Data Quality Management Framework for Smart Grid Systems

Preprint · June 2019

Stanislav Chren
Masaryk University

Bruno Rossi
Masaryk University

Tomáš Pitner
Masaryk University

Data Quality Management Framework for Smart Grid Systems

Mouzhi Ge, Stanislav Chren, Bruno Rossi, and Tomas Pitner

Faculty of Informatics, Masaryk University, 602 00 Brno, Czech Republic
{mouzhi.ge,chren,brossi,tomp}@muni.cz,

Abstract. New devices in smart grid such as smart meters and sensors have emerged to become a massive and complex network, where a large volume of data is flowing to the smart grid systems. Those data can be real-time, fast-moving, and originated from a vast variety of terminal devices. However, the big smart grid data also bring various data quality problems, which may cause the delayed, inaccurate analysis of results, even fatal errors in the smart grid system. This paper, therefore, identifies a comprehensive taxonomy of typical data quality problems in the smart grid. Based on the adaptation of established data quality research and frameworks, this paper proposes a new data quality management framework that classifies the typical data quality problems into related data quality dimensions, contexts, as well as countermeasures. Based on this framework, this paper not only provides a systematic overview of data quality in the smart grid domain, but also offers practical guidance to improve data quality in smart grids such as which data quality dimensions are critical and which data quality problems can be addressed in which context.

Key words: smart grid, data quality, data quality problem, smart meter

1 Introduction

Smart grids are developed to optimize the generation, consumption, and management of energy via intelligent information and communication technology. Its research involves smart meters, user-end smart appliances, renewable energy resources, digitalization in electricity supply networks, as well as new technologies to detect and react to the changes in electricity supply networks. As [1] stated, a smart grid reflects a combination between Information and Communication Technologies (ICT) and Internet of Things (IoT), whereby data services such as aggregation of sensor data and analysis of voltage consumption from smart meters [28] offer a foundation for the concept of smartness. Since the quality of data can directly affect the output of data services, the security, quality, reliability, and availability of an electric power supply depends on the quality of data in the power system [19]. Thus, data quality has been considered as a prominent

Pre-print of the paper Ge, M., Chren, S., Rossi, B., Pitner, T. (2019). Data Quality Management Framework for Smart Grid Systems, to appear in the 22nd International Conference on Business Information Systems (BIS2019).

issue in smart grids [8]. In a broader context, data quality has become a critical concern to the success of organizations [15]. Numerous business initiatives have been delayed or even canceled, citing poor-quality data as the main reason [16]. Therefore, data quality management can be regarded as an indispensable component in smart grid applications.

The current data quality problems in smart grid are addressed still in an ad-hoc style. For example, Chen et al. [8] focused on the outlier detection of electricity consumption data. Their solution tackles a specific quality aspect of electricity consumption data. However, this will obstruct practitioners to foresee the other data quality problems and delay the reaction on time for potential data quality problems. Thus, it is valuable to obtain a big picture of the different data quality problems in the smart grid network. Also, some of the data quality problems in smart grids may be interconnected. One data quality problem may be caused by another data quality problem. For example, an outlier in the electricity consumption data may be caused by missing data items or data attacks. Therefore, focusing on specific quality aspects can mislead the root causes of the data quality problems. Based on our review, there is a lack of a systematic framework for managing data quality in smart grids. Also, data quality is critical in the smart grid domain, as invoices of end users depend for example on the collected power consumption data.

In this paper, we propose a systematic data quality management framework for smart grids. It can not only profile a variety of data quality problems in the smart grid context, but also show how to categorize and organize the data quality problems based on data quality dimensions. In this framework, different data quality problems are identified and assigned to the related dimensions. It can therefore indicate which data quality dimensions are critical in the smart grid data quality improvement. Furthermore, the data quality problems assigned in the same dimension may need to considered together.

The remainder of the paper is organized as follows. Section 2 reviews the general data quality management and the state-of-the-art data quality research in Smart Grid. Section 3 identifies and summarizes a comprehensive set of the possible data quality problems in smart grid. Based on the identified data quality problems, Section 4 proposes a framework to categorize the data quality problems into established data quality dimensions. Finally, section 5 concludes the paper and outlines the future research.

2 Data Quality and Smart Grid Research

Since the data quality problems are usually domain-specific, the importance of data quality dimensions may vary in different application domains. For example, Ge et al. [17] conducted a study to rank the overall importance of different data quality dimensions used in a variety of data quality studies. They further emphasized that prioritizing the importance of data quality can determine the focus of data quality improvement and management. Therefore, to find out which dimensions are important in the smart grid domain, assigning the data quality problems to dimensions can be used to facilitate the data quality measurement process.

There exists some research work that tends to classify the data quality problems in smart grids. for example, Chen et al. [8] proposed that the data quality issues in electricity consumption data can be divided into noise data, incomplete data and outlier data. Noise data refer to the data with logical errors or the data violating certain rules or specifications. These data that can in turn affect data analysis results. Incomplete data mean the missing values in the data sources, and outlier data are the data that deviate from standard data variation ranges. However, the scope of data in smart grids is broader than electricity consumption data. For example, [24] specified various types in smart grids such as sensor data, battery status data, and device downtime data. Therefore, the classification of the data quality problems in smart grids can be extended to a larger scale.

The smart grid domain encounters the Big Data Quality problems. Due to the massive number of smart meters, various sensors and other customer facilities, the smart grid network has been generating Big Data [11]. Zhang et al. [35] further described the big data characteristics in smart grids such as large amount of meter and sensor data (volume), real-time data exchange (velocity), and extensive data sources in a smart grid (variety). They further stated that data quality is a critical issue in processing the big smart grid data, where the data quality management is usually positioned in the data preprocessing phase. Zhang et al. [35] classified the big data quality problems by using three countermeasures, which are data integration, data cleansing, and data transformation. While data integration deals with the entity resolutions and data redundancy, data cleansing can be used to alleviate missing and abnormal data. Finally, data transformation serves to provide high-quality data formats for data analytics such as correcting data distribution and constructing new data attributes. This classification is especially designed for smart grid data analytics. In this paper, we will outline big and normal data quality problems in smart grids.

3 Data Quality in Smart Grid

When we discuss the term "data" in the context of smart grids, we cannot ignore the overall complexity of the infrastructure and the communication needs [10]. Due to the complexity, data in smart grids comes from a variety of sources, and can be structured, unstructured, but very often a mixture of both, making the analysis more complex [35].

3.1 Smart Grid Infrastructure

The smart grid infrastructure comprises several parts, each of them with different responsibilities regarding the energy and data transfer. The smart grid architecture depends on the standards used. According to the NIST standard, the smart grid has a hierarchical structure that includes the following domains:

Wide Area Network (WAN) is responsible for communication between power generation plants, substations and transformer equipment. Neighbourhood Area Network (NAN) serves as a bridge between customer premises and the substations. This level focuses on the collection of data from the smart meters, which are aggregated by the data concentrator and further transferred to the data centers [10]. Furthermore, the customer premises network (CPN) consists of networks at the customer location. Depending on the type of customer, we can distinguish between Home Area Networks (HAN), Industrial Area Network (IAN) and Business Area Network (BAN). This layer enables communication between the smart meters, intelligent appliances and their connectivity to NAN. An overview of communication technologies in smart grids is shown in Fig. 1.

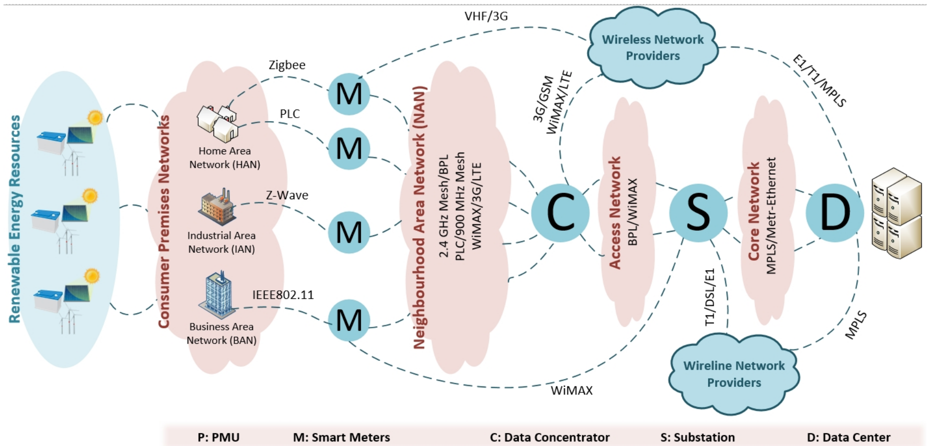

Fig. 1. Communication infrastructure in smart grids (from Al-Omar et al. [2])

3.2 Data in Smart Grid

Generally, smart grid data can be classified in three categories: measurement data (e.g., smart meters data), business data (e.g., customer data), and external data (e.g., weather data) [35]. In this context, we focus our analysis on *measurement data*, that is the type of data that can, more than other types, characterize SG data analysis needs. We focus in particular on data derived from two devices: Smart Meters [10] and PMUs [25, 5].

In the smart grid infrastructure, there are two main components which produce the measurement data essential for the grid operation: smart meters and phasor measurement units (PMUs). *Smart meters* are devices which serve as replacements of traditional power meters installed at customer premises (e.g. households, industrial buildings, etc...). They record data about customer's power consumption (and possibly production if the customer utilizes renewable power sources). Smart meters enable two-way communication and a power distributor is also able use them to remotely control appliances such as water heaters. This becomes useful in various load management programs to balance the power flow in the grid. Besides the measurement data, smart meters are also able to report various events, for example meter failures, unexpected manipulation with the device or occurrence of over/under-voltage states [31]. On the other hand, *PMUs* are devices which measure phasor information in the power distribution, such as voltage and current. The PMUs collect the measurement data from many points in the power grid at very high frequency (up to 120 samples per second). The data are time synchronized based on the GPS radio clock. Measurement data are transmitted to various monitoring systems using them to analyze the current state of the power grid to discover potential stability issues.

There is a number of systems in smart grids that ensure reliability of the power supply and the availability of critical services and which rely on high quality data collected from smart meters or PMUs [9]: (1) Blackout Prevention Systems protect the grid from instabilities and failures. They cover the whole power grid, using the data from PMUs to obtain relevant information from the grid. (2) Supervisory Control and Data Acquisition Systems (SCADA) are one of the core systems of a Smart Grid that provide monitoring and support to operation activities and functions in transmission automation, dispatch centers and control rooms. In a SCADA system, a remote terminal unit collects data from smart meters or devices in a substation and delivers the data to a central Energy Management System. (3) Flexible Alternating Current Transmission Systems are responsible for reliable and secure transmission of power. They allow dynamic voltage control, increased transmission capability and capacity, and support fast restore of the grid after failure. (4) Feeder Automation Systems are responsible for the operation of medium-voltage networks including fault detection.

3.3 Data Quality Problems

Issues in data collected by smart grid devices are usually referred by literature as either bad data [29], corrupted [6], or missing data [6]. However, such definitions do not capture the diversified facets of smart grid data issues that are more refined in terms of specific issues. For example, Shishido and Solutions [32] discuss issues in smart meter data quality during the consumption data collection process. The main issues reported are duplicate items from the meter readings, zero record periods, and large spikes over periods of time. There are some issues in Smart Meters/PMU data that are peculiar of the smart grid context: non-trustful data points, data aggregation issues due to privacy concerns, timing issues with skewed timestamps of recorded events. We summarize the main Data Quality Problems in Table 1, as a series of issues that are derived from literature on smart meters and PMU data analysis.

Duplicate records from multiple devices (DQP1) mean that the same record is stored multiple times in the same way or with different values, causing duplication in the data [32]. A suggested strategy for the identification is cross-linking records across different devices to look for possible duplicated values, as well to search for repeating sequences [32].

Table 1. Data Quality Problems in the Smart Grid context.

DQ Problem	**Description**	**Context**	**Countermeasures**
DQP1. Duplicate data	Duplicate records from Smart Meter reading, can be caused by upgrading of Smart Meters (e.g., same reading from the old and new SM) [32].	SM	Cross-linking data from multiple devices and examining repeating sequences [32]
DQP2. Missing/ incomplete data	Some data can be expected to be available (e.g. regular smart meter reading) but due to some reasons (e.g. technical failure) they are not [32].	PMU/SM	Linear interpolation (short periods), creation of daily load profiles for historical patterns recreation (longer periods) [22, 29]
DQP3. Zero Records Semantics	Detecting differences between data that was not transmitted/recorded by sensors and stand-by periods. All lead to difficulties in interpreting zero-valued ranges [32, 29].	SM	Creation of daily load profiles [22, 29], reasonability tests for allowed ranges and comparison of values from other devices [34]
DQP4. Data Outliers (out-of-range)	Large bursts (spikes), or low values compared to the average over a period of time [32, 20]	PMU/SM	Reasonability tests for allowed ranges [34], application of anomaly detection algorithms, context-, collective-based [31, 20].
DQP5. Measurement Errors	Datapoints that represent measurement errors due to hardware failures, signal interference, etc... [29]	PMU/SM	Reasonability tests for allowed ranges and comparison of values from other devices [34], signal analysis of smart meters for outliers detection [30]
DQP6. Non-trustful data-points	Datapoints that were manipulated intentionally (e.g. data injection attacks: alter the measurements of SMs to manipulate the operations of the smart grid [23, 7, 27])	PMU/SM	Using historical data, statistical-based detection [23]
DQP7. Data anonimization	Aggregation of attributes/features for privacy preservation / anonimization can lead to issues for data analysis [12]	SM	Preserving data integrity for smart grid data aggregation, e.g. by hashing/signature checking against data tampering [26], Smart Metering data de-pseudonymization [21]
DQP8. Timing issues	Timing in which an event is recorded by PMUs / Smart Meters is not precise, causing difficulties in the integration of data, or in case of PMUs, wrong computations [34, 13]	PMU/SM	Comparing values recorded by different systems, e.g. PMU and SCADA [34]

Missing/incomplete data (DQP2) represents the case in which data recordings are missing for some periods of time, making this a problem of data imputation research [22]. Strategies in such cases go in the direction of linear interpolation for short periods of missing records, or the creation of daily load profiles for historical patterns recreation in case of longer periods [22, 29].

Zero record periods (DQP3) constitute a distinct case from the aforementioned missing/incomplete data scenario [32]. In this case, data is present but with zero recordings, making difficult to understand if such records were not recorded/transmitted, or missing values were due to some stand-by period [29]. There are different strategies that can be applied to understand the semantic of zero-record periods of time, like the creation of daily load profiles from smart meter data [22, 29], or reasonability tests for allowed ranges and comparison of values from other devices for PMUs [34].

Data outliers or out-of-range values (DQP4) represent large bursts of data spikes or low values compared to the average values over periods of time [32]. Detection of these value ranges is part of the anomaly detection area, determining outliers based on context-, or collective-based algorithms [20, 31]. For PMUs, reasonability tests for allowed ranges are important [34]

Measurement errors (DQP5) can represent a relevant issue for both smart meters and PMUs [29]. There can be many sources of such issues in smart grids data. According to Chen et al. [6], measurement errors can derive from smart meter problems, communication failures, equipment outages, lost data, interruption/shut-down in electricity use, but also components degradation and operational issues [3]. To address measurement errors, reasonability tests for allowed ranges and comparison of values from other devices can be used in the context of PMUs [34], while signal analysis of smart meters for outliers detection can be applied to smart meter data [30].

Non-trustful data points (DQP6) derive from potential cyber-attacks to the smart grid infrastructure. Such attacks do not only involve authentication issues, but also false data injection attacks to provide fake data-points as they were real recorded events [23, 7]. The non-trustful data-points injected/modified by means of cyber-physical attacks, are meant to manipulate the overall operations in the SG by leading operators into false beliefs about the current state of the infrastructure [27].

Data aggregation issues due to privacy concerns (DQP7) come from the needs to preserve the privacy of data collected from smart meter readings. Some features collected by the different types of devices might be obscured or aggregated into other features, making the analysis process more difficult. Over the last years, many techniques have been developed to preserve the statistical properties of aggregated features [12], preserve data integrity from tampering [26], and algorithms that attempt at data de-pseudonymization [21]. While the removal of some features might be seen as a way to anonimize data from smart meters, this is however ineffective, as customers can be re-identified by other features [4].

Timing issues with skewed timestamps of recorded events (DQP8) can be an issue in both smart meters and PMUs. While in smart meters such errors might just involve issues in later data attribution between different devices [13], for PMUs such issues can involve subsequent wrong computations [34]. Comparing data from different devices can be a strategy to detect and correct timing issues, such as comparing timestamps from PMUs and SCADA systems [34].

Furthermore, there are two main peculiarities in smart grid data cleansing activities: (1) data are mostly generated from sensors, hardware-devices: root causes can be found in hardware failures, communication related problems [14]; (2) data cleansing in the data mining domain usually assume structural data, while in the smart grid context, mostly time-series approaches are needed for the identification of patterns/anomalies.

4 Data Quality Management Framework in Smart Grid

To propose a data quality management framework for smart grids, we have adopted the general data quality framework from Wang and Strong [33]. Thus, intrinsic, contextual, representational and accessibility are adopted to categorize the data quality concept. Further, the relations between data quality categories and data quality dimensions are also adopted from Wang and Strong [33]. However, since not all the data quality dimensions are important for smart grids, we have used the identified data quality problems to select the data quality dimensions in our framework. Therefore, our framework is intended to be domain-specific for smart grids. Based on the typical data quality problems that we revisited in the smart grid context, we derive the data quality framework in smart grid that classifies these data quality problems into dimensions, categories, as well as into contexts. This data quality framework is divided into five layers. The first layer from the top is the overall data quality in smart grid. This layer is usually used as one step in the whole big data analytics process e.g. before or after the data integration. The second layer divides the smart grid data quality into four quality aspects. Under each data quality aspect, it is the dimension layer that indicates which data quality dimensions are related to which smart grid data quality problems. Therefore, the third layer and fourth layer are data quality dimensions and specific problems. As data quality problems are derived from different contexts, the context in smart grids is the final layer.

It can be seen that there are seven data quality dimensions that are particularly important for the smart grid domain. These seven data quality dimensions are accuracy, consistency, timeliness, completeness, believability, accessibility, and interpretability. Accuracy is mainly defined as the data points falling into a normal range or interval. Thus, data outliers in smart grids belong to this dimension and detected data outliers can be considered as inaccuracy. The consistency dimension is used when there are different sources of smart grid data. In the smart grid domain, data can be generated from different devices. On one hand, this creates the data redundancy, on the other hand, the cross-reference approach can be used to validate the data consistency. Since time series analysis is usually used in smart grids, timing issues like wrong timestamps may cause problems to construct the time series data. Therefore, the timeliness dimension

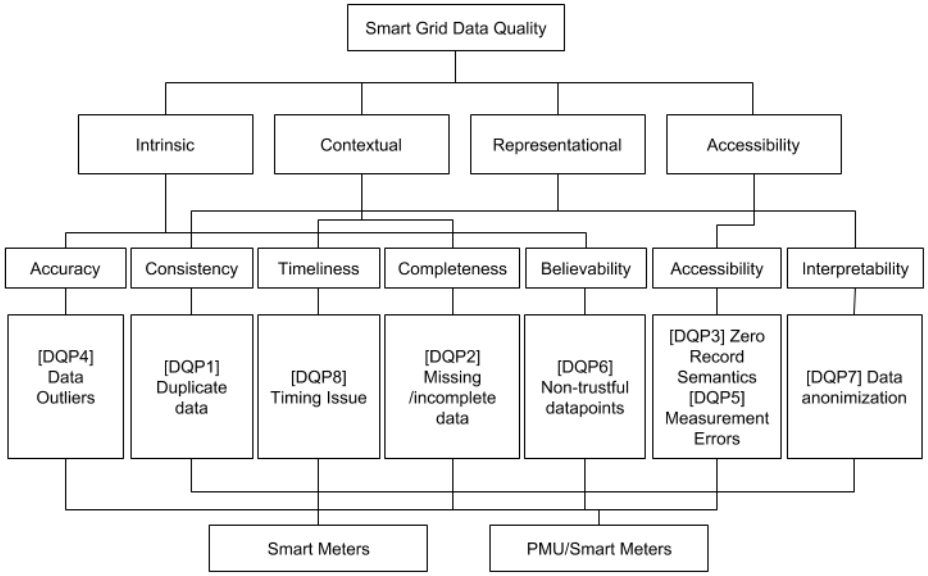

Fig. 2. Data quality framework in smart grid

is to control if the data are recorded in a precise time or time interval. Completeness dimension involves the problems of missing and incomplete smart grid data. As there is a large number of devices in smart grids, the data completeness issue can be regularly caused by device malfunction. Although the believability dimension is not well discussed in other domains, trustful data are important in smart grids because the data manipulation in smart grid can be directly related to economical benefits. The accessibility dimension is related to the hardware and infrastructures in smart grid. Therefore, accessibility can be used to measure if the data can be accessed. Finally, interpretability is defined as how well the data can be interpreted. This can be balanced between the data privacy and the analytic details. Overall, not all the data quality dimensions from previous research are critical for the smart grids domain.

Our framework is proposed in an operational and measurable level. For example, each data quality dimension can be measured by its related data quality problems. Likewise, the data quality categories such intrinsic or contextual data quality can be further measured by aggregating the related dimensions. Since most of the general data quality management frameworks are not domain-specific, their model granularity is refined only to the dimension level: it is therefore difficult to apply other data quality management frameworks and measure data quality in a domain. Our framework tackles this problem and relates the data quality measurement to specific quality problems.

Our framework can be further integrated into other data quality management frameworks. Since most of the existing data quality models or frameworks are based on the data quality dimensions [18], our framework is centralized by data quality dimensions and can be easily integrated into other frameworks or models by replacing the dimensions from this framework. Furthermore, our proposed framework can locate the root cause of low data quality dimensions by concrete quality problems in smart grids. After the assessment, the contexts and countermeasures can then be used for data quality improvement. For example, to determine which data quality problems occurred in which context and to determine the countermeasures to improve some data quality dimensions.

5 Conclusions

In this paper, we have proposed a systematic and practical taxonomy of data quality problems in smart grids. We have then proposed a new data quality management framework that adapted the data quality aspects and refined them to seven critical data quality dimensions for smart grids. Thus, the data quality assessment and improvement in smart grids can be more focused on the derived dimensions. Each data quality dimension is connected to concrete smart grid data quality problems. On one hand, the framework enables the data quality measurement for data quality dimensions. On the other hand, since the data quality problems are linked to specific smart grid contexts, it can facilitate to identify the root causes of low quality data and establish a data quality improvement plan. Compared to other general data quality frameworks, our framework is designed to be domain-specific and limited to the smart grid. The framework contributes towards automatically controlling the data quality in smart grid. As future work, we plan to further extend the framework by automating the data quality measurement processes.

6 Acknowledgements

The research was supported from ERDF/ESF "CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence" (No. CZ.02.1.01/0.0/0.0-/16_019/0000822).

References

[1] Abo, R., Even, A.: Managing the quality of smart grid data research in progress. In: IEEE Int. Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies. pp. 5–8 (2016)
[2] Al-Omar, B., Al-Ali, A., Ahmed, R., Landolsi, T.: Role of information and communication technologies in the smart grid. Journal of Emerging Trends in Computing and Information Sciences 3(5), 707–716 (2012)
[3] Alahakoon, D., Yu, X.: Smart electricity meter data intelligence for future energy systems: A survey. IEEE Transactions on Industrial Informatics 12(1), 425–436 (2016)

[4] Buchmann, E., Böhm, K., Burghardt, T., Kessler, S.: Re-identification of smart meter data. Personal and ubiquitous computing 17(4), 653–662 (2013)

[5] Burnett, R.O., Butts, M.M., Sterlina, P.S.: Power system applications for phasor measurement units. IEEE Computer Applications in Power 7(1), 8–13 (1994)

[6] Chen, J., Li, W., Lau, A., Cao, J., Wang, K.: Automated load curve data cleansing in power systems. IEEE Transactions on Smart Grid 1(2), 213–221 (2010)

[7] Chen, P.Y., Cheng, S.M., Chen, K.C.: Smart attacks in smart grid communication networks. IEEE Communications Magazine 50(8) (2012)

[8] Chen, W., Zhou, K., Yang, S., Wu, C.: Data quality of electricity consumption data in a smart grid environment. Renewable and Sustainable Energy Reviews 75, 98 – 105 (2017)

[9] Chren, S., Rossi, B., Buhnova, B., Pitner, T.: Reliability data for smart grids: Where the real data can be found. In: 2018 Smart City Symposium Prague. pp. 1–6 (2018)

[10] Chren, S., Rossi, B., Pitner, T.: Smart grids deployments within eu projects: The role of smart meters. In: Smart Cities Symposium Prague, 2016. pp. 1–5. IEEE (2016)

[11] Daki, H., El Hannani, A., Aqqal, A., Haidine, A., Dahbi, A.: Big data management in smart grid: concepts, requirements and implementation. Journal of Big Data 4(1), 13 (2017)

[12] Efthymiou, C., Kalogridis, G.: Smart grid privacy via anonymization of smart metering data. In: Smart Grid Communications, First IEEE International Conference on. pp. 238–243. IEEE (2010)

[13] Eichinger, F., Pathmaperuma, D., Vogt, H., Müller, E.: Data analysis challenges in the future energy domain. Computational Intelligent Data Analysis for Sustainable Development pp. 181–242 (2013)

[14] Gao, J., Xiao, Y., Liu, J., Liang, W., Chen, C.P.: A survey of communication/networking in smart grids. Future Generation Computer Systems 28(2), 391–404 (2012)

[15] Ge, M., Helfert, M.: A framework to assess decision quality using information quality dimensions. In: Proceedings of the 11th International Conference on Information Quality, MIT, USA. pp. 455–466 (2006)

[16] Ge, M., Helfert, M.: Effects of information quality on inventory management. International Journal of Information Quality 2(2), 177–191 (2008)

[17] Ge, M., Helfert, M., Jannach, D.: Information quality assessment: validating measurement dimensions and processes. In: 19th European Conference on Information Systems, Helsinki, Finland. p. 75 (2011)

[18] Ge, M., O'Brien, T., Helfert, M.: Predicting data quality success - the bullwhip effect in data quality. In: Perspectives in Business Informatics Research - 16th International Conference, Copenhagen, Denmark. pp. 157–165 (2017)

[19] Gellings, C.W., Samotyj, M., Howe, B.: The future's smart delivery system. IEEE Power and Energy Magazine 2(5), 40–48 (2004)

[20] Jakkula, V., Cook, D.: Outlier detection in smart environment structured power datasets. In: Sixth International Conference on Intelligent Environments. pp. 29–33. IEEE (2010)

[21] Jawurek, M., Johns, M., Rieck, K.: Smart metering de-pseudonymization. In: 27th Annual Computer Security Applications Conf. pp. 227–236 (2011)

[22] Kim, M., Park, S., Lee, J., Joo, Y., Choi, J.K.: Learning-based adaptive imputation methodwith knn algorithm for missing power data. Energies 10(10), 1668 (2017)

[23] Kosut, O., Jia, L., Thomas, R.J., Tong, L.: Malicious data attacks on the smart grid. IEEE Transactions on Smart Grid 2(4), 645–658 (2011)

[24] Leonardi A, Ziekow H, S.M.K.P.: Dealing with data quality in smart home environments lessons learned from a smart grid pilot. Journal of Sensor and Actuator Networks 5(1) (2016)

[25] Li, F., Qiao, W., Sun, H., Wan, H., Wang, J., Xia, Y., Xu, Z., Zhang, P.: Smart transmission grid: Vision and framework. IEEE transactions on Smart Grid 1(2), 168–177 (2010)

[26] Li, F., Luo, B.: Preserving data integrity for smart grid data aggregation. In: Smart grid communications, 2012 IEEE third international conference on. pp. 366–371. IEEE (2012)

[27] Liu, Y., Ning, P., Reiter, M.K.: False data injection attacks against state estimation in electric power grids. ACM Transactions on Information and System Security 14(1), 13 (2011)

[28] Matta, N., Rahim-Amoud, R., Merghem-Boulahia, L., Jrad, A.: Putting sensor data to the service of the smart grid: From the substation to the ami. Journal of Network and Systems Management 26(1), 108–126 (2018)

[29] Peppanen, J., Zhang, X., Grijalva, S., Reno, M.J.: Handling bad or missing smart meter data through advanced data imputation. In: IEEE Innovative Smart Grid Technologies Conference. pp. 1–5. IEEE (2016)

[30] Rao, R., Akella, S., Guley, G.: Power line carrier (plc) signal analysis of smart meters for outlier detection. In: Smart Grid Communications, IEEE International Conference on. pp. 291–296. IEEE (2011)

[31] Rossi, B., Chren, S., Buhnova, B., Pitner, T.: Anomaly detection in smart grid data: An experience report. In: Systems, Man, and Cybernetics, IEEE International Conference on. pp. 2313–2318. IEEE (2016)

[32] Shishido, J., Solutions, E.U.: Smart meter data quality insights. In: ACEEE Summer Study on Energy Efficiency in Buildings (2012)

[33] Wang, R.Y., Strong, D.M.: Beyond accuracy: What data quality means to data consumers. Journal of Management Info. Sys. 12(4), 5–33 (1996)

[34] Zhang, Q., Luo, X., Bertagnolli, D., Maslennikov, S., Nubile, B.: Pmu data validation at iso new england. In: Power and Energy Society General Meeting (PES), 2013 IEEE. pp. 1–5. IEEE (2013)

[35] Zhang, Y., Huang, T., Bompard, E.F.: Big data analytics in smart grids: a review. Energy Informatics 1(1), 8 (2018)