Design of Overall Framework of Self-Service Big Data Governance for Power Grid

Lin Qiao[1], Qiaoni Zhou[1], Chunhe Song[2,3](✉), Hao Wu[1], Biqi Liu[1], and Shimao Yu[2,3]

[1] State Grid Liaoning Electric Power Co., Ltd., Shenyang 110000, People's Republic of China
[2] Key Laboratory of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, People's Republic of China
songchunhe@sia.cn
[3] Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, China

Abstract. At present, power grid companies have not formed a complete data quality control system and a comprehensive and effective data quality assurance mechanism, which restricts the deep mining of data value. In this paper, based on the full-service unified data center of power Grid Company, a general framework of self-service power grid big data governance is presented. Firstly, the related work of big data governance is reviewed; secondly, the architecture and characteristics of the grid company's full-service unified data center are analyzed; thirdly, the related requirements of self-service grid big data governance are analyzed; fourthly, the overall framework of self-service grid big data governance is proposed. Compared with other general big data governance framework, the proposed framework considers the specification of power grid as well as the features of self-service, making the framework more feasible.

Keywords: Data governance · Power grid · Self-service · Framework design

1 Introduction

In recent years, with the global energy problem becoming more and more serious, the research work of smart grid has been carried out all over the world. The ultimate goal of smart grid is to build a panoramic real-time system covering the whole production process of power system, including power generation, transmission, substation, distribution, power consumption and dispatch [1]. The basis of supporting the safe, self-healing, green, strong and reliable operation of smart grid is the panoramic real-time data acquisition, transmission and storage of power grid, and the rapid analysis of accumulated massive multi-source data [2]. With the development of smart grid construction, the amount of data generated by grid operation and equipment inspection/monitoring increases exponentially, which gradually constitutes the big data concerned by today's information academia, which needs corresponding storage and fast processing technology as support [3].

X. B. Zhai et al. (Eds.): MLICOM 2019, LNICST 294, pp. 222–234, 2019.
https://doi.org/10.1007/978-3-030-32388-2_19

With the continuous construction and deepening application of information technology in power supply enterprises, various businesses of power supply enterprises have been preliminarily integrated with information technology [4]. The number and types of business data in information systems are gradually increasing, and the need for data sharing is urgent. The data quality and data sharing utilization layer are not high [5]. First, the support degree of data for analysis and decision-making is low, and there are many sources and different statistical calibers for the same data. Second, the support degree of data for operation and management needs to be improved, the data quality is uneven, some data do not have business system support, lack of uniform norms, standards and clear data accountability; third, the number of front-line personnel. Data input is a huge workload, data duplicate input, business functions duplicate. Fourth, data quality control lags behind, management and control work is one-sided, there is no complete data quality control system and a comprehensive and effective data quality assurance mechanism, restricting the deep mining of data value. Therefore, it is necessary to focus on the enterprise data life cycle, closely integrate the requirements of the company to promote the innovation of management system and working mechanism, based on the current situation of the construction and application of information support system in operation monitoring center, draw lessons from the company data management experience, realize the whole process quality management of power supply enterprise data, consolidate the data base, improve the data quality, ensure the accuracy of data, which provides a strong guarantee for data integration and mining applications [6].

In order to solve the above problems, based on the full-service unified data center of power Grid Company, this paper puts forward the general framework of self-service power grid big data governance. In this paper, firstly, the related work of big data governance is reviewed; secondly, the architecture and characteristics of the grid company's full-service unified data center are analyzed; thirdly, the related requirements of self-service grid big data governance are analyzed; fourthly, the overall framework of self-service grid big data governance is proposed. The last part is the conclusion of this paper.

2 Related Works

In the theoretical field of data governance, many organizations have made pioneering contributions, especially ISO38500, DAMA, DGI, IBM DG Council, ISACA and Gartner [7]. Their main work is to analyze, summarize and refine data governance elements such as principles, scope and contributing factors, and to establish a self-contained data governance framework on this basis. ISO 38500 proposes an IT governance framework (including goals, principles and models) and considers that the framework is also applicable to the field of data governance [8]. In terms of objectives, ISO38500 considers that the goal of IT governance is to promote the efficient and rational use of IT by organizations. In terms of principles, ISO38500 defines six basic principles of IT governance: responsibilities, strategies, procurement, performance, conformity and personnel behavior. This principle describes the recommended behavior guiding decision-making. Each principle describes the measures to be taken, but does not explain how and what to do [9]. In terms of the model, ISO38500 believes that the leaders of an organization should focus on three core tasks: first, to assess the current and future IT utilization; second, to guide the policies and plans for the implementation of governance readiness; and third, to establish a cycle model of "evaluation, guidance and supervision" [10]. DAMA summarizes three main functions of data governance, including data governance, data architecture management, data development, data operation management, data security management, reference data and master data management, data warehouse and business intelligence management, document and content management, metadata management and data management, and puts data governance at the core [11]. Then it elaborates data governance in detail. Seven major environmental elements, namely goals and principles, activities, major deliverables, roles and responsibilities, technologies, practices and methods, have finally established the corresponding relationship between the ten functions and the seven major environmental elements. It is believed that the key point of data governance is to solve the matching between the ten functions and the seven major environmental elements. DAMA believes that data governance is the exercise of power and control over data asset management, including planning, monitoring and implementation. It also distinguishes between data governance and IT governance: IT governance targets IT investment, IT application portfolio and IT project portfolio, while data governance targets data; IT is like water pipe, data is like water in water pipe, water governance methods are obviously different from water pipe governance methods [12]. DGI believes that different IT governance should establish an independent data governance theory system. DGI summarizes ten key points of data governance from three levels of organization, rules and process, and proposes a DGI data governance framework. DGI data governance framework shows the logical relationship among ten basic components in the form of intuitive access paths, and forms a self-contained and complete system from method to implementation. Components are divided into three groups according to their functions: rules and collaborative work norms, personnel and organizational structure, and processes. Rules and collaborative work norms, i.e. establishing, coordinating and standardizing data governance rules (including policies, requirements, standards, responsibilities, controls and data definitions) to guide different departments to work together to formulate and implement rules of collaborative work norms [13]. It includes the following six components: mission and vision; objectives, measures of governance effectiveness, financial strategies; data rules and definitions; decision-making power; division of responsibilities and control. Personnel and organizational structure, that is, the organizational structure to formulate and implement data governance rules and norms. It includes the following three components: data stakeholders; data governance committee; and data managers. Processes, that is, the steps and processes that data governance should follow, should be formal, written, repeatable and recyclable [14]. It mainly includes the following contents: active, passive and ongoing data governance process. IBM Data Governance Committee puts forward a maturity model of data governance according to the characteristics of data. In the aspect of building a unified data governance hub, this paper puts forward an essential model of data governance, and considers that business objectives or results are the most critical proposition of data governance. In the factor model, there are several contributing factors that affect the achievement of business objectives, namely organizational structure and awareness, policy and data-related accountability; besides contributing factors, we must focus on the core elements and supporting elements of data governance [15]. Specifically, it includes data quality management, information life cycle management, information security and privacy, data architecture, classification and metadata, as well as auditing, logging and reporting. COBIT is a process-oriented information system audit and evaluation standard formulated by ISACA. It is an internationally recognized authoritative information technology management and control framework [16]. The current version has been updated to 5.0. COBIT 5 is a principle-based top-down framework that makes a strict distinction between governance and management. COBIT 5 puts forward five basic principles of data governance: meeting stakeholder needs, covering enterprises from end to end, adopting a single integration framework, introducing a comprehensive approach, and differentiating governance from management [17]. On the basis of this principle, COBIT 5 elaborates on relevant data governance theories, including stakeholders, contributing factors, scope, key areas of governance and management. The theory of data governance proposed by COBIT 5 is a principle-driven approach. It deduces the complete system of data governance through five basic principles, so that enterprises can establish an effective governance and management framework.

3 Full Service Unified Data Center of Power Grid Company

The full-service unified data center of Power Grid Company includes three parts: data management domain, data analysis domain and data processing domain. Figure 1 gives the architecture of full service unified data center.

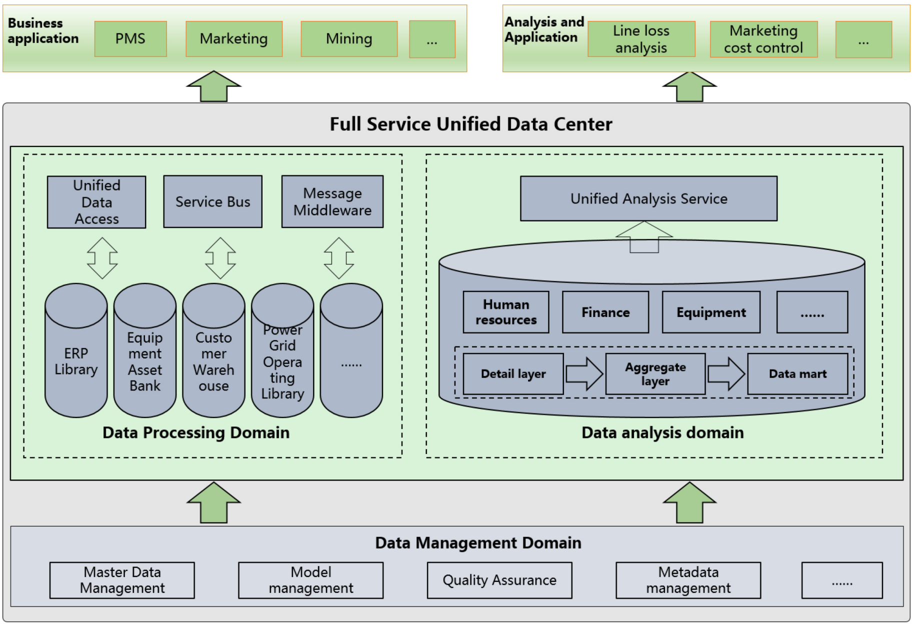

Fig. 1. Architecture of full service unified data center

3.1 Data Management Domain

Data management domain is the center of data model management and master data application in power Grid Company, and it is the key and guarantee to realize data standardization, security and correctness. The data management domain unifies the planning and control of the definition, storage and access of enterprise data in order to ensure the consistency, accuracy and reliability of enterprise-wide data. Data management domain mainly includes unified data model, enterprise-layer master data management system, data resource management component, and master data management component.

Unified data model is divided into two parts: enterprise information model (SG-CIM 3.0) and enterprise data warehouse model. Based on SG-CIM 2.0, the enterprise information model completes the design of core business objects in related subject domains. Based on the enterprise information model, the enterprise data warehouse model corresponds to the hierarchical structure of data analysis domain data warehouse, and completes the physical model design of data warehouse.

Enterprise-layer master data management system is used to sort out and identify business objects (such as organizations, personnel, projects, equipment, etc.) that are used and managed jointly by different professions, and clarify their definitions, attributes, source departments, maintenance methods, application processes and necessary technical implementation requirements.

Data Resource Management Component aims at realizing the unified management and control of the whole business data life cycle, realizing the whole management and control of model, the whole process of data account, the guarantee of data quality and the complete monitoring of data integration. It supports the whole process management of data model design, test and operation of all business systems in the processing domain, and supports the real-time online monitoring of data resources and flow in the analysis domain.

Master data management component realizes the unified management of all master data objects, provides authoritative and reliable data analysis objects for analysis domain, and provides standardized and unified core data resource information for processing domain.

3.2 Data Analysis Domain

Data analysis domain is the center of all kinds of data cleaning, conversion, aggregation and integration of power grid companies, which mainly supports the application of data acquisition, monitoring and analysis and decision-making. Data analysis domain includes data access, data storage computing and unified analysis services.

Data access is mainly responsible for collecting structured data, measurement data, unstructured data and external data from the source business system into the analysis domain.

Data storage computing includes data computing component and data storage layer, in which data computing component (stream computing component, memory computing component and batch computing component) provides distributed running engine and collaborative computing function. The data storage layer includes the basic data layer, the integration detail layer (SG-CIM), the slight summary layer and the data mart layer, in which the basic data layer includes the paste source historical area and the vertical historical area.

Unified analysis service provides unified data interface service, data mining service and self-service analysis service for all kinds of analysis applications. Data interface service component encapsulates different data storage types and computing requirements, realizes multi-table or view join and merge mechanism across libraries and different storage types, and realizes standardization and standardization of data service interface calls. Data mining service component supports data mining and trend prediction oriented to historical data, real-time discrimination and real-time analysis oriented to future prediction and Simulation in a service way. The self-service analysis component provides an intuitive and easy-to-use drag-and-drop interface. By choosing the tables related to the subject and the corresponding tables, charts, words and other presentation forms, setting layout, style and other information, it can realize the centralized, dynamic, real-time and interactive analysis and display of data information.

3.3 Data Analysis Domain

Data processing domain realizes the comprehensive integration and sharing of business data, and supports the construction of integrated business applications. Data processing domain realizes service integration architecture between systems through enterprise service bus and message middleware, realizes cross-departmental business collaboration and logical data sharing; constructs unified data access service, realizes unified access to different types of databases, provides flexible access to multiple data sources for business monitoring applications, and realizes unified data monitoring; follows the unified data model of the company and follows the unified data model of the company. Data architecture requirements, design business processing database, according to the main line of business reasonable division, deployment.

4 Demand Analysis of Big Data Governance in Power Grid

4.1 Deficiencies in Data Management of Power Grid Company

Data management domain is the center of data model management and master data application in power Grid Company, and it is the key and guarantee to realize data standardization, security and correctness. The data management domain unifies the planning and control of the definition, storage and access of enterprise data in order to ensure the consistency, accuracy and reliability of enterprise-wide data. Data management domain mainly includes unified data model, enterprise-layer master data management system, data resource management component, and master data management component.

According to the current situation of industry informatization development and the requirements of data governance in today’s industry, power grid companies have the following shortcomings in data management at this stage:

(1) Data multi-head management, lack of organizations specializing in data management supervision and control. The construction and management functions of information systems are dispersed in various departments, which results in the decentralization of responsibilities and unclear powers and responsibilities of data management. Organizational departments pay attention to data from different perspectives. Lack of an organization to manage data from a global perspective makes it impossible to establish unified data management rules and standards, and the corresponding data management supervision measures cannot be implemented. The data assessment system of the organization has not yet been established, which cannot guarantee the effective implementation of data management standards and procedures.

(2) Decentralized multi-system construction, there is no standardized and unified provincial data standards and data models. In order to meet the rapidly changing market and social needs, organizations have gradually established their own information systems. Each Department produces, uses and manages data from its own standpoint, which makes the data dispersed in different departments and information systems. It lacks unified data planning, reliable data sources and data standards, resulting in data irregularity, inconsistency, redundancy and non-sharing. Problems arise, and it is difficult for organizations and departments to use consistent language to describe the understanding of data, resulting in inconsistent understanding.

(3) Lack of unified master data, the main information such as personnel between the core systems of organizations is not stored in an independent system, or maintained between systems through a unified business management process. Without the management of master data of power grid units, it is impossible to guarantee the consistency, integrity and controllability of master data in the whole business scope, resulting in the failure to guarantee the correctness of business data.

(4) Lack of a unified group data quality management process system. In the current situation, data quality management is mainly carried out separately by various organizations and departments; the mechanism of data quality communication between different departments and departments is not perfect; there is a lack of clear standards and standards for data quality control between different departments and departments; the randomness of data analysis is strong, and the business needs are unclear, which affects the data quality; the automatic data collection has not yet been fully realized, and the process of data processing is human-made. Previously, there are many problems in many departments, such as insufficient data quality management personnel, insufficient knowledge and experience, incomplete supervision methods, and lack of perfect data quality control process and system support ability.

(5) Data life cycle management is incomplete. At present, the norms and processes of data life cycle management for power grid companies are not perfect, which cannot identify expired and invalid data, and unstructured data is not included in the scope of data life cycle management; there is no information tool to support the query of data life cycle status, and metadata management is not effectively used.

4.2 Demand Analysis of Big Data Governance

In view of the problems existing in large data governance of power grid, this paper summarizes various governance scenarios and puts forward the following basic requirements for the framework of large data governance of power grid:

Firstly, the standard of large data for power grid should be established. Data standard is a set of standardization system which conforms to the reality of the group and covers the definition, operation and application of multi-layer data. The establishment of data standards is an important work in the informationization and digitization construction of group units. All kinds of data in the industry must be organized according to a unified standard in order to form a circulating and sharing information platform. The requirements for standards in big data governance can be divided into two categories: basic standards and applied standards. The former is mainly used to form a coordinate reference system for consistent understanding and unification of information among different systems, which is the basis of information collection, exchange and application, including data classification and coding, data dictionary and digital map standards; the latter is to provide certain standard specifications for all links involved in the platform function, so as to ensure the efficient collection and exchange of information, including data classification and coding, data dictionary and digital map standards. Metadata standards, data metadata standards for main entities, data classification and coding standards, data quality standards, and data processing process specifications (Fig. 2).

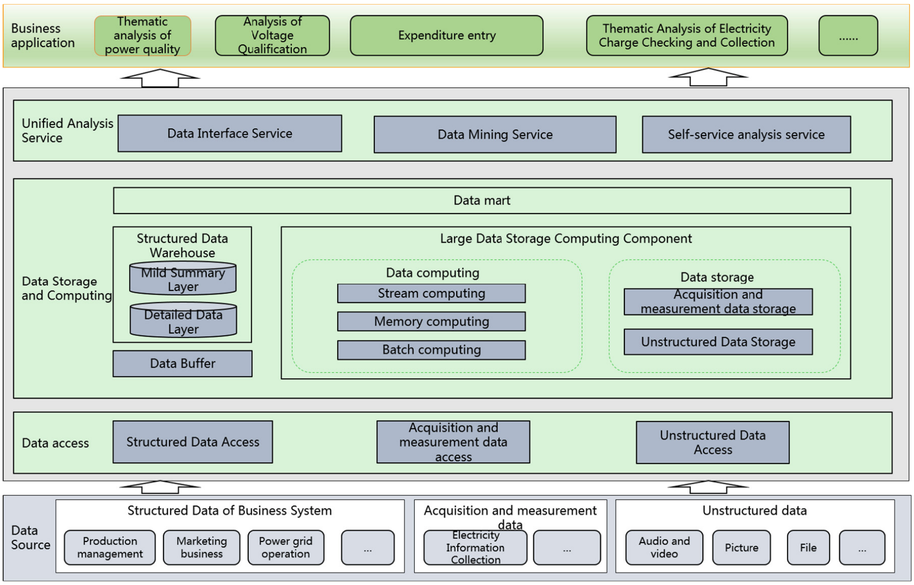

Fig. 2. Demands of data processing in the full service unified data center

Secondly, establish a total data quality management system. Low data quality will affect the application of data warehouse. Low data quality often results in the development of systems that are quite different from users expectations. Data quality is related to the success or failure of the construction of analytical information systems. At the same time, data resources are the strategic resources of the group units. Rational and effective use of correct data can guide the group units to make correct decisions and improve the comprehensive competitiveness of the province. The unreasonable use of incorrect data (i.e. poor data quality) can lead to the failure of decision-making, which can be said to be millions of millions and thousands of miles. Data quality management includes absolute data quality management and process quality management. Absolute quality is the authenticity, completeness and autonomy of data itself. Process quality refers to the quality of use, storage and transmission of data, and the quality of use of data refers to the correct use of data. If the correct data is used incorrectly, it is impossible to draw the correct conclusion. Data storage quality index data are safely stored in appropriate media. The so-called storage in an appropriate medium means that when data is needed, it can be taken out in time and conveniently. Data transmission quality refers to the efficiency and correctness of data in the process of transmission. In the whole life cycle of large data in power grid, a data quality control system is established, which runs through data acquisition, processing, fusion and application, to meet the data quality closed-loop management of "discovery-feedback-correction" of problem data. To provide quantitative automatic data quality evaluation and report for promoting data quality improvement and designing quality evaluation system.

Thirdly, the traceability mechanism of data governance should be established. Focusing on the establishment of a closed-loop control system for data quality to quickly discover and solve the problem data, mining and analyzing the problem data in depth, introducing the necessary data quality control fields in the data modeling stage to realize traceability and feedback to the source of the problem data. The relevant basic functions include: accurately locating the source unit of the problem data, giving the classification and solution of the problem. It is suggested that the source unit be fed back with the problem data sheet and the solution of the problem be tracked. In addition, the workflow mechanism of closed-loop control of problem data is given.

Fourthly, data quality evaluation and improvement plan should be established. The main influencing factors of data quality in large data environment are analyzed. According to the four key characteristics of data quality: data consistency, data timeliness, and data integrity and data accuracy, data quality evaluation indexes are established to guide and assess the system data quality layer under large data environment. This paper studies the data quality evaluation model based on large data, realizes real-time and automatic processing of quality index calculation, statistical analysis and comprehensive evaluation, and meets the requirement of quantitative diagnosis and evaluation of data quality in dynamic and real-time system. It mainly includes data quality index definition model, data quality evaluation method and data quality diagnosis and evaluation results. Quantitative index definition model, mainly studies data quality hierarchical evaluation index tree design, index weight design and index score calculation; data quality evaluation method research, mainly through a certain data algorithm and calculation rules to establish evaluation model, realize the automatic calculation and analysis of index weight and index score, and generate diagnosis and evaluation results. Data administrators regularize data quality auditing, quality alteration processing, data quality assessment and evaluation, feedback to business systems, monitor the whole process from data source to data storage on the ground, solve problems in data quality, improve data quality, and form closed-loop data management.

Finally, the visualization of data governance is realized. The data stored in business system has the remarkable characteristics of wide coverage, large amount of data and long storage time. With the advancement of the practical process of business system, it is urgent for business system to provide a multi-dimensional, multi-form and visualized data display platform to facilitate data mining and statistical analysis of the platform. According to business analysis and data display requirements, multi-dimensional data view can be established. Managers can visually display and analyze large-scale data stored in business systems in multi-dimensional and diversified ways according to needs, and assist management to make more accurate decisions based on statistical analysis results. Combination of Visual Display Platform for Business System.

5 The Framework of Big Data Governance for Full Service Unified Data Center of Power Grid

According to the characteristics and scenario requirement analysis of large data governance in power grid, a data governance framework is proposed from the perspective of innovation of large data application in power grid. The whole governance framework is shown in the Fig. 3.

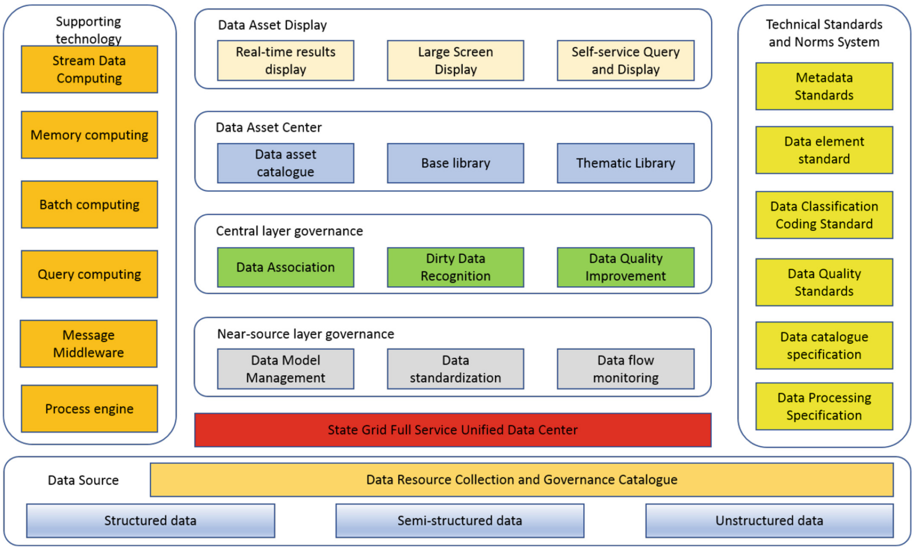

Fig. 3. The framework of big data governance for full service unified data center of power grid

The core idea of the proposed framework are near-source layer governance and central layer governance.

For near-source layer governance, first, in data modeling, master data and general data should be distinguished according to the degree of impact on business attributes, so as to meet the different requirements of data entities in business processes and data quality control. Master data refers to data that is critical to business impact, such as corporate registration information from industry, commerce, civil affairs and compilation; conversely, general data. At the same time, according to the impact of record attributes on data entity recognition, we should design weights for each attribute and distinguish between core and common attributes to meet the quality control requirements of deep data. Secondly, data standardization refers to format conversion, dictionary mapping and preliminary data specification of the collected source data (hereinafter referred to as source data) in accordance with metadata or metadata standard specifications. Thirdly, data verification is the core measure of data quality control. According to the existence of business relevance, it can be divided into technical verification and business verification. Among them, data technology checking refers to checking and checking data which does not involve business. That is to say, according to the data quality standard, using the data checking engine to check the quality of the source data such as format, range, repeatability, integrity and accuracy, so as to find and eliminate the problem data to the greatest extent and lay a solid foundation for the follow-up quality control. Fourthly, data quality assessment is the main output of source-affiliated governance, which is usually output in the form of data quality reports. Data quality report is composed of pre-defined quality evaluation indicators in the standard specification system, which can be used to feedback data governance stakeholders and trigger related business processes of data quality control.

For central layer governance, firstly, data association refers to the series of related data models based on business master data to form a holographic data portrait of the entity, and to preserve the association among these data through related attributes. Data association plays a decisive role in the implementation of government big data application. Usually the data that can be correlated is the actual data available. Secondly, data fusion means that on the basis of data association, the same kind of data is de-duplicated and aggregated to change “one number and multiple sources” into “one number and one source”; or different data fragments of the same entity are constructed to form a new and more complete data description. Data fusion is usually oriented to specific application scenarios, and is one of the most common data operations in data applications. Thirdly, data business verification is based on business attributes of data to check business logic compliance. Business verification is an indispensable part of data quality verification, which is as important as technical verification. Taking the data of personal identity card number as an example, technical verification can only check the compliance of the length, format, specific bit value (area code, age) of the ID card number, but cannot identify the true or false of the number; business verification is to confirm the true or false of the number by comparing the number with the database of the ID card registration authority.

Besides the above two aspects, data standard specification is also very important. Data standard specification is the basic prerequisite for implementing data governance, and plays a decisive role in the effectiveness of data governance. In short, there is no standard specification, no data governance; standard specification is incomplete, data governance is incomplete. For the big data of government affairs, the following norms need to be established and perfected in order to achieve good governance: first, metadata standards. It is necessary to establish metadata standards in an all-round way so as to cover the global data. Secondly, data element standard. It is necessary to establish data element standards selectively for main data entities. Third, data classification and coding standards. It is important to establish classification and coding standards for important data and to establish coding dictionary tables for basic data. Fourth, data catalogue specification. To the greatest extent possible, we should establish a unified directory specification for government data resources and standardize directory coding and operation to the greatest extent. Fifth, data quality standards. From the perspective of accuracy, compliance, consistency, repeatability, timeliness, integrity and other indicators, a comprehensive data quality standard should be established, and evaluation indicators and evaluation methods should be given. Sixth, data governance process specification. Processing is the guarantee of orderly governance. Data governance should be streamlined, corresponding process specifications should be established, and the layer of orderly governance should be improved by process specifications.

6 Conclusion

Based on the full-service unified data center of power Grid Company, this paper proposes a general frame-work of self-service power grid big data governance. In this paper, firstly, the related work of big data governance is reviewed; secondly, the architecture and characteristics of the grid company's full-service unified data center are analyzed; thirdly, the related requirements of self-service grid big data governance are analyzed; fourthly, the overall framework of self-service grid big data governance is proposed. Compared with other general big data governance framework, the proposed framework considers the specification of power grid as well as the features of self-service, making the framework more feasible.

Acknowledgments. This work was supported by the State Grid Corporation Science and Technology Project (Contract No.: SGLNXT00YJJS1800110).

References

1. Zaveri, A., Rula, A., Maurino, A., et al.: Quality assessment for linked data: a survey. Semant. Web **7**(1), 63–93 (2015)
2. Zhiqiang, W., Qinghai, Y.: Research on the quality and standardization of scientific data. Stan. Sci. **03**, 25–30 (2019)
3. Mohan, L., Jianzhong, L., Honghong, G.: Solution algorithm for data timeliness determination. J. Comput. Sci. **35**(11), 2348–2360 (2012)
4. Fan, W., Geerts, F.: Relative information completeness. ACM Trans. Database Syst. (TODS) **35**(4), 27 (2010)

5. Fan, W., Li, J., Ma, S., et al. Interaction between record matching and data repairing. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, Athens, Greece, 12–16 June 2011. ACM (2011)
6. Fan, W., Ma, S., Tang, N., Yu, W.: Interaction between record matching and data repairing. J. Data Inf. Q. (JDIQ) **4**(4), 16 (2014)
7. Quercia, D., Hogan, B.: Proceedings of the Ninth International AAAI Conference on Web and Social Media, ICWSM 2015. AAAI Press (2015)
8. Ding, X., Wang, H., Zhang, X., et al.: Research on the relationship among various properties of data quality. J. Softw. **27**(7), 1626–1644 (2016)
9. Cheng, H., Feng, D., Shi, X., et al.: Data quality analysis and cleaning strategy for wireless sensor networks. EURASIP J. Wireless Commun. Netw. **2018**(1), 61 (2018)
10. Li, C., Liang, Yu., Yangyong, Z., et al.: Historical evolution and development trend of data quality. Comput. Sci. **45**(4), 1–10 (2018)
11. Saha, B., Srivastava, D.: Data quality: the other face of big data. In: IEEE International Conference on Data Engineering. IEEE (2014)
12. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manage. Inf. Syst. **12**(4), 5–33 (1996)
13. Sidi, F., Panahy, P.H.S., Affendey, L.S., et al.: Data quality: a survey of data quality dimensions. In: International Conference on Information Retrieval & Knowledge Management (2012)
14. Zhao, W., Li, C.: A review of the research on quality evaluation methods of associated data. Intell. Theory Pract., **39**(02), 134–138 + 128 (2016)
15. Liu, H.: Analysis of statistical data quality. In: International Joint Conference on Computational Sciences & Optimization. IEEE (2014)
16. Alpar, P., Winkelsträter, S.: Assessment of data quality in accounting data with association rules. Expert Syst. Appl. **41**(5), 2259–2268 (2014)
17. Song, C., Jing, W., Zeng, P., et al.: Energy consumption analysis of residential swimming pools for peak load shaving. Appl. Energy **220**, 176–191 (2018)