Automating Substation Modeling using Labeled Images

Arjun Kumar Madhusoodhanan*, Julian Hoffmann*, Pascal Emanuel Schmidt†, Marco Schott†,
Uwe Kühnapfel*, Veit Hagenmeyer*
* Institute of Automation and Applied Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany
arjun.madhusoodhanan@kit.edu
† Department of Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany

***Abstract*—This paper presents a comprehensive workflow for automating substation modeling using labeled aerial images. Our approach streamlines the process into three key steps: generating the input dataset as labeled images, inferring an electrical substation model based on standard design principles and exemplarily importing the model into DIgSILENT PowerFactory as a simulation environment. We demonstrate that this new automated process significantly enhances speed, improves consistency, and facilitates easier updation of models, outperforming traditional manual methods.**

***Keywords*—Labeled Images, Powerfactory API, Electrical Substation Model, Automation.**

I. Introduction

In the rapidly evolving landscape of renewable energy integration and smart grid technologies, expanding and modernizing the electrical grid has become critical. This task is increasingly relevant as we address challenges such as ensuring grid robustness against disruptive events and adapting to dynamic shifts in political, social, and technological landscapes [1]. Accurate modeling of electrical power grids is a fundamental and well-established practice in energy system analysis and design, playing a key role in navigating these complexities [2], [3]. Comprehensive grid simulations enable the analysis and optimization of various scenarios, such as integrating new energy sources, predicting demand, and managing grid stability. These models are essential for assessing the potential impacts of policy decisions, investment strategies, technological innovations, and possible grid expansions. Despite the critical role of grid modeling, existing manual methods are time-consuming, error-prone, and difficult to maintain, making it challenging to keep models accurate and up-to-date.

A. Motivation

Although Distribution System Operators (DSOs) have access to detailed data of the electrical grid, including the structural and electrical parameters such as network topology, load distribution, voltage levels, and real-time operational metrics, this information is typically not available to researchers [4]. As a result, researchers often rely on generic or black-box models to emulate grid behavior [5], [6]. These models focus on the grid topology and do not include the electrical configurations of the substations such as installed transformers, busbars with single or double configurations, busbar splits, cross-couplings, and switches. Having such detailed grid models will greatly enhance the precision and reliability of electrical grid simulations, especially in the case of adverse events, such as line faults, where the grid topology needs to adjust. In recent years, there has been a significant enhancement in the quality of publicly available aerial imaging, with resolutions reaching down to 2cm per pixel [7]. The majority of substations are of the open-air type, where insulation is primarily achieved through air and by maintaining physical distance between the conducting components. Due to the required insulating distances for substations operating at 380 kV, 220 kV, and 110 kV levels, these substations have large dimensions and follow standardized designs [8], [9]. This characteristic makes it possible to model substations in unprecedented detail by inferring electrical configurations from aerial images, potentially filling the gap left by existing manual and generic grid modeling methods.

B. Contribution

We present a new method for automatically generating electrical grid models from aerial images, particularly suited for large structures like the German 380/220/110 kV grid. Our approach uses labeled images to infer substation layouts, creating detailed electrical models that are then imported into PowerFactory as an exemplary case. This method establishes a verifiable workflow for automating electrical grid modeling, reducing initial modeling time, and keeping models up-to-date.

The remainder of this paper is organized as follows: section II reviews related work, and section III details our methodology. In section IV, we evaluate the proposed method, followed by results and discussion in section V. Finally, section VI concludes our findings and proposes future work.

II. Related works

The challenge of power grid simulation is not new, as electrical grid models are used for multiple applications. Consequently, various software tools for simulating the electrical grid have been developed. Notable software solutions include ETAP, PowerFactory, GridLAB-D, and NEPLAN, among others. An overview of these can be found in Egert et al. [10]. The quality of simulation results depends not only on the simulation environment but also on the level of detail and completeness of the model [11]. Below, we provide an overview of common approaches.

As mentioned in the motivation, the restrictive policies of DSOs to share information regarding critical grid infrastructure necessitate that models be created based on publicly available data. To address the lack of detailed models, transparent planning tools have been developed for electrical grids, emphasizing open-source and open-data methods [4]. Hoffrichter et al. [12] developed a 110 kV model using publicly available OpenStreetMap data [13]. This model integrates into an existing comprehensive European transmission grid model to analyze congestion management but simplifies substation representation by assuming coupled busbars, limiting its accuracy in substation configurations. Heitkoetter et al. [14] explored different methodologies, with and without heuristics, to deduce connections between electrical grid nodes. They compared models from SciGrid [15], GridKit [16], and OpenStreetMap transmission grid models [17], noting that these approaches can only generate simplified abstractions of substations. Riveria et al. [18] adopted a hybrid approach, combining traditional data inferred from topological information with data from the available Common Information Model (CIM); however, their method still lacked detailed substation modeling. Several works have utilized automated approaches to model electrical grids, primarily relying on OpenStreetMap as their primary data source [19]–[22]. In summary, to the best knowledge of the authors, no project or method has created electrical substation models using aerial images at the desired level of detail.

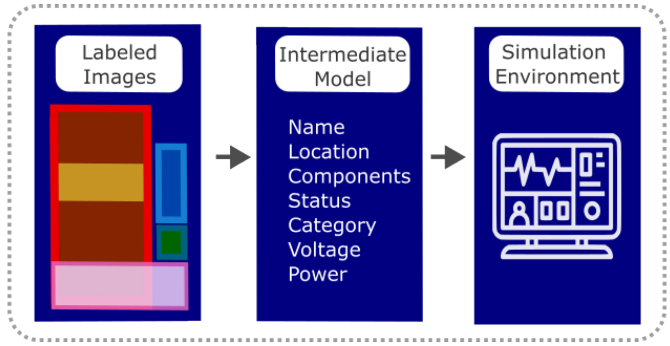

Figure 1: Workflow
From left to right: Labeling substation components in Label Studio, creating an intermediate model, and importing it into the simulation environment (PowerFactory)

III. Methodology

This section describes the methodology used to derive electrical substation models from aerial images. The goal of this work is to develop electrical models of substations and their components, including installed transformers, busbar configurations, busbar splits, cross-couplings, and switches. We have divided the entire process into four independent steps. The steps are processed sequentially and collectively to form the automated workflow, as shown in Figure 1:

1) Manual identification and labeling of substation components in aerial images.
2) Automatic inference of the electrical configuration of the substation based on the labeled dataset and standard design principles.
3) Automatic generation of an electrical model in PowerFactory.

1) Step 1 - Manual identification and labeling of substation components in aerial images: Label Studio [23] is an open-source software project designed to streamline the manual creation of labeled datasets. In this study, we used Label Studio to annotate images of substations obtained from publicly available repositories, namely Bing Maps and Google Maps. The resolution and clarity of these images vary depending on the provider and location, influenced by factors such as geographical location, weather conditions, and viewing angles [7], [24]. To ensure optimal quality, the available images (one from each provider) are displayed in Label Studio, allowing the annotator to select the best image for each substation based on clarity and resolution. Additional information sources, such as links to 3D or street views, are accessible through the user interface to assist in the identification of components.

For image annotation, a set of predefined labels corresponding to the electrical components of the substation is provided to the annotator. Each image is manually marked with these components and assigned the appropriate labels. Figure 2 illustrates an example of such a labeled image of a substation. In our analysis, we identified the busbar system as the central connecting component of a substation. A busbar system is typically built in single, double, or triple configuration, where the busbars run parallel. Cross-couplings connect parallel busbars, facilitating flexibility in grid operations. A busbar can also be divided into segments by busbar splits. These components (parallel busbars, cross-couplings, and splits) allow for the segmentation of the electrical grid. A use case is, for example, to bypass a transmission line when the complete substation needs to undergo maintenance. Electrical components, such as transformers and transmission lines, are aligned along the busbar system and connected to the busbars via bays. Transmission lines and underground cables are labeled by their associated connection bays, which hold the circuit breakers, disconnectors, voltage and current measurement, and earthing devices. The connection bays for overhead lines and underground cables can be differentiated by the presence or absence of visible lines and towers.

2) Step 2 - Automatic inference of the electrical configuration of the substation based on the labeled dataset and standard design principles: In open-air substation designs, sufficient insulation between current-carrying components is ensured by arranging them in a two-dimensional grid-like manner (compare to Figure 2). The design is further simplified to one dimension, with the busbar system serving as the central element, running straight with connection bays for transmission lines and transformers arranged on both sides. Busbar splits, indicated by discontinuities, segment the busbar, while cross-couplings connect busbar segments above them. Therefore the arrangement of components, derived from aerial images, provides information about electrical interconnections.

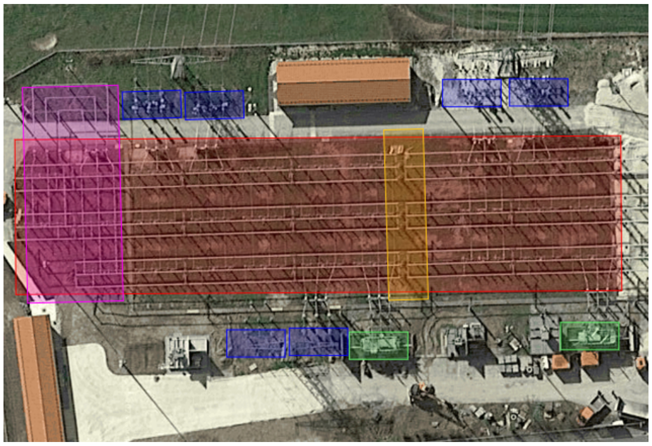

Figure 2: The substation components in Label Studio are marked in different colors: busbar (red), split-coupling (yellow), cross-coupling (pink), transmission line (blue), and transformer (green).

These component positions, defined as rectangular labels, are exported from Label Studio into a JSON file. Initial label positions are normalized using an affine transformation to account for busbar rotation and translation. Busbar segments and cross-couplings are placed along the busbar's length, and connection bays are positioned relative to these segments, maintaining their order and connection side. This information, along with other attributes listed in Table 1, is stored in an intermediate model consisting of Python classes for each electrical component. This intermediate model serves as the data structure for subsequent model generation in a suitable simulation environment.

Attributes	Explanation
Name	Name of the substation
Short name	Short name of the substation
Substation category	Type of substation (generic, industrial, generation)
Busbar category	Busbar type (single, double, triple)
Geographical location	Latitude and longitude of substation
Status	Operational or construction, or wrong location
Covering	Can be open air, partially in-house or fully in-house
Voltage	Voltage on which the substation runs

Table 1: The attributes in the intermediate model for each substation.

The busbar serves as the central element for defining the interconnection of electrical components. However, due to variations in substation designs, certain special cases require alternative approaches. If a substation omits a busbar to reduce costs, a dummy busbar is introduced to standardize the model representation without affecting electrical simulations. Substations may have multiple busbars, especially when they are placed at different angles or locations to accommodate the spatial constraints of the surrounding area. An example of this is illustrated in Figure 3. In such cases, the main busbar is identified as the one closest to the components.

Figure 3: In the substation in Ettringen the busbar is split into multiple sections (redbox)

3) Step 3 - Automatic substation generation in PowerFactory: PowerFactory is a leading software for power system analysis, known for its ability to evaluate generation, transmission, distribution, and industrial systems [25]. It is widely used by DSOs for preliminary studies, grid planning, and operations due to its robust scripting and interface capabilities, which support automated grid modeling. PowerFactory features a logical representation of electrical components and interactive graphical diagrams [26]. The logical representation connects all grid components for simulations and analysis, while the interactive diagrams provide user interaction with a subset or abstraction of these components. Since each graphical component references a logical component, generating the logical representation of the grid is a prerequisite.

The logical model includes terminals (such as busbars and junctions) and bays (for components like transformers and transmission lines). Each bay features detailed switch connections to its components. Cubicles link components to terminals and additional bays represent couplings between terminals. The intermediate model is parsed and a simulation model created by using the simulation environments scripting API. Transformers and generic loads are added to the substation for inclusion in the detailed diagram generation. Additionally, a naming scheme for components is applied, and a default datatype for transformers is created. For the graphical representation, a simplified overview diagram features busbars, transformers, loads, and key circuit breakers, each depicted with distinct symbols. An example of this representation, using the Bobingen substation in Germany, is depicted in Figure 4. The space between the two transformers accommodates an additional bay, which can be used for connecting overhead lines or cables in subsequent modeling steps.

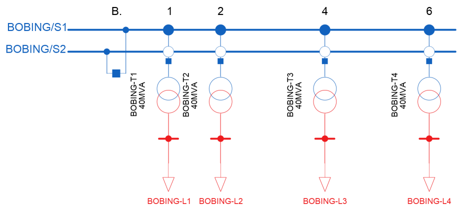

Figure 4: Graphical representation of Bobingen substation

Metric	Value
Number of substations	127
Number of revisions	166
Number of annotations	1689
Substations skipped	0
Renamed substations	1
Missing attributes	1
Sectioned substations	2
Import time (model)	0,0625s
Import time (PowerFactory logic)	8.375s
Import time (PowerFactory graphic)	3.25s
Import time (whole process)	11.688s
Size of the grid	841mm × 1189mm

Table 2: Test summary of substation import to Powerfactory

IV. Evaluation

Before implementing the automated workflow, a manual process was used to model the substations within the simulation environment, based on labeled images. The performance of the automated workflow is evaluated by comparing it to the manual workflow using the same labeled dataset. This dataset includes 110 kV substations operated in the control zone D76 in Germany. Performance indicators for the comparison include time efficiency, accuracy, and data variation. Testing was conducted on a Windows PC with 16GB RAM and a 12th Gen Intel Core i7-1255U processor, utilizing Python 3.11 and PowerFactory 2023 SP1 with a research license.

The results, shown in Table 2, cover 127 substations, each with 1-2 revisions. Revisions refer to annotation sets created on different dates, and attributes were sourced from these various revisions, successfully condensing the annotations into attributes. Two substations were flagged as ambiguous due to multiple definitions of distinct busbars in different locations. Other errors, such as empty annotations and duplicate components, were identified and excluded from further processing. This exclusion did not impact the final model, as only redundant information was discarded.

V. Results and Discussion

The new automated approach to generate electrical models of substations with the relative position of components in labeled aerial images shows significant time savings compared to established manual workflow. The manual approach involves customizing templates and placing them on the overview grid, which takes about 10 minutes per substation. With about 4,800 substations in the German 110 kV grid, this results in an estimated total of 800 hours of work. In contrast, the automated method creates each component (transformers, connection bays, busbars, splits, cross-couplings, and dummy loads), applies a standardized naming scheme, connects the components, and generates a graphical representation of the substation. The processing time depends on the number of substations and components. For example, importing and placing 127 stations takes just 11.7 seconds. Extrapolating this, processing the 4,800 substations of the German 110 kV grid would require approximately eight minutes, representing a substantial reduction from the estimated 800 hours required for manual model generation. The automatic approach also minimizes human error in model creation, such as errors in the naming conventions of substation components and the risk of overlooking components in images and failing to include them in the model. It handles minor errors automatically and flags significant issues for user review, although complex busbar layouts still require manual inspection. This limitation pertains only to errors that would make a substation semantically incorrect. For example, in the case of sectioned busbars (see Figure 3), the automated methods can still identify the main busbar, which connects to a larger number of components and generates the corresponding substation layouts. However, due to the complexity of the design, these cases are flagged for manual review. This affects only a small percentage of substations (2 out of 127 in the sample dataset).

A limitation of using aerial images is that substations located partially or entirely indoors, such as Gas Insulated Substations (GIS), cannot be identified. While this issue was rare in the rural test dataset, it could be significant in urban areas like Frankfurt in Germany, where in-house substations are more common. This limitation affects both manual and automated workflows based on aerial imagery. Another advantage of the automated approach is the ease of maintaining the model. As power grids evolve, due to factors such as the integration of renewable energies and the electrification of transport and heating, it is essential to keep grid models up-to-date. The automated workflow allows for easy updates of the labeling and direct modifications to the electrical model, eliminating the need for manual component identification.

VI. Conclusion and Future Work

This paper introduces a new automated method for importing labeled images into electrical substation models in Powerfactory. This approach can be used to import models into any similar simulation environment with scripting capabilities, such as ETAP, GridLAB-D, NEPLAN, and others. This new method significantly reduces the time required compared to the manual model creation. By implementing error prevention mechanisms, we enhance the reliability and accuracy of the resulting models. The method can automatically infer most electrical substation configurations using only labeled aerial images. Instances with uncertainty are flagged for manual review, ensuring robustness and reliability. The primary constraint is the correctness of the manually created input dataset. As aerial imaging quality improves, this methodology promises even greater precision and speed for power grid modeling and analysis. Additionally, this method creates a transparent workflow for model creation, allowing for easier evaluation and critique by model reviewers.

We propose three key improvements to enhance the approach. First, we recommend automating the import of overhead lines and cables using publicly available map data, such as the 'Digitales Basis-Landschaftsmodell' [27], to reduce the time spent connecting substations. Second, since image labeling is currently done manually, automating this process could significantly speed up and improve the model's reliability. This could be accomplished by using image processing algorithms, such as machine learning. Recent developments in machine learning have highlighted the importance of large training sets, for which manual labeling is already an essential step. Finally, integrating power plants from databases like the Marktstammdatenregister [28] and incorporating electrical load estimates for each substation would enhance the model's accuracy and comprehensiveness through continuous updates.

VII. Acknowledgements

This work is funded by the Helmholtz Platform for the Design of Robust Energy Systems and their Supply Chains (RESUR).

References

[1] Helmholtz Energy System Design, Helmholtz platform for the design of robust energy systems and their supply chains (RESUR), Available: https://www.esd.kit.edu/422.php.

[2] C. Singh, Chanan, P. Jirutitijaroen, J. Mitra, Electric Power Grid Reliability Evaluation: Models and Methods, 978-1-119-48627-5, John Wiley & Sons, December 2018.

[3] F. Calero, C. A. Cañizares, K. Bhattacharya, C. Anierobi, I. Calero, M. F. Z. De Souza, M. Farrokhabadi, N. S. Guzman, W. Mendieta, D. Peralta, B. V. Solanki, N. Padmanabhan, and W. Violante, A review of modeling and applications of energy storage systems in power grids, *Proceedings of the IEEE*, vol. 111, no. 7, pp. 806–831, Jul. 2023, Available: https://ieeexplore.ieee.org/document/9743285/.

[4] U. P. Mueller, L. Wienholt, D. Kleinhans, I. Cussmann, W.-D. Bunke, G. Pleßmann, and J. Wendiggensen, "The eGo grid model: An open source approach towards a model of German high and extra-high voltage power grids," *Journal of Physics: Conference Series*, vol. 977, no. 1, pp. 012003, Feb. 2018. Available: https://dx.doi.org/10.1088/1742-6596/977/1/012003.

[5] D. Hewes, I. Boiarchuk, and R. Witzmann, Estimation of reactive power compensation in the European transmission system, in *NEIS Conference 2016*, D. Schulz, Ed. Wiesbaden: Springer Fachmedien Wiesbaden, ISBN: 978-3-658-15029-7 pp. 15–20, 2017.

[6] P. Villella, PEGASE pan-European test-beds for testing of algorithms on very large scale power systems, in *Proc. 2012 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, Oct. 2012, Available: https://ieeexplore.ieee.org/document/6465783.

[7] J. Schwartz, Bing Maps Tile System - Bing Maps, Microsoft Learn, Jun. 2022, Available: https://learn.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system.

[8] VDE Verlag, VDE-AR-N 4121:2018-04 - Planungsgrundsätze für 110-kV-Netze, VDE-AR-N 4121, VDE Verlag, 2018, Available: https://www.vde.com/de/fnn/arbeitsgebiete/netzbetrieb-sicherheit/netzplanung/planungsgrundsaetze-hochspannung-vde-ar-n-4121.

[9] VDE Verlag, VDE-AR-N 4141-2:2022-09 - Technische Regeln für den Betrieb und die Planung von elektrischen Netzen, VDE-AR-N 4141-2, VDE Verlag, 2022, Available: https://www.vde-verlag.de/normen/0100700/vde-ar-n-4141-2-anwendungsregel-2022-09.html.

[10] R. Egert, A. Tundis, and M. Mühlhäuser, On the simulation of smart grid environments, in *Proc. 2019 Summer Simulation Conf. (SummerSim '19)*, San Diego, CA, USA, pp. 1–12, Jul. 2019, Available: https://dl.acm.org/doi/abs/10.5555/3374138.3374155.

[11] U. Kühnapfel and V. Hagenmeyer, On the Art of Electric Power System Modelling and Simulation for Integrated Transmission-Distribution Analysis, in *Modelling the Energy Transition. Cultures – Visions – Narratives*, R. M. Erdbeer, V. Hagenmeyer, and K. Stierstorfer, Eds. London: Palgrave Macmillan, 2024, ISBN 978-3-031-69030-3, in print.

[12] A. Hoffrichter, H. Barrios, J. Massmann, B. Venkataramanachar, and A. Schnettler, Impact of considering 110 kV grid structures on the congestion management in the German transmission grid, *Journal of Physics: Conference Series*, vol. 977, no. 1, pp. 012004, Feb. 2018, Available: https://dx.doi.org/10.1088/1742-6596/977/1/012004.

[13] OpenStreetMap Contributors, OpenStreetMap, 2004. Available: https://www.openstreetmap.org/.

[14] W. Heitkoetter, W. Medjroubi, T. Vogt, and C. Agert, Comparison of Open Source Power Grid Models—Combining a Mathematical, Visual and Electrical Analysis in an Open Source Tool, *Energies*, vol. 12, no. 24, p. 4728, Jan. 2019, Available: https://www.mdpi.com/1996-1073/12/24/4728.

[15] W. Medjroubi and C. Matke, SciGRID—an open source reference model for the European Transmission Network (v0.2), presented in Nov. 2016, Available: https://doi.org/10.13140/RG.2.2.14596.12161.

[16] B. Wiegmans, GridKit: GridKit 1.0 'for Scientists', v1.0, Zenodo, Mar. 2016, Available: https://doi.org/10.5281/zenodo.47263.

[17] M. Scharf and A. Nebel, osmTGmod: Open Source German Transmission Grid Model Based on OpenStreetMap v0.1.3, 2017, Available: https://github.com/wupperinst/osmTGmod.

[18] J. Rivera, J. Leimhofer, and H.-A. Jacobsen, OpenGridMap: towards automatic power grid simulation model generation from crowdsourced data, *Computer Science - Research and Development*, vol. 32, no. 1, pp. 13–23, Mar. 2017, Available: https://doi.org/10.1007/s00450-016-0317-4.

[19] J. Hörsch, F. Hofmann, D. Schlachtberger, and T. Brown, PyPSA-Eur: An open optimisation model of the European transmission system, *Energy Strategy Reviews*, vol. 22, pp. 207–215, 2018, Available: https://doi.org/10.1016/j.esr.2018.08.012.

[20] M. Weber, L. Janecke, H. K. Çakmak, and V. Hagenmeyer, Open Data-Driven Automation of Residential Distribution Grid Modeling With Minimal Data Requirements, *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2024, Available: https://ieeexplore.ieee.org/document/10540632.

[21] H. K. Çakmak, L. Janecke, M. Weber, and V. Hagenmeyer, An Optimization-based Approach for Automated Generation of Residential Low-Voltage Grid Models Using Open Data and Open Source Software, in *Proc. 2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 2022, pp. 1–6, Available: https://ieeexplore.ieee.org/document/9960483.

[22] D. Sarajlić and C. Rehtanz, Low voltage benchmark distribution network models based on publicly available data, in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 1–5, Sep. 2019, Available:https://ieeexplore.ieee.org/document/8905726.

[23] M. Tkachenko, M. Malyuk, A. Holmanyuk, and N. Liubimov, Label Studio: Data Labeling Software, GitHub, 2020. Available: https://github.com/heartexlabs/label-studio.

[24] Google, Google Maps, Google Maps, 2024. Available: https://www.google.com/maps.

[25] DIgSILENT GmbH, *PowerFactory*, Available: https://www.digsilent.de/.

[26] DIgSILENT, PowerFactory 2023 User Manual, 2023.

[27] Bundesamt für Kartographie und Geodäsie, Digitales Basis-Landschaftsmodell, Jan. 2024, Available: https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/basis-dlm.pdf.

[28] Federal Network Agency, Marktstammdatenregister, Available: https://www.marktstammdatenregister.de/MaStR/.