The Journal of Open Source Software

open-mastr: A Python Package to Download and Process the German Energy Registry Marktstammdatenregister

Florian Kotthoff [1], **Christoph Muschner** [2], **Deniz Tepe** [1], **Guido Pleßmann** [2], **and Ludwig Hülk** [2¶]

1 fortiss, Research Institute of the Free State of Bavaria, Guerickestraße 25, 80805 München, Germany **2** Reiner Lemoine Institut gGmbH, Rudower Chaussee 12, 12489 Berlin, Germany ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 11 December 2023
Published: unpublished

Summary

The Python Package `open-mastr` provides an interface for accessing and cleaning the German Energy Unit dataset called *Marktstammdatenregister* (MaStR). The MaStR is a central registry with detailed information about renewable and conventional power plants in the German energy system. `open-mastr` enables the creation and updating of a local database of the entire registry, as well as processing the data for further evaluation. Ultimately, the package offers methods to reduce the registry's parsing time and thus enables energy system researchers to start working with the entire dataset right away.

Statement of need

`open-mastr` has been built to facilitate the process of downloading, parsing, and cleaning the MaStR dataset. The MaStR is a German registry provided by the German Federal Network Agency (Bundesnetzagentur / BNetzA) (Bundesnetzagentur, 2019). It was first published in 2019 and includes detailed information about more than 8.2 million data points covering electricity and gas production units, electricity and gas consumers, storages, grids, and energy market participants. Tepe et al. (2023) found 54 papers in the fields of sustainability studies, energy politics, energy data, energy system analysis, and energy economics that used the MaStR dataset in their research.

Besides its relevance in research, the raw MaStR dataset provided by BNetzA bears some obstacles: First, the documentation of the data model and download methods are only provided in German. Second, many entries in the dataset are encoded. Third, information that belongs together is distributed over several tables. And finally, the provision of the raw MaStR dataset as xml files is not optimal for fast and simple access. The Python Package `open-mastr` addresses those issues as shown in Table 1.

Table 1. Summary of benefits provided by 'open-mastr'

Benefit	Description
Data download and parsing	Download, decode, and write data to a local database
Translation to English	Translate table names and columns from German to English as well as an English documentation page of the dataset

Benefit	Description
Data processing	Merge relevant information about different technologies to single csv files

Besides `open-mastr`, no other software solution exists that provides an interface to download and clean the MaStR dataset. For other energy-related data, similar solutions exist: The *iotools* module from `pvlib` implements access to different raw data sources via its *get* methods (Holmgren et al., 2018). Data platforms offer another approach to provide cleansed datasets: This is done by the Open Energy Platform (Hülk, Glauer, et al., 2022), the Open Power System Data (Wiese et al., 2019), the Global Power Plant Database (Byers et al., 2018), or the Public Utility Data Liberation Project (Selvans et al., 2020). The advantage of aforementioned data platforms is their simplicity in accessing data for end users, as users can simply download files in standardized formats, such as csv. The disadvantage is that end users have to rely on platform maintainers for data currency and correctness. Here, `open-mastr` comes at hand by providing direct access for end users to the original data.

Package description

The first and main use-case of the `open-mastr` package is to download and parse the MaStR registry to a local database. The `open_mastr.Mastr` class and its `download` methods are used to achieve this. The whole MaStR registry is downloaded from the MaStR website as zipped xml files when running the `download` method.
The downloaded xml files are extracted and parsed to a sqlite database. This results in a local database of the MaStR with a size larger than 5GB. However, many textual data points are still encoded by IDs. Thus, as a last step, the `Mastr.download` method decodes the IDs to their actual meaning.

The local database is then ready for further processing. Its columns can be translated to English with the `Mastr.translate` method. Relevant information about different technologies, like wind turbines or PV systems, can be merged from multiple tables and written to csv using the `Mastr.to_csv` method.

A second use-case is the wrapper for the MaStR SOAP API. Calling the SOAP API directly can be interesting for specific users who do not need to download the whole registry. All possible API requests, as described in the official documentation are callable as methods of an `soap_api.download.MaStRAPI` object. The classes `soap_api.download.MaStRDownload` and `soap_api.mirror.MaStRMirror` use the API to download some tables or the whole registry. Both classes offer very similar functionalities to the basic `Mastr.download` function with the differences, that they require an API key and a daily API request limit exists. Hence, the use-cases of the `MaStRDownload` and `MaStRMirror` are limited since BNetzA offers a way to download the whole registry as zipped xml files, as implemented in aforementioned `Mastr.download`.

As an extra service for people that are not familiar with Python, the developers offer the cleaned and reduced dataset created with `Mastr.to_csv` on Zenodo (Hülk, Pleßmann, et al., 2022).

Conclusion

In summary, `open-mastr` gathers community developed code to work with the Marktstammdatenregister. It simplifies the data parsing and cleaning process and thus facilitates data-based research and energy system planning. In the future, a steady maintenance of the Python

package is needed to handle BNetzA induced changes in the dataset and its data model. It is also planned to enhance MaStR's metadata in the future to further comply with FAIR data standards.

CRediT Authorship Statement

FK: Writing original draft, creating and maintaining code for bulk download, writing documentation page CM: Creating and maintaining code for API download, Review of draft, writing documentation page DT: Maintaining code for API, bulk download, and csv export GP: Creating code for API download LH: Creating code for API download, Review of draft

Acknowledgements

FK and DT acknowledge support by "Bayerische Staatsministerium für Wirtschaft, Landesentwicklung und Energie" as part of "Bayerischen Verbundförderprogramms (BayVFP) – Förderlinie Digitalisierung – Förderbereich Informations- und Kommunikationstechnik".

References

Bundesnetzagentur. (2019). *Marktstammdatenregister*. https://www.marktstammdatenregister.de/MaStR

Byers, L., Friedrich, J., Hennig, R., Kressig, A., Li, X., McCormick, C., & Valeri, L. M. (2018). A global database of power plants. *World Resources Institute*, *18*.

Holmgren, W. F., Hansen, C. W., & Mikofski, M. A. (2018). Pvlib python: A python package for modeling solar energy systems. *Journal of Open Source Software*, *3*(29), 884. https://doi.org/10.21105/joss.00884

Hülk, L., Glauer, M., Huber, J., & Winger, C. (2022). *Open Energy Family - Open Energy Platform (OEP)* (Version 0.14.2). https://github.com/OpenEnergyPlatform/oeplatform/

Hülk, L., Pleßmann, G., Muschner, C., Kotthoff, F., & Tepe, D. (2022). *Open-MaStR - marktstammdatenregister* (Version 2022-12-01_B) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7387843

Selvans, Z., Gosnell, C., Sharpe, A., Winter, S., Rousik, J., Welty, E., Bush, T., & Norman, B. (2020). *The Public Utility Data Liberation (PUDL) Project* (Version 0.5.0). https://doi.org/10.5281/zenodo.5677623

Tepe, D., Kotthoff, F., Muschner, C., Vogt, E., & Hülk, L. (2023). *Improving data reliability in germany's energy system: A validation of unit locations of the marktstammdatenregister*. https://arxiv.org/abs/2304.10581

Wiese, F., Schlecht, I., Bunke, W.-D., Gerbaulet, C., Hirth, L., Jahn, M., Kunz, F., Lorenz, C., Mühlenpfordt, J., Reimann, J., & others. (2019). Open power system data–frictionless data for electricity system modelling. *Applied Energy*, *236*, 401–409.