Adapting the CO_2-Compass Architecture to Further Optimize Data Generation Methods

Enhancing CO_2 Emission Forecasts by Minimizing the Area of Observation

Lucas Hüer, Hans-Jürgen Pfisterer
Faculty of Engineering & Computer Science
University of Applied Sciences Osnabrück
Osnabrück, Germany
Email: {l.hueer | j.pfisterer}@hs-osnabrueck.de

Helge Fischer, Sebastian Lawrenz
Institute for Software & Systems Engineering
Technische Universität Clausthal
Clausthal-Zellerfeld, Germany
Email: {helge.fischer | sebastian.lawrenz}@tu-clausthal.de

Oliver Thomas
Chair of Information Management & Information Systems
University of Osnabrück
Osnabrück, Germany
Email: oliver.thomas@uni-osnabrueck.de

***Abstract*—Climate change is one of the most important social issues of recent years. Every new scientific insights and political decisions make it clear that innovative ways of attacking the climate change are needed. Minimizing emissions are especially important in order to stop the greenhouse effect and thus a source for global warming. Therefore, the software of the CO_2-Compass was developed to provide a transparent overview of the electricity production and related CO_2 emissions. Containing the service of a 24 hour forecast of these data, the CO_2-Compass serves as a control tool to decide when electrical devices should be used from an ecological point of view. This paper strives to improve the existing architecture of the tool, by adding new sources of data collection and thus optimize the outcome of all offered services by the CO_2-Compass. Therefore, the main goal of this paper is to improve the existing architecture of the first monomythical prototype towards a flexible and expandable microservice based architecture.**

Keywords-CO_2 Emission Reduction; Software Engineering; Energy Production; Renewable Energies; Energy Transition; Data Warehouse; Microservices.

I. Introduction

In the wake of an undergoing energy transition, several parts of the German energy industry are continuously changing. This mainly implies a shift from centralized to decentralized energy production, as well as a shift from conventional towards renewable energy sources [1]. Moving towards an increased utilization of renewable energy sources has multiple advantages, such as security of supply due to unlimited sources or higher sustainability levels due to decreased CO_2 emissions [2] [3]. However, to better facilitate all advantages that go hand in hand with using renewable energies, it is necessary to align the electricity usage of electrical devices with a more resource-saving energy production [4] [5]. Thus, electrical devices can be used at times in which the regenerative part of created energy is high and CO_2 emissions are low. To determine the best time of a day for this scenario, the software tool "CO_2-Compass" was created [6]. By using this tool companies, as well as households can get a transparent overview of the electricity production and CO_2 emissions that go along with it. Further, a 24-hour-forecast will be provided to show the likely development of electricity production and belonging CO_2 emissions. This source of information supports a user's decision when to use energy-intensive hardware (like heat pumps, air conditioners, charging stations or production machines) by determining the point of time at which CO_2 emissions of the energy production are lowest. In a first version of the CO_2-Compass, data collection is limited to the aggregated information that is made available by the four major power grid operators in Germany (50Hertz, Amprion, TenneT, Transnet BW) [7]. Decentralized power supply information and detailed data broke down to local energy providers are not yet included. However, to have a reliable knowledge base about power generation and its CO_2 emissions, it is necessary to have as much local information as possible, which in turn would optimize the data in terms of relevance, reliability and accuracy. A growing share of renewables and a fluctuating, decentralized power production will require flexible and open interfaces in order to process the accruing data. In order to tackle this problem and further optimize the data generation of the CO_2-Compass, this paper is tackling the following research question:

What kind of changes can be made to the existing CO_2-Compass architecture to improve its data generation in terms of data quality while keeping it expandable?

To answer this question thoroughly, following structure will guide the reader through this article: Once a scientific and political background is given in Section 2, there will be a description of the current CO_2-Compass architecture and its limitations in Section 3. Those limitations will then serve as an explanation for the refactoring of the existing architecture. All changes that are necessary to optimize the data generation of the CO_2-Compass will then be described in Section 4, which is followed by an elaboration on the creation of timelines in terms of gathering data in Section 5. Once all changes are explained, a discussion and conclusion will complement this paper.

II. Scientific and Political Background

The 21st century has so far been largely shaped by scientific and political discussions and decisions relating to climate change [8]. Government representatives from most countries meet regularly and reach agreements on various actions to protect the environment and society from negative consequences of the climate change. One of the biggest factors that has been addressed in previous climate summits is the emission of greenhouse gases [9]. A main objective of the Paris Agreement is to limit global warming to 1.5°C [10].

In order to curb the emission of these gases and thus to combat the greenhouse effect that primarily leads to global warming, the global community is setting ever more ambitious goals. The German government has, for instance, decided to reduce greenhouse gas emissions by 40% between 1990 and 2020 [11]; goals for the following decades are even more ambitious. It is therefore necessary to develop innovative products and services that support companies and consumers in reducing CO_2 emissions. A promising field in which CO_2 emissions can be vastly minimized is the electrical power generation [12]. The saving potential can be seen when looking at the development of direct CO_2 emissions per kilowatt hour of electricity related to the German electricity mix. Since 1990 the direct CO_2 emissions per kilowatt hour of electricity (in g/kWh) was reduced from 764 g/kWh to 523 g/kWh in 2016. This is equivalent to a reduction of 31% [12]. This reduction is explained by Petra Icha as follows: "If the proportion of an energy source with a high CO_2 emission factor, such as brown or hard coal, falls in favor of an energy source with a lower CO_2 emission factor, such as a renewable energy sources […] the emission factor of the electricity mix also decreases" [12]. In other words, different energy sources are associated with different levels of CO_2 emissions. Therefore, improved transparency in terms of the energy mix and its forecasted development over the next hours is needed, to base decisions on electricity usage of electrical devices on the decrease of CO_2 emissions. Based on our knowledge, there are some associations that transparently show the electricity mix and the associated CO_2 emissions. Good examples of this are Agora Energiewende, electricityMap or KlimAktiv Consulting. However, there are some drawbacks when using these services and thus the CO2-Compass was developed in order to offer new solutions in the field. The CO_2-Compass has two major advantages over potential competitors: On the one hand, there is the possibility of displaying the CO_2 emissions per individual transmission network provider, which leads to more relevant information for individual households and companies. On the other hand, there is a REST interface with which the software can be coupled to any hardware in the form of devices and machines. However, in order to add new functions, interfaces and solutions or to optimize all existing ones, the CO_2-Compass is under continuous improvement. One of these improvements is the change of its architecture to increase the data quality.

III. Existing Architecture of the CO2-Compass and it's Limitations

Initially, the CO_2-Compass software was developed on basis of the SCRUM method. By setting up, processing and completing functional and non-functional system requirements, the software was created incrementally. The system requirements were based on an interview with an expert in the field of electrical engineering. However, to optimize the CO_2-Compass, further software-development methods were used. The decision to use as an agile, user centred approach aimed to generate a first prototype.

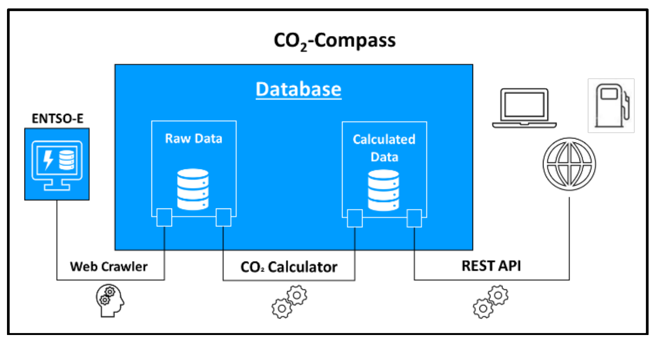

Figure 1 – Initial Architecture of the CO_2-Compass

The initial architecture of the CO_2-Compass consists of three main system-parts (see Figure 1): The Crawler, the CO_2-Calculator and the REST API. Within a first step, the raw data from the European Network of Transmission System Operators for Electricity (ENTSO-E) is collected by a self-developed crawler at five-minute intervals, before being stored in a separate database. The electricity production data (in MW) for every high-voltage power grid operator in Germany (50Hertz, Amprion, TenneT, Transnet BW) and for Germany as a whole are now stored in this database. There are currently 18 different types of energy sources for producing electricity. After the crawler updates and collects the raw data from the transmission networks, the specific CO_2 values for the associated production figures are calculated by the CO_2-Calculator. These values are also stored in the database and can thus be assigned to the network operator and the type of production. This calculation takes place for each individual provider in five-minute intervals and provides

ISBN: 978-1-61208-771-9

values every 15 minutes. Furthermore, the CO_2-Calculator creates a CO_2 forecast for each production type for the next 24 hours every day at midnight, which is also stored in the internal database. An integrated REST API enables public access to the generated data and the forecast values. This interface can be used in a variety of ways to couple the emission information with electrical devices. One example is the connection of the CO_2-Compass to intelligent charging stations, which continuously query the current electricity mix (including the associated emission values). In combination with the predicted values from the forecast, charging stations can thus be switched in such a way that they only start charging when there is a low-emission electricity mix. However, a fast-charging option instead of the eco-friendly charging is still possible and customers can use an approach which suits them best.

The current implementation of the CO_2-Compass is based and limited on an input interface which collects data from ENTSO-E. Information about power generation and CO_2 emissions are published through their transparency platform at the abstract level of control areas. They hereby fulfill an EU Regulation, where data must be published by control areas. By definition, these control areas are “a coherent part of the interconnected system, operated by a single system operator” [13]. These areas usually reflect the high-voltage grid of a whole country (e.g., in France or Portugal) or supra-regional network operators (e.g., in Germany the four control areas of the in Section 1 mentioned transmission system operators). Since TenneT’s geographical coverage, to name an example, ranges from the Danish border to the Austrian border (see Figure 2 and [7]), a precise indication of the CO_2 emissions of locally consumed power is hardly possible.

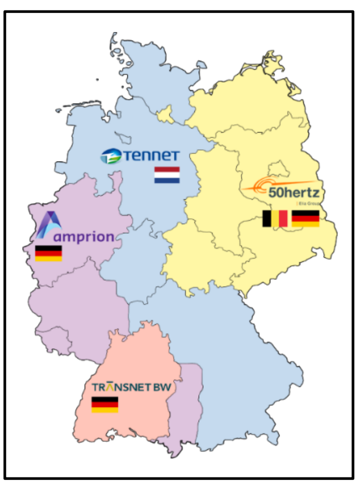

Figure 2 – Control Areas in Germany

One reason for this is the so called “copper plate illusion” (translated from the German term: Illusion einer Kupferplatte), which describes the problem of assuming that electricity producers and consumers act without restrictions, and can thus generate and consume electricity as they please without transmission bottlenecks or energy loss [14] [15]. However, for a lossless transmission of electricity, the power grid would have to be a superconducting (copper) plate - hence the name - which is not the case. Therefore, an extension of the input interfaces in order to have more specific, locally relevant data is needed. Those changes and their relevance will be presented and explained in the coming sections.

IV. PROPOSED ARCHITECTURE

The actual state of the architecture (see Figure 1) was explained in the previous section. An agile procedure was chosen to prove the functionality of the CO_2-Compass at short term. However, after a successful establishment of the prototype, the next step will be a change in the architecture to a more modular system, within the goal of “low coupling, high cohesion”. This paradigm means that modules or services inside the software architecture should be as self-contained and independent as possible, and thus depending as less as possible on other components [16]. Based on this, following objectives for the architecture have been defined:

1. Modular design and expandable splitting of the monolith and change the architecture to micro-services [17]
2. A clear division of components into layers to define security and sovereignty of data
3. Data-driven design for analytic services

The proposed architecture is shown in Figure 3 consisting of separated layers. On the left hand side there are various data sources, such as IoT-Devices and (smart) meters to measure the current power production and calculate related CO_2 emissions at specific points. In addition to the ENTSO-E database and other sources of forecast (such as local weather and solar radiation data) should be imported by a (web)-crawler. Based on this information local power generation which is not metered online (e.g., solar panels on rooftops) can be forecasted and taken into account. For all own and third parties’ generators in a local grid with measuring devices, the available data should be transferred via a clear defined interface into the databases. Now, based on these extended data collection, additional substitute values can be generated and processed together with metered raw data in the CO_2-Calculator. Data integration into the data warehouse takes place in the **Data Integration Layer.** This layer manages the incoming data flows, filters and cleans it (if necessary) and transfers it into a data warehouse. Services, like a broker or the web crawler, receive and structure the data beforehand. The following **Data Warehouse Layer** stores and manages the data for all services and handles incoming raw data. The raw data are divided into at least two

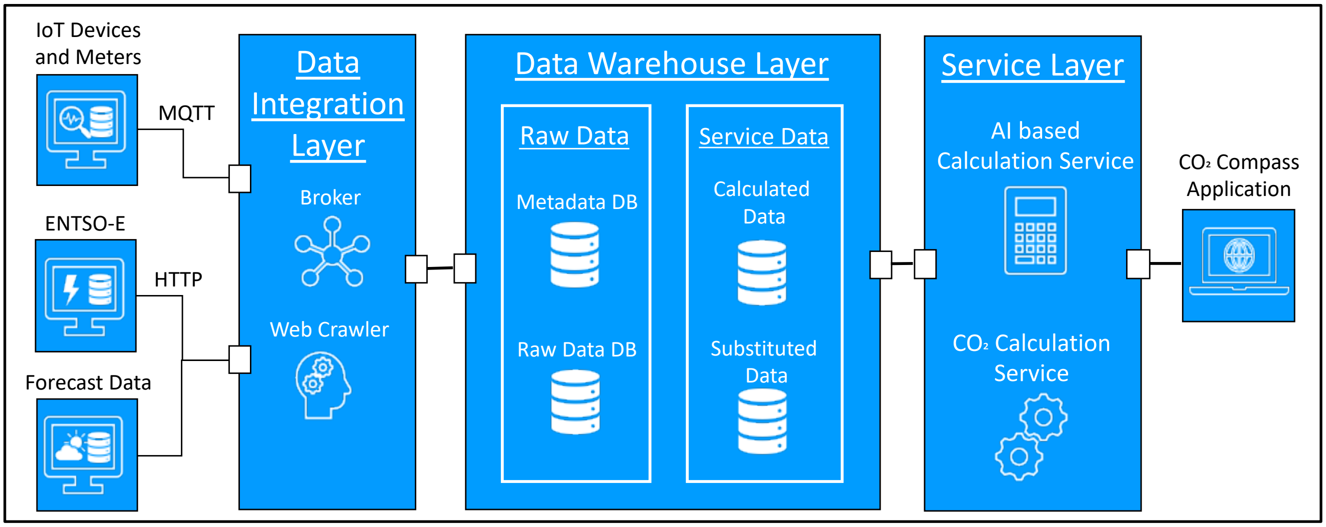

Figure 3 – Enhanced Architecture of the CO_2-Compass

different databases. The first one is the *Metadata DB*, which contains meta data and information about the data itself, such as the origin of this data. The *Raw Data DB* contains the values themselves (or based on the DIKW-Pyramide so called data [18]) that means for example the measured electric power (*12.5* megawatts per hour at *2019-05-08 12:35:00*).

However, even if the *Metadata DB* and the *Raw Data DB* are independent at first sight, they have a close relation. This comes out of the fact that every column with specific values in the Raw Data DB has a meta description in the Metadata DB (see Table 1).

TABLE I. SMALL EXTRACT FROM TWO RAW DATA COLUMNS

Power Volumes	Timestamp
12.5	2019-05-08 12:35:00
15.5	2019-05-08 12:37:00
24.4	2019-05-08 12:39:00

The associated metadata in the *Metadata DB* contains, for example, the link to the corresponding columns, the *unit* (here megawatts per hour and datetime in the format YYYY-MM-DD HH-MM-SS), information about the sensor (time interval between two measuring points [here 2 Minutes]) and the *description* of the data source (e.g., metering device of a wind park).

Moreover, as shown in Figure 3 the service data are separated from the raw data. The service data (here shown as *Calculated Data* and *Substituted Data*) are the data sources which belong to a specific service inside the Service Layer. This separation ensures to have a clear separation on the one hand and on the other hand to improve the data access depending on the intended use. This ensures that the generated data (which might be wrong when the service or the AI fails) does not get mixed up with the raw data.

The **Service Layer** contains the modules and services for the application of the CO_2-Compass itself. One service for example calculates the CO_2 component of the current power mix (*CO_2-Calculation Service)* and proposes services to generate missing data (indicated here as *AI based Calculation Services*) to increase the accuracy (see also Section 5). Via a REST Interface the user can then interact with the frontend of the *CO_2-Compass Application*.

In summary, it can be stated that a clearly separated but extendable architecture was introduced, which allows to extend the use cases of the CO_2-Compass and to further develop mechanism to control and reduce the CO_2 emissions. By introducing different layers, a high level of data security and a clean separation of incoming data is made possible. Thus, it can be differentiated easily between public non-critical data sources like ENTSO-E and private data sources like smart meters. Furthermore, the *Data Integration Layer* will be designed similarly dynamic and expandable as the *Service Layer*. This enables an uncomplicated extension of the data warehouse with further data sources.

V. Creation of Timelines

To fulfill the requirements of more detailed information at local level, more timelines have to be implemented. Local data, as well as artificial data will give the necessary added value to have an in-depth view on a distinct power grid (hereafter simplified called "distribution grid"). The local data may be generated by measuring the power production like windmills or Combined Heat and Powerplants (CHPs) in the distribution grid. These metering devices have to be connected "online" and submit the data continuously. Missing values (see crosshatched bars in Figure 4) can be substituted by replacement values and discrepancies to the defined time patterns (e.g., hourly values instead of expected 15 minutes) may be aligned immediately in the service layer.

Estimating and generating artificial data opens the opportunity to gain a holistic view as well as compensate poor data quality from certain metering devices or other input sources. If not every power production (especially the small ones) or battery storages in cars or households are metered online a precise allocation of CO_2 emissions is hardly possible. Even, for example, if a field of solar panels is equipped with metering devices, but poorly connected online,

 ISBN: 978-1-61208-771-9

it may make sense to calculate artificial data and short-term prognoses based on solar radiation forecasts. It would also open opportunities to simplify the technical requirements and perform a cost-benefit consideration concerning metering in the context of an organically growing distribution grid driven by the energy transition (‘Energiewende’).

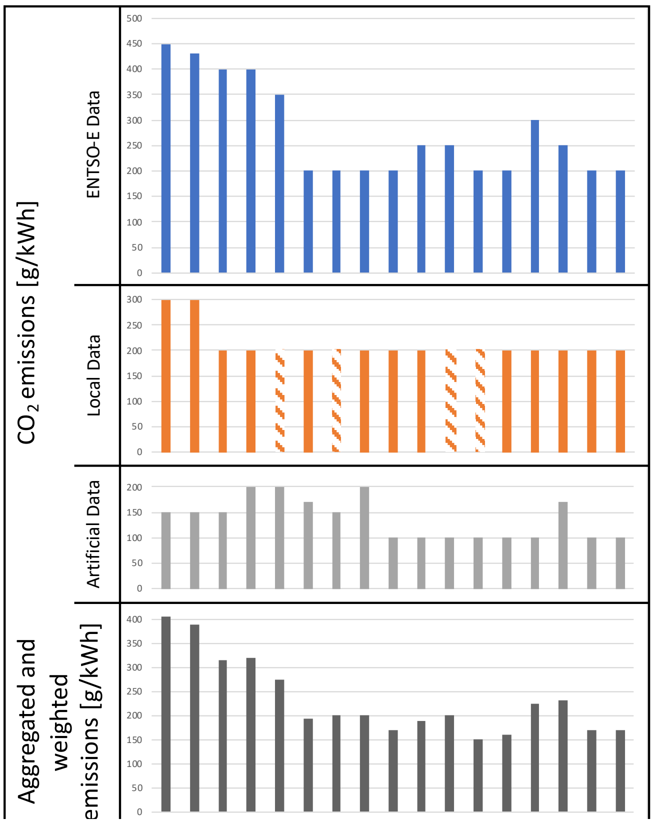

Figure 4 – Timelines with CO_2 emissions

An AI based calculation service (see Figure 5) can be used to generate the artificial data. It will base on AI methods like the recurrent network algorithm Long Short-Term Memory (LSTM) [19] and considers former results, non-online metered devices which are submitted with delay, as well as relevant input variables (e.g., solar radiation, wind, energy prices or time of day) in order to train a neuronal network.

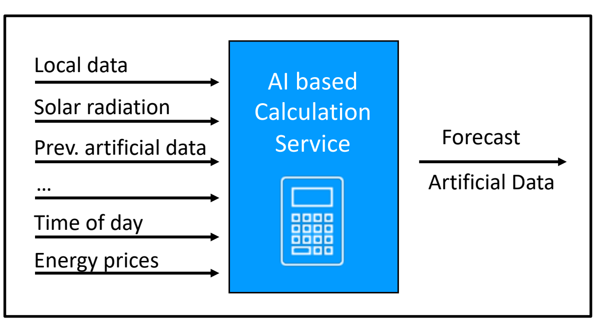

Figure 5 – Generator for Artificial Data

The output will be an estimation of produced power volumes and their dedicated CO_2 emissions reflecting the characteristics of grid topologies and individual influencing factors in the local energy mix. Finally, the three data timelines containing specific CO_2 emissions have to be weighted with allocated power volumes at the connection from the upstream grid, metered local power generators and estimated miscellaneous devices. As a result, this aggregation will now give a more precise, average CO_2 emissions factor for the distribution grid in comparison to the rough proxy based on the ENTSO-E data.

VI. Conclusion

Scientific warnings, as well as political decisions with regard to global problems arising from the climate change call for innovative solutions to minimize CO_2 emissions. One of these solutions was developed in 2019 and is since being continuously improved to optimize its functions: the so-called CO_2-Compass.

Initially, the architecture of the CO_2-Compass was focused on basic data collection in order to transparently present all CO_2 emissions from the power mix based only on the data of ENTSO-E and thus from the four big power grid operators in Germany: 50Hertz, Amprion, TenneT, and Transnet BW. However, as explained in Sections 3 and 4, the data quality can be improved significantly when additionally taking local data sources, as well as artificial data sources into account. While local data can be gained out of decentral energy sources, such as solar panels or wind parks, the generation of artificial data is based on calculations of artificial intelligence algorithms. The artificial data thus opens the opportunity to gain a holistic view by closing gaps in which no local data collection is possible due to pure connection for instance. These new data sources improve the total data quality in terms of accuracy and completeness and thus enable the possibility to generate more precise information.

In total, a new data-driven architecture with a modular design has been developed, which can still be extended in case of future adaptions and include more and new data sources. Via a new separation into three layers based on microservices, a high level of data security and a clean separation of incoming data is made possible. First, the Data Integration Layer manages all incoming data flows, filters and cleans it (if necessary) before transferring it into a data warehouse. This is where the second layer, the Data Warehouse Layer, stores and manages all relevant data in different databases. In the final Service Layer, all transferred and stored data is then used for creating solutions for different customers.

Future research will elaborate even further on the CO_2-Compass, its optimization potential and especially the testing in different environments. On the one hand, it is planned that the software will then act as part of a product-service system, by creating and utilizing interfaces with different kinds of electrical devices, such as charging stations or heat-pumps.

 ISBN: 978-1-61208-771-9

On the other hand, there will be additional research related to trigger IoT devices and to control conventional power generators via results of the CO_2-Compass with the aim of minimizing the CO_2 emissions to meet individual emission reduction paths.

ACKNOWLEDGMENT

This research was conducted in the scope of the research project SmartHybrid – Electrical Engineering (ID: ZW 685003732), which is partly funded by the European Regional Development Fund (ERDF) and the State of Lower Saxony (Investitions- und Förderbank Niedersachsen – NBank). We would like to thank them for their support.

REFERENCES

[1] M. Guidolin and R. Guseo, "The German energy transition: Modeling competition and substitution between nuclear power and Renewable Energy Technologies.," *Renewable and Sustainable Energy Reviews,* no. 60, pp. 1498–1504, 2016.

[2] G. Resch, A. Held, T. Faber et al., "Potentials and prospects for renewable energies at global scale.," *Energy policy*, vol. 36, no. 11, pp. 4048–4056, 2008.

[3] S. Shafiei and R. A. Salim, "Non-renewable and renewable energy consumption and CO2 emissions in OECD countries: A comparative analysis.," *Energy policy*, vol. 66, pp. 547–556, 2014.

[4] P. Finn, C. Fitzpatrick, D. Connolly et al., "Facilitation of renewableelectricity using price based appliance control in Ireland's electricity market," *Energy*, vol. 36, no. 5, 2952_2960, 2011.

[5] A. Pina, C. Silva, and P. Ferrão, "The impact of demand side management strategies in the penetration of renewable electricity," *Energy*, vol. 41, no. 1, pp. 128–137, 2012.

[6] L. Hüer, N. Stadie, S. Hagen et al., "Der CO2-Kompass: Konzeption und Entwicklung eines Tools zur emissionsarmen Stromnutzung," *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik-Informatik für Gesellschaft*, 2019.

[7] K. Heuck, K. D. Dettmann, and D. Schulz, eds., *Überblick über die geschichtliche Entwicklung der elektrischen Energieversorgung*, Springer Vieweg, Wiesbaden, 2013.

[8] G. P. Brasseur, D. Jacob, and S. Schuck-Zöller, *Klimawandel in Deutschland: Entwicklung, Folgen, Risiken und Perspektiven.*, Springer, 2017.

[9] M. Prys-Hansen, M. Lellmann, and M. Röseler, "Die Bedeutung der Klimafinanzierung für den Pariser Klimagipfel 2015," 2015.

[10] J. Rogelj, M. den Elzen, N. Höhne et al., "Paris Agreement climate proposals need a boost to keep warming well below 2 °C," *Nature*, vol. 534, no. 7609, pp. 631–639, 2016.

[11] B. Schlomann and W. Eichhammer, eds., *Energieverbrauch und CO2-Emissionen industrieller Prozesstechnologien: Einsparpotenziale, Hemmnisse und Instrumente.*, Fraunhofer-Verlag, 2013.

[12] P. Icha, "Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990 - 2018," *Umweltbundesamt*, 2019.

[13] European Parliament and Council, *Commission Regulation (EU) No 543/2013 of 14 June 2013 on submission and publication of data in electricity markets and amending Annex I to Regulation (EC) No 714/2009*, 2013.

[14] M. Kahles, "Überprüfung der einheitlichen deutschen Stromgebotszone nach der Elektrizitätsbinnenmarkt-Verordnung," *Würzburger Berichte zum Umweltenergierecht,* no. 44, 2019.

[15] J. Aengenvoort and H. Sämisch, "Die Illusion einer Kupferplatte," https://www.next-kraftwerke.de/energie-blog/kupferplatte-stromnetz.

[16] S. Newman, "Building microservices: designing fine-grained systems," *O'Reilly Media Inc.*, 2015.

[17] S. Newman, "Monolith to Microservices: Evolutionary Patterns to Transform Your Monolith," *O'Reilly Media*, 2019.

[18] J. Rowley, "The wisdom hierarchy: Representations of the DIKW hierarchy.," *Journal of Information Science*, vol. 33, no. 2, pp. 163–180, 2007.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation (Journal)*, vol. 9, no. 8, pp. 1735–1780, 1997.