Method for Calculating the Weights of Internet + Government Service Data Quality Assessment Indexes Based on Analytic Hierarchy Process

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{1}15201290448@163.com, \textsuperscript{b}*18610137786@163.com

Abstract. In this paper, research and analysis have been carried out on the theories related to Internet + government service data quality. Based on the principles of scientificity, systematicness, development, and operability, an Internet + government service data quality assessment index system suitable for China's big data era has been proposed, this index system is proposed from five aspects: inherent quality of data, quality of data expression, system-related scenario quality, utility quality of data, and user experience quality of data. The analytic hierarchy process method has been used to analyze the relationship between indexes at the criterion layer, the sub-criterion layer and the objective layer, and the weights of the data quality assessment indexes at each layer were determined. The method proposed in this paper provides an idea for the assessment of government service data quality and has certain practical significance, laying a theoretical foundation for the related work of Internet + government service data quality assessment.

1. Introduction
Since 2016, the Chinese government has attached great importance to the "Internet + government service" work and put forward the requirement of "continuously improving data quality and service availability". With the vigorous advancement of the "Internet + government service" work, many data issues have aroused widespread concern, such as difficulty in data sharing due to large data differences and inconsistent data quality, and "information silos". Therefore, improving the quality of government service data and the utilization rate of government information has become the focus of scholars' research in China. Many factors can affect the quality of government service data, and the relationship between these factors is complex. This paper draws on relevant research conclusions at home and abroad, and combines with the actual situation of big data of government services in China to build an Internet + government service data quality assessment index system. With reference to the successful application cases of analytic hierarchy process (AHP) in data quality management, scheme optimization, risk management and other fields [1-5], the AHP method has been used to determine the weight of each assessment index, providing an effective data quality assessment tool for government departments.

2. Construction of Internet + Government Service Data Quality Assessment Index System
With reference to the relevant research conclusions of data quality assessment at home and abroad and based on the principles of scientificity, systematicness, development, and operability [6-9], this paper
proposes the Internet + government service data quality assessment index system from five aspects: inherent quality of data, quality of data expression, system-related scenario quality, utility quality of data, and user experience quality of data, as shown in Figure 1.

![Diagram of Internet + Government Service Data Quality Assessment Index System](image)

**Figure 1.** Internet + Government Service Data Quality Assessment Index System

### 2.1. Inherent Quality of Data
(1) Accuracy: It means that the data should have the expected attribute of correct representation of concept or time, mainly including the correctness of data content, the compliance of data format, the data repetition rate, the uniqueness of data, and the occurrence rate of dirty data.

(2) Authenticity: It means that the data should be authentic and credible.

(3) Completeness: It means that the data should be able to display all the required information in a specific data set, mainly including the completeness of data elements and data records.

### 2.2. Quality of Data Representation
(1) Normalization: It means that the generation of data should conform to the data standards in a specific field, mainly including the normalization of data standards, data models, metadata, business rules, and authoritative reference data.

(2) Understandability: It means that the expression of data should be understandable to the user.

### 2.3. System-related Scenario Quality
(1) Consistency: It refers to the degree to which the attributes of data are consistent in different system environments, mainly including the consistency of the same data and associated data.

(2) Traceability: The subject, content, time and other elements involved in the process of data generation, disclosure, acquisition and use should be traceable after the data format conversion.

### 2.4. Utility Quality of Data
(1) Accessibility: It refers to the degree to which the data can be accessed in a specific use environment, mainly including accessibility and availability.

(2) Timeliness: It means that in a specific use environment, the data should have appropriate age attributes, mainly including time period-based correctness, time point-based timeliness, and time series.

### 2.5. User Experience Quality of Data
(1) Security: It means that the data should meet security and privacy rules, mainly including data rights management and data desensitization.
3. AHP-based Weight Calculation Method

Analytic hierarchy process (AHP) is a simple and practical weighting method for decision making that combines qualitative and quantitative analysis by applying network system theory and multi-objective comprehensive evaluation method. Based on an in-depth analysis of the nature, influencing factors, and internal relationships of complex decision-making problems, this method quantifies the decision-making process by using a small amount of quantitative information [10], and determines the degree of importance of each element at the same layer compared with the elements at the previous layer by pairwise comparison, to finally obtain the weight of each index. The steps for determining the weight of the Internet + government service data quality assessment index system based on AHP are as follows.

3.1. Construction of Judgment Matrix

AHP requires to judge the relative degree of importance of each index at each layer and scale the judgment with numerical values to form a judgment matrix. In this study, 10 experts in related fields were invited to assign values to the judgment matrix according to the scale in Table 1.

Table 1. Meaning of Scale

| Scale Value | Degree of Importance of One Factor over the Other Based on Pairwise Comparison |
|-------------|--------------------------------------------------------------------------------|
| 1           | Equally important                                                              |
| 3           | Slightly more important                                                        |
| 5           | Obviously more important                                                       |
| 7           | Highly more important                                                          |
| 9           | Absolutely more important                                                       |
| 2, 4, 6, 8  | Median of adjacent judgments above                                             |

The scores given by all the experts were summarized construct the decision matrix of criterion layer and sub-criterion layer for the Internet + government service data quality assessment index system, as shown in Tables 2-6.

Table 2. Criterion Layer Judgment Matrix

| Criterion Layer | B₁   | B₂   | B₃   | B₄   | B₅   |
|-----------------|------|------|------|------|------|
| B₁              | 1.000|      |      |      |      |
| B₂              |      | 5.000|      |      |      |
| B₃              |      |      | 3.000|      |      |
| B₄              |      |      |      | 4.000|      |
| B₅              |      |      |      |      | 4.000|

Table 3. Sub-criterion Layer Judgment Matrix

| Sub-criterion Layer | C₁   | C₂   | C₃   |
|---------------------|------|------|------|
| C₁                  | 1.000|      |      |
| C₂                  |      | 1.000|      |
| C₃                  |      |      | 1.000|
Table 4. Sub-criterion Layer Judgment Matrix

| Sub-criterion Layer B_3 | C_4    | C_5    |
|-------------------------|--------|--------|
| C_4                     | 1.0000 | 2.0000 |
| C_5                     | 0.5000 | 1.0000 |

Table 5. Sub-criterion Layer Judgment Matrix

| Sub-criterion Layer B_3 | C_6    | C_7    |
|-------------------------|--------|--------|
| C_6                     | 1.0000 | 3.0000 |
| C_7                     | 0.3333 | 1.0000 |

Table 6. Sub-criterion Layer Judgment Matrix

| Sub-criterion Layer B_3 | C_8    | C_9    |
|-------------------------|--------|--------|
| C_8                     | 1.0000 | 3.0000 |
| C_9                     | 0.3333 | 1.0000 |

3.2. Calculation of Weights of Indexes at Each Layer

Calculate the weight of each data quality assessment index at the criterion layer and the sub-criterion layer according to Table 2-6. The specific algorithm is as follows with the weight calculation process of inherent quality of data, quality of data expression, system-related scenario quality, utility quality of data, and user experience quality of data at the criterion layer as an example:

3.2.1. Normalization of Judgment Matrix Columns. Determine and construct the consumer goods criterion layer judgment matrix - the relative importance comparison judgment matrix

\[ B = (b_{ij})_{5 \times 5} \]

according to the scale of 1-9 in Table 1, and normalize each column of the judgment matrix:

\[
\begin{bmatrix}
1 & 5 & 3 & 4 & 4 \\
0.2000 & 1 & 2 & 2 & 4 \\
0.3333 & 0.5000 & 1 & 3 & 3 \\
0.2500 & 0.5000 & 0.3333 & 1 & 2 \\
0.2500 & 0.2500 & 0.3333 & 0.5000 & 1
\end{bmatrix}
\]

Therefore,

\[
\bar{B} = \begin{bmatrix}
0.4918 & 0.6897 & 0.4500 & 0.3810 & 0.2857 \\
0.0984 & 0.1379 & 0.3000 & 0.1905 & 0.2857 \\
0.1639 & 0.0690 & 0.1500 & 0.2857 & 0.2143 \\
0.1230 & 0.0690 & 0.0500 & 0.0952 & 0.1429 \\
0.1230 & 0.0345 & 0.0500 & 0.0476 & 0.0714
\end{bmatrix}
\]
3.2.2. Adding of Rows of the Normalized Judgment Matrix. Add the normalized judgment matrix by rows to get the vector \( \overline{W} = (\overline{w}_1, \overline{w}_2, \ldots, \overline{w}_n)^T \), as follows:

\[
\overline{W} = \begin{bmatrix}
2.2981 \\
1.0125 \\
0.8829 \\
0.4800 \\
0.3265
\end{bmatrix}
\]

3.2.3. Obtaining of Eigenvector. Normalize the vector \( \overline{W} = (\overline{w}_1, \overline{w}_2, \ldots, \overline{w}_n)^T \) to obtain the eigenvector \( W = (w_1, w_2, \ldots, w_n)^T \), i.e., the weight vector of the criterion layer index:

\[
W = \begin{bmatrix}
0.4596 \\
0.2025 \\
0.1766 \\
0.0960 \\
0.0653
\end{bmatrix}
\]

Therefore,

\[
BW = \begin{bmatrix}
2.6470 \\
1.1008 \\
0.9149 \\
0.5016 \\
0.3377
\end{bmatrix}
\]

3.2.4. Obtaining of Maximum Eigenvalue. Calculate the maximum eigenvalue \( \lambda_{\text{max}} \) of the judgment matrix with \( BW \) and \( W \):

\[
\lambda_{\text{max}} = \sum_{i=1}^{n} \left( BW \right)_{ni} w_i = \frac{2.6470}{5 \times 0.4596} + \frac{1.1008}{5 \times 0.2025} + \frac{0.9149}{5 \times 0.1766} + \frac{0.5016}{5 \times 0.0960} + \frac{0.3377}{5 \times 0.0653} = 5.3546
\]

3.2.5. Consistency Check. Obtain the consistency check index \( CI \) based on the maximum eigenvalue of the judgment matrix:

\[
CI = \frac{\lambda_{\text{max}} - n}{n - 1} = \frac{5.3546 - 5}{5 - 1} = 0.0887
\]

Then query Table 7 to get the consistency index \( RI \), and calculate the consistency ratio \( CR \) as follows:

\[
CR = \frac{CI}{RI} = \frac{0.0887}{1.12} = 0.0792 < 0.1
\]
Table 7. Calculated RI by Repetition for 1,000 Times

| N  | 1     | 2     | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
|----|-------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| RI | 0.00  | 0.00  | 0.52| 0.89| 1.12| 1.26| 1.36| 1.41| 1.46| 1.49| 1.52| 1.54|

Based on this, it can be determined that the judgment matrix shows satisfactory consistency, so that the relative weights of the five indexes at the criterion layer can be determined. Repeat the above process and calculate the weight of each index at the sub-criterion layer to finally determine the weight relationship between the indexes, as shown in Table 8:

Table 8. Weight of the Internet + Government Service Data Quality Assessment Index System

| Objective Layer | Criterion Layer | Sub-criterion Layer |
|-----------------|-----------------|---------------------|
| Internet + government service data quality assessment index system A | B₁ = 0.4596 | C₄ = 0.1532 |
|                  | B₂ = 0.2025 | C₅ = 0.1532 |
|                  | B₃ = 0.1766 | C₆ = 0.1532 |
|                  | B₄ = 0.0960 | C₇ = 0.1532 |
|                  | B₅ = 0.0653 | C₈ = 0.1532 |

3.3. Analysis of Assessment Results

The weight coefficient of the Internet + government service data quality assessment index was determined by using AHP method. It can be seen from the calculation results that the weight value of inherent quality of data > the weight value of quality of data expression > the weight value of system-related scenario quality > the weight value of utility quality of data > the weight value of user experience quality of data, which indicates that the inherent quality of data is the primary influencing factor that affects the results of data quality assessment. Among the secondary indexes, the accuracy, authenticity, completeness, etc. of the data have a larger weight. They are important indexes for the Internet + government service data quality assessment and also the basis for judging the quality of Internet + government service data.

4. Summary

Scientific and reasonable data quality index weight values are the premise of the Internet + government service data quality assessment. This paper proposes the Internet + government service data quality assessment index system from five aspects: inherent quality of data, quality of data expression, system-related scenario quality, utility quality of data, and user experience quality of data. AHP method was used to analyze the relationship between indexes at the criterion layer, the
sub-criterion layer and the objective layer, and the weights of indexes at each layer were determined. The method of calculating the weights of the Internet + government service data quality assessment indexes based on AHP proposed in this paper provides an idea for the assessment of government service data quality and has certain practical significance.

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] Cai Li, Li Yongxuan, et al. Research on Quality Assessment of Crowdsourced Geographic Data Based on Analytic Hierarchy Process [J]. Journal of Geomatics, 2020.
[2] Li Chen, Yin Zili, Wang Xiaohui, et al. Evaluation of Distribution Network Dispatching Based on Analytic Hierarchy Process and Entropy Weight Method [J]. Proceedings of the CSU-EPSA, 2019: 1-8.
[3] Zhu Jingyan, Wen Xiuying, Yang Ming. Research on Operation Risk Assessment of Comprehensive Pipe Corridor Based on Analytic Hierarchy Process [J]. Shanxi Architecture, 2020, 46 (2): 179-180.
[4] Qi Aijie, Liu Lun, et al. Determination of Annual Priority Improvement Indexes of Hospital Quality and Safety by Delphi Method and Analytic Hierarchy Process [J]. Chinese Journal of Hospital Statistics, 2019, 26 (6): 445-467.
[5] Wang Mingzheng, Meng Weiqiang, et al. Education Security Data Management Using Hierarchical Data Quality Assessment Method [J]. Technology Research, 2019, 12: 79-87.
[6] Ma Yiming. Research on the Construction of Government Big Data Quality Assessment System [D]. Changchun: Jilin University, 2016.
[7] Shao Yanhong. Research on the Construction of the Open Government Data Quality Assessment Index System in China [D]. Baoding: Hebei University, 2019.
[8] Lang Yanna. Research on Quality Assessment of Open Platform Data of Provincial Government in China [D]. Baoding: Hebei University, 2019.
[9] Ren Zhaobo. Research on Aviation Economy Big Data Quality Assessment Index System [D]. Zhengzhou: Zhengzhou University of Aeronautics, 2019.
[10] Zhu Jiayong, Su Hang. Discussion on Port Construction Mode Based on Analytic Hierarchy Process [J]. Port & Waterway Engineering, 2019, 554 (4): 89-94.