] https://en.wikipedia.org/wiki/Open_Database_License

[5] https://creativecommons.org/about/program-areas/open-data/

[6] https://opendefinition.org/licenses/. France: https://en.wikipedia.org/wiki/Open_licence_(French). Eandis : https://www.eandis.be/sites/eandis/files/documents/gebruikslicentie_-_open_data_van_eandis.pdf

[7] This means that the party who is the natural owner retains final responsibility for the dataset in its own back-end.

- the well understanding of what the data is and how it can be used: data glossary, definitions, use of shared ontologies (see also the concept of "linked data"[8]) and standardized representations (units, geocoding, etc.)
  the easiness to combine it with other datasets – in particular, the superposition of layers over a geographical map can be very useful for data representation and visualization.

Data visualization

Data visualization ('Dataviz') has become an expertise domain in itself; it is a powerful tool to make open data more accessible and understandable. Data visualization is a valid example of what an extra layer on top of an open dataset can imply. It adds comprehensive value to open data sets by making it more accessible to a wider public.

One very common Dataviz application is the geographical representation, and the superposition of one or several informations on one single geographical map. The use of Googlemaps API illustrates the application of this kind of tool. As alternative to Googlemaps, the OpenStreetMaps platform[9] gives the possibility to freely use maps based on information collected by crowdsourcing. Specific information layers are also available - e.g. energy infrastructure. Some open data platforms offer integrated basic Dataviz functionalities and give a first-level tool to valorize opened datasets. Additional visualization effects can be brought in by specific tools and libraries; they enable the representation of the most specific datasets.

Enedis use case: open data platform

Enedis initiated an open data project in 2015 and started an agile approach using a Platform as a service offer ("PaaS"). It allowed the company to easily and quickly publish the first datasets with educational data visualisations: Enedis electricity balance with aggregated consumption/generation/losses. This quick win brought a demonstrative force to Enedis' willingness to commit to open data. The new publications increased every year with currently (October 2018) 53 open datasets available in the following categories:

- Enedis electricity monthly balance: generation, consumption, losses (monthly);
- consumption profiles and number of generation/delivery points (yearly);
- aggregated anonymized meter data (consumption, generation) at different scales (yearly);
- generation capacity, installed power plants, waiting queue for new installations (quaterly);
- new connections by type of customers (yearly);
- EV charging capabilities by region (quaterly);
- performance indicators: SAIDI and other indicators (yearly);
- network infrastructure: km of lines, overhead lines and substations position (yearly).

Data visualisation is still used to help understand the dynamics of electricity distribution.

[8] See https://en.wikipedia.org/wiki/Linked_data and http://linkeddata.org/.
[9] See https://www.openstreetmap.org.

As for data access, Enedis chose to use the Free License originated by the French government[10]. Moreover, data access conditions have been stated in "general terms of use" associated to the use of the open data portal and cover:

- quality of service: availability;
- limitation of liability;
- conditions of use ("fair use" of the data);
- cookies / tracking information.

6. Final thoughts

Data collected by DSOs in the context of their legally imposed tasks and work area can bring many benefits when made publicly available. Open data can contribute to better decision making, spark innovation and services and increase transparency. However, not all data is suitable for this purpose as there are important trade-offs involved, for example data that can harm critical infrastructure or essential services, personal data or privacy-sensitive data. The examples provided in this paper show that open data is a relevant topic for many European DSOs and other stakeholders. They also highlight the different approaches currently adopted across the continent, each shaped by their own specific national context.

[10] https://en.wikipedia.org/wiki/Open_licence_(French).

E.DSO

SHAPING SMARTER GRIDS FOR YOUR FUTURE

E.DSO is a European association gathering leading electricity distribution system operators (DSOs) **shaping smart grids for your future.**

www.edsoforsmartgrids.eu