Studies on Data Quality Evaluation Index System for Internet Plus Government Services in Big Data Era

Li Ya\textsuperscript{1, a}, Song Heliang \textsuperscript{1, b*} and Xu Yingcheng\textsuperscript{1}

\textsuperscript{1}China National Institute of Standardization, Beijing 100191, China

Correspondence author: \textsuperscript{a}15201290448@163.com; \textsuperscript{b*}18610137786@163.com

Abstract. This paper studies and analyzes the related theories of data quality and data quality of Internet plus Government Services in big data era, summarizes and references the data quality evaluation indexes at home and abroad, puts forward the primary and secondary indexes suitable for evaluating the data quality of Internet plus Government Service in big data era in China, and constructs a scientific, systematic, developmental and operable data quality evaluation index system. The index system determines, according to the characteristics of big data, 5 primary data quality evaluation indexes (inherent quality of data, quality of data expression, situation quality related to the system, utility quality of data, and user experience quality of data) and 10 secondary data quality evaluation indexes. The construction of the system provides an effective data quality evaluation tool for government departments. Through data quality evaluation, relevant data quality problems can be found, analyzed and improved, which is conducive to improving the level of data quality.

1. Introduction
The Chinese government attaches great importance to the work of “Internet Plus Government Services”\textsuperscript{[1]} which was proposed for the first time in the Report on the Work of the Government (2016) and mentioned again in the Report on the Work of the Government (2017), Report on the Work of the Government (2018) and Report on the Work of the Government (2019). During this period, the Government has issued documents such as Guiding Opinions of the State Council on Vigorously Advancing the “Internet Plus” Action, Guidelines for the Construction of Technical System of “Internet Plus Government Service” and Implementing Program for Further Improving Internet Plus Government Services and Promoting the Reform for Providing Online, Offline and In-person Access to One-stop Government Services, and has proposed that governments at all levels should construct and perfect government data resource systems, continuously improve data quality and improve service availability. Vigorous promotion of “Internet Plus Government Services” and increasingly accumulating government service data volume promote the formation of government big data.

In big data era, government service data is a huge fortune. The mining and application of government big data will help innovate government service methods, effectively improve the efficiency of government public service and control the cost of operating government management. However, at present, China's online government service platforms are mostly constructed separately, with large regional differences, and are faced with problems such as large data differences and different data quality among departments and regions, which lead to the difficulty of data sharing, “information isolated island” and low utilization rate of information. Therefore, the construction of data quality evaluation index system for Internet Plus Government Services (hereinafter referred to as...
the Index System) in big data era provides an effective data quality evaluation tool for government departments, which will help standardize the scientific management of government big data, accelerate the standardization of data resources, improve the quality of government service data in big data era, improve the work efficiency of government staff, and promote the modernization of government governance capability.

2. Study Status of Data Quality Evaluation Index System for Internet Plus Government Services in Big Data Era

2.1. Basis for Studies on Big Data Quality

With the advent of the big data era, the quality of big data has attracted increasing attention from scholars at home and abroad.

Wang and Strong[2] believe that the significance of data quality to data users lies not only in accuracy, and propose that in the data quality study, a conceptual model of data quality should be considered and constructed from four aspects: quality of intrinsic data, of situational data, of expressive data and of accessible data. Y. Demchenko et al. describe the definition of big data and its “5V” characteristics, point out the needs of different scientific research teams for big data management, acquisition control and security, and put forward suggestions on unstructured factors in big data. D. Boyd et al. explain the impact and significance of big data from the aspects of culture, technology, academic phenomenon and data analysis. J. M. Tien outlines the basic theories of data collection, data access, data analysis and data application in big data era and their improvement methods [3]. B. Saha points out that there are a large number of low-quality data in large databases and networks, which will seriously affect the results of data analysis. T. Lukoianova et al. believe that the differences in the sources, processing techniques, data collection and scientific discovery methods of big data lead to the quality problems of big data such as deviation, ambiguity and inaccuracy. They elaborate the concept of “big data accuracy” in detail and explore the “accuracy” of big data from three main dimensions: objectivity and subjectivity, honesty and deception, trust and disbelief respectively. T. Taleb et al. propose to solve big data quality problems in the data item processing stage, including subprocesses such as cleaning, integration, filtering and normalization and propose a big data quality model combined with processes to support data quality selection and adaptation [4]. ISO 8000 series of standards is an international standard system related to data quality, and it describes the principle of data quality, defines the characteristics of data quality, and gives a data quality certification process similar to ISO9000 series of standards [5].

Song Wei [6] et al. organize the basic characteristics of big data and discuss the challenges, importance and countermeasures of enterprise data quality in big data era from the perspectives of process, technology and management. Mo Zuying [7] studies the definition of big data quality, and proposes the quality dimensions of big data according to the connotation and characteristics of big data quality, including data authenticity, data integrity, data consistency, data accuracy, data effectiveness, data security, data availability, data value, and understandability. Liu jinjing [8] et al., considering the characteristics of the big data platform in the public security big data field and various influencing factors such as data processing flow, data source, user data and data model, respectively, put forward an evaluation index framework from three aspects of data quality, data processing quality and data efficiency to evaluate and measure the data quality in the whole life cycle. Liu Yandong [9] analyzes the data quality problems in big data, and divides the big data quality problems into six aspects from the perspective of collection, application and users, including accuracy, integrity, consistency, relevance, effectiveness, credibility and interpretability. GB/T 36344 establishes a data quality evaluation index framework, which determines the data quality evaluation index from the aspects of standardization, integrity, accuracy, consistency, timeliness and accessibility.

2.2. Study Status of Data Quality Evaluation Index System for Internet Plus Government Services in Big Data Era
With the vigorous promotion of “Internet Plus Government Services” and the construction and application of government service platforms at all levels across the country, the importance of government service data resources has become increasingly prominent. The study on the quality of Internet Plus Government Services data has also become the focus of study in the field of practice and academia.

Zhai Jun [10] et al. perform statistics on the distribution of data quality problems in Beijing, Shanghai and Harbin, and analyze the current situation of data quality problems from the aspects of data sets and data files, definition of data patterns, consistency of data in various formats, timeliness judgment through metadata, reference to external authoritative data, privacy disclosure, etc. Zhai Jun [11] et al. organize the study progress of open government data quality evaluation, and summarize the dimensions and index of quality evaluation from two aspects (metadata and data sets) and at two levels (mode level and example level). Zhang Nannan [12] analyzes the quality problems of open data at the data itself, technical level, process level and management level starting with the definition of the quality dimension of government open data. Liu Bohao [13] uses SPDQM model for reference, and establishes a data quality evaluation system consisting of data intrinsic quality attributes and user use attributes from the data point of view. Ma Yiming [14] determines, according to the existing data quality evaluation dimensions and the characteristics of big data, 5 primary data quality evaluation indexes (data source quality, data scale quality, data structure quality, data actual effect quality, data value density quality) and 17 secondary data quality evaluation indexes. Shao Yanhong [15] determines, combining the principles and characteristics of government's open data, 4 primary data quality evaluation indexes (data content, data expression, data utility and user experience) and 12 secondary data quality evaluation indexes. Lang Yanna [16] conducts an objective and effective analysis on the data quality of China's provincial government's open data platform, and establishes four primary data quality evaluation indexes (data completeness, metadata, data format and update, data acquisition) and 10 secondary data quality evaluation indexes.

3. Construction of Data Quality Evaluation Index System for Internet Plus Government Services in Big Data Era

3.1. Principles of Index System Construction

Building a scientific and reasonable data quality evaluation index system is the premise to improve the quality of government service data and user satisfaction. In the design process of the index system, relevant principles should be followed in combination with the characteristics of study objects to ensure that the data quality evaluation index for Internet Plus Government Services in the big data era can comprehensively, accurately, objectively and scientifically evaluate the big data quality in this field. In the process of designing the data quality evaluation index for Internet Plus Government Services in big data era, the following principles should be followed [14-17]:

1) Principle of scientificity: The construction of the index system should first meet the requirements of scientificity, which is an important prerequisite to ensure the rigor and authority of the index system. For construction of the index system, clear objectives should be set, being supported by solid scientific theories, and in reference with the existing efficient data quality evaluation system. The index system should be provided with scientific and reasonable basis during index selection and screening to ensure a scientific and reasonable evaluation index system for Internet Plus Government Services data quality in big data era.

2) Principle of systematicness: The internal structure of the index system should be an organic whole with a certain level structure and complement each other. With Internet Plus Government Services data quality evaluation in big data era as the goal, it is decomposed into primary index and secondary index to form a logically rigorous, comprehensive and hierarchical indicator system.

3) Principle of development: The index system should follow the developmental principle and keep pace with the times and continuously improve. The selected indexes can be used for comprehensively and completely evaluating the quality of Internet Plus Government Services data. At
the same time, the index system should be kept open and continuously adjusted to guide the healthy and orderly development of Internet Plus Government Services data.

(4) Principle of operability: The purpose of index system construction is to apply to Internet Plus Government Service data quality management in big data era, and to analyze and evaluate the quality of big data in this field. Therefore, the overall framework structure and specific indexes of the index system must be operable and can be applied in practice.

3.2. Route Data Quality Evaluation Index System for Internet Plus Government Services in Big Data Era

Referring to the relevant study conclusions of big data quality evaluation, this paper puts forward an index system from five aspects of inherent quality of data, quality of data expression, situation quality related to the system, utility quality of data, and user experience quality of data, focusing on the data itself, to evaluate and measure the quality of Internet Plus Government Services data in big data era, as shown in Figure 1.

**Figure 1. Data Quality Evaluation Index System for Internet Plus Government Services in Big Data Era**

3.2.1. Inherent Quality of Data. (1) Accuracy. Refers to the fact that data should have the expected attribute of correctly representing concepts or time, mainly including data content correctness, data format compliance, data repetition rate, data uniqueness and dirty data occurrence rate. Accuracy is an important evaluation index of data quality and the basis for judging data quality.

(2) Authenticity. Refers to the fact that data should have authentic attributes. Only the real data can reflect the objective facts. Authenticity requires the data content to be objective and true, and the way of data generation is not processed or falsified.

(3) Integrity. Refers to the fact that data can be used for displaying all required information in a specific data set, mainly including data element integrity and data record integrity. Incomplete data will affect the data analysis, causing the data analysis results to fail to meet the accuracy requirements.

3.2.2. Quality of Data Expression Forms. (1) Standardization. Refers to that the generation of data should conform to the data standards in specific fields, mainly including the standardization of data standards, data models, metadata, business rules and authoritative reference data.

(2) Understandability. The expression of data meets the degree that users can understand. Understandability requires that the expression form of data should be machine-processable, the expression of data should be intuitive and understandable, statements should be unambiguous, and relevant documents should be available to explain the symbols and abbreviations of data.
3.2.3. Context Quality Related to the System. (1) Consistency. Refers to the degree to which data attributes are consistent in different system environments, mainly including the consistency of the same data and the consistency of associated data. The consistency of data plays an important role in data integration, data sharing and data coordination of government service platforms at all levels.

(2) Traceability. It is required that the main body, content, time and other elements involved in the process of generating, disclosing, acquiring and using data can be inverted after data format conversion. For some non-original data, processed data should provide traceability information, and data changes should be audited and tracked to facilitate data query and use.

3.2.4. Utility Quality of Data. (1) Accessibility. Refers to the degree to which data can be accessed in a specific use environment, mainly including accessibility and usability. Data shall be available when needed and usable within the set effective life cycle. Data accessibility is a prerequisite for data to be used by users.

(2) Timeliness. Refers to that in a specific use environment, data should have appropriate age attributes, mainly including correctness based on time period, timeliness based on time point and timing. The timeliness of data can help the government to solve problems or formulate plans in a timely and effective manner, and especially for some emergencies, timeliness is of great significance.

3.2.5. User Experience Quality of Data. (1) Security. Refers to that data meets security and privacy rules, including data rights management and data masking. Different from simply obtaining information, “Internet Plus Government Services” takes serving the people as its starting point and emphasizes that people can handle affairs through government websites, especially the functions of personal information exchange, formalities handling and electronic payment. Once there is a data security problem, it not only endangers the interests of citizens, but may even endanger national security. Therefore, data security is the lifeline to ensure data quality.

4. Brief Summary
This paper studies and summarizes the study status of the data quality evaluation index system. Following the principles of scientificness, systematicness, developability and operability, it comprehensively analyzes the influencing factors of government service data quality from five aspects including the inherent quality of data, the quality of data expression forms, the context quality related to the system, the utility quality of data, and the user experience quality of data. It puts forward the data quality evaluation index system for Internet Plus Government Services in big data era, which will contribute to the scientific management of government big data and improve the quality level of government big data.

Acknowledgements
We would like to acknowledge that this Study is supported and funded by the Market Supervision Technology Assurance Project under Grant No. 2020YJ043, the National Science Foundation of China under Grant No. 91646122, 91746202, the Basic Scientific Research Business Projects 552018Y-5927-2018.

References
[1] Li Zhen. Studies on “Internet Plus Government Services” in Big Data Era [D]. Jinan: Shandong Normal University, 2019.
[2] Lab for Digital & Mobile Governance website of Fudan University. China Open Data Index (H2, 2018) [EB/OL]. [2019-01-02].http://www.dmg.fudan.edu.cn/?p=6210
[3] G Decandia, D Hastorum, M Jampani. Dynamo: amazon’s highly available key-value store[J].Acm Sigops Operating Systems Review.2007
[4] G Fauconnier, M Turner. The way we think: conceptual blending and the mind’s hidden complexities [J]. Wired. 2002
[5] Wang Junling, Li Hua, Wang Qiang. Analysis of ISO 8000 Data Quality Series Standards [J]. Standard Science, 2010 (12).

[6] Song Wei, Wu Feng. Challenges of Data Quality in Big Data Era [J]. Journal of Xi 'an Jiaotong University (Social Science), 2012, 05: 38-43.

[7] Mo Zuying. Analysis of Data Quality Impact in Big Data Processing Process [J]. Modern Intelligence, 2017 (03): 69-72.

[8] Liu Jinjing, Wang Mei. Construction Practice of Data Quality Evaluation Index in Big Data [J]. Computer Technology and Development, 2019, 10 (29): 46-50.

[9] Liu Yandong. Analysis on the Quality of Data in Big Data [J]. Modern Business Trade Industry, 2020 (4): 193.

[10] Ma Yiming. Studies on the Construction of Government Big Data Quality Evaluation System [D]. Changchun: Jilin University, 2016.

[11] Shao Yanhong. Studies on the Construction of Evaluation Index System for Open Data Quality of Chinese Government [D]. Baoding: Hebei University, 2019.

[12] Lang Yanna. Studies on Data Quality Evaluation of Open Platform of Provincial Government in China [D]. Baoding: Hebei University, 2019.

[13] Zhai Jun, Li Xiaotong, et al. Study on the Problem of “Dirty Data” in China's Open Government Data and Its Countermeasures [J]. Library, 2019 (01): 42-51.

[14] Zhai Jun, Tao Chenyang, et al. Study Progress and Enlightenment of Open Government Data Quality Evaluation [J]. Library, 2018 (12): 74-79.

[15] Zhang Nannan. Study on the Quality Control Mechanism of China's Open Government Data [J]. Journal of Intelligence, 2019 (01): 136-141.

[16] Liu Bohao. Study on Quality Evaluation of China's Open Government Data [D]. Zhengzhou: Zhengzhou University, 2019.

[17] Ren Zhaobo. Studies on Quality Evaluation Index System of Aviation Economy Big Data [D]. Zhengzhou: Zhengzhou University of Aeronautics, 2019.