

There are the following problems with open datasets for machine learning in the power industry:

- **Data privacy.** There is an ongoing debate about whether it is truly possible to anonymize data like this or whether any system of anonymization can ultimately be defeated. If successful technologies for anonymization are developed, they will open up new opportunities for data analysis and publication. On the other hand, if experts and the public come to believe that individuals' identities can always be deciphered from the data, then several paths to innovation will be cut off.
- **Poor quality data.** Power industry data are often stocked in obsolete data management systems and may have errors, gaps in the data, or out-of-date information. These are not easy problems to solve, but without some solutions, the potential of open data in power industry will be underused.

The first problem can be solved with the help of government. In this case the government can create a thriving ecosystem of scientists, coders, and application developers. Also, it is necessary to describe main business scenarios of open data usage in regulatory documents. It will stimulate power industry to share their datasets and to connect to different stakeholders to promote data-driven innovations.

The solution of second problem follows the solution of the first one. Once there is an open data ecosystem, it becomes possible to establish common requirements for data quality and to create platform solution to check the data quality and make it better if it is needed.

The proposed open data ecosystem includes the following stakeholders:

- government
- data suppliers (energy power companies, government)
- aggregators (gather data in one organized and well protected space)
- application and algorithm developers
- enablers (data preparers)

Possible benefits for them are as follows:

- 1) **Advertising.** For example, most weather forecast services rely on data from observation stations operated or sponsored by the National Weather Service and gain revenue through geographically targeted advertising.
- 2) **Subscription models.** Some of the forecast services offer a subscription service with better data and deeper data history.
- 3) **Lead generation.** Stakeholders can collect and analyze data on the power industry. Users can request necessary information from contractors.
- 4) **Analytics fees.** Stakeholders (enablers and algorithm developers) can provide services for load or generation forecasting or predictive analytics of power equipment.
- 5) **Consulting fees.** Stakeholders can analyze proprietary data and advise their power industry clients to improve the efficiency of power supply system.

Moreover, data suppliers such as DSO, TSO, and ISO can get access to **ML-driven innovations** if they give an access to their data.

To make all this possible, we propose to bring together data providers with data users for a structured, action-oriented dialogue to identify the most important datasets for business and public use and find ways to improve them.

The architecture of the proposed Open Data platform is presented on the figure below.

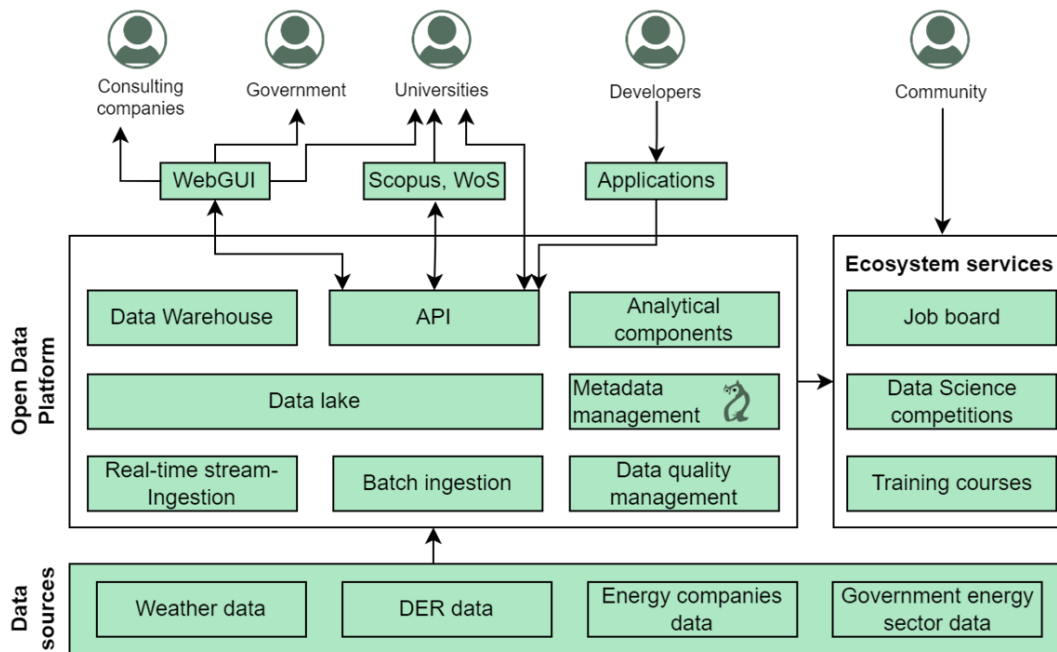


Figure 1 – Open Data Platform architecture for power industry

Open Data platform allows:

- 1) **to organize a “Job board”**, where the power industry can describe their problems and propose different tasks to community members.
- 2) **to promote data science** in the power industry by stimulating community members to use ML-driven algorithms.
- 3) **to organize data using the ontology approach** and to sell it as a service. Through this, it becomes possible to improve the data quality and reduce costs for application modernization.
- 4) **to speed up ML-driven innovations** by giving an access to application developers. It will allow to improve ML-based algorithm before its implementation.
- 5) **to contribute to the formation of ESG frameworks.** It gives businesses a rigorous and transparent framework to deliver carbon neutrality for their business, product and/or activities.
- 6) **to accelerate RES integration** by increasing the investment attractiveness of projects with RES.

In conclusion, we would like to underline that it is necessary to set out some basic principles for the government to use in promoting Open Data in the power industry. Ultimately, creating a better Open Data ecosystem will take both public and private resources and funding.