

# Research on Theory and Practice of Data Quality Management

Yundan Liang<sup>1</sup>, Yinghui Xu<sup>2\*</sup>, Hongbin Xie<sup>2</sup>, Yi Huang<sup>1</sup>, Zhi Yang<sup>2</sup>, Chuang Li<sup>2</sup>, Lianzhu Ge<sup>2</sup>, and Tianying Jiang<sup>3</sup>

<sup>1</sup>State Grid Corporation of China, Beijing, 100031, China

<sup>2</sup>China Electric Power Research Institute Co., Ltd, Beijing, 100192, China

<sup>3</sup>Beijing Wuzi University, Beijing, 101149, China

**Abstract.** Data quality management includes both theory and technology, as well as practice and management. Through data quality management, enterprises can obtain high-quality data, which is an important prerequisite for developing data products, providing data services, and realizing the value of data. It is also the core goal of enterprise data governance. This article first reviews and compares the relevant theories of data quality management; Then summarize the data quality practices carried out by typical case enterprises; Finally, based on the theory and practical cases of data quality management, valuable experiences for data quality management are proposed, and prospects for data quality management are presented.

## 1. Introduction

Strengthening data quality management is the core task of building a strong quality country. Data quality is the core element of data value and the development of the digital economy, and data quality management is of utmost importance in comprehensively improving the level of quality management.

Strengthening data quality management is an important measure to promote the development of enterprise data. Data is the fundamental strategic resource for enterprise development, and data governance is a necessary foundation for enterprises to achieve digital transformation. The level of data quality is an important benchmark for measuring the level of data governance and a key indicator for measuring the degree of digital transformation. Only continuous improvement of data quality can promote the improvement of data governance level, and only by formulating scientific evaluation standards can the direction of data quality improvement be clarified, providing strong data element guarantee for enterprise digital transformation.

## 2. Data Quality Management Theory

### 2.1. DAMA-DMBOK2

DAMA-DMBOK2 provides a detailed introduction to the business drivers, goals and principles, activities, tools, methods, implementation guidelines, and other aspects of data quality management. The content covers the actions, tasks, techniques, and methods required to achieve data quality management goals, providing comprehensive and detailed guidance for data quality

management<sup>[1]</sup>. DAMA defines data quality as "the application of data management techniques for planning, implementing, and controlling management activities to ensure that the needs of data consumers are met," and believes that data quality management is a continuous task that involves the entire data lifecycle.

In terms of data quality dimensions, DAMA-DMBOK2 summarizes common data quality dimensions, including accuracy, completeness, consistency, integrity, reasonability, timeliness, uniqueness/deduplication, validity, etc.

In terms of data quality business rule types, DAMA-DMBOK2 has summarized some common business rule types, such as definition consistency, data existence and record completeness, format compliance, value range matching, range consistency, mapping consistency, consistency rules, accuracy verification, uniqueness verification, timeliness verification, etc.

In terms of data quality improvement, DAMA-DMBOK2 points out that the most commonly used method is the Deming Cycle, which is the P-D-C-A (Plan-Do-Check-Act) model. From the perspective of organizational optimization, data quality in the entire data lifecycle should be ensured, rather than improving data quality in the process.

### 2.2. The ten steps process for data quality

The ten steps process of data quality provides a practical and feasible approach for enterprises to improve data quality. It is independent of business and provides clear action guidelines for implementing data quality management by selecting appropriate steps, actions, processes, and technologies. It can be applied in many situations<sup>[2]</sup>. The ten steps process for data quality includes the following steps: Defining business need and

\* Corresponding author: [xcfw@epri.sgcc.com.cn](mailto:xcfw@epri.sgcc.com.cn)

approach; Analyze information environment; Assess data quality; Assess business impact; Identify root causes; Develop improvement plans; Prevent future data errors; Correct current data errors; Implement controls; Communicate actions and results.

In step 3, a detailed introduction was given to the dimensions of data quality, including data specifications, data integrity fundamentals, duplication, accuracy, consistency and synchronization, timeliness and availability, ease of use and maintainability, data coverage, presentation quality, perception, relevance and trust, data decay, transactability, etc.

### 2.3. Data Management Capability Maturity Assessment Model (DCMM)

National standard GB/T36073-2018 data management capability maturity assessment model, DCMM, is the first national standard officially released in the field of data management in China<sup>[3]</sup>, which provides a maturity assessment model for data management capabilities and corresponding maturity levels. It defines 8 capability areas and 29 capability items, including data strategy, data governance, data architecture, data application, data security, data quality, data standard, and data lifecycle.

In the data quality capability area of DCMM, there are four capability items: data quality requirements, data quality inspection, data quality analysis, and data quality improvement. Starting from the requirements of data quality, determine the data quality objectives and corresponding data quality rules. Then, through data quality inspection, identify data quality problems, analyze the data quality problems, identify the causes, and finally improve the data quality based on the results of data quality analysis.

### 2.4. Comparative analysis of data quality management theories

From the above theories, it can be seen that the three theories have different emphases. As can be seen from Table 1, DAMA-DMBOK2 focuses on the problems of the theoretical system, the ten steps process focuses on the problems of method practice, and DCMM solves the problems of evaluation.

**Table 1.** Comparative Analysis of Data Quality Management Theories.

| Theory                | Release time | Key issues to be solved                                |
|-----------------------|--------------|--------------------------------------------------------|
| DAMA-DMBOK2           | 2020         | The overall framework of data quality management       |
| The ten-steps process | 2010         | Implementation process for improving data quality      |
| DCMM                  | 2018         | Evaluation of maturity of data management capabilities |

Compared to others, DAMA-DMBOK2 provides a more comprehensive explanation of data quality

management, not only with authoritative theories but also practical experience, which can provide framework guidance for enterprise data management.

The ten steps process is a clear guide for planning and executing information quality improvement projects, combining data quality dimensions and tool technologies to comprehensively evaluate data and information quality. A business independent system methodology is provided to improve information quality, with a greater emphasis on specific implementation processes.

The data management capability maturity assessment model (DCMM) is the first national standard officially released in the field of data management in China. The DCMM model is used to define the maturity level of organizational data management capabilities, providing a basis for the construction and improvement of organizational data management capabilities.

## 3. Typical case practice of data quality

### 3.1. National Bureau of Statistics

In order to further improve the quality of government statistical data, the National Bureau of Statistics has formulated the National Statistical Quality Assurance Frameworks(2021),which is used to guide statistical work in various departments<sup>[4]</sup>.

In the aspect of comprehensive evaluation of statistical data quality, the framework is based on the whole process of statistical data production and measures from nine aspects, including facticity, accuracy, completeness, timeliness, applicability, economy, comparability, coordination and availability.

At the same time, it emphasizes the whole process control of statistical quality, including quality control in ten aspects: determining needs, survey design, approval and filing, task deployment, data collection, data processing, data evaluation, data publication and dissemination, statistical analysis and project evaluation.

In terms of measures to ensure statistical quality, it is emphasized that it is necessary to strengthen the construction of the rule of law in statistics, improve the statistical system and mechanism, standardize the statistical system and methods, optimize the allocation of statistical resources, consolidate the basic level foundation of statistics, strengthen information technology support, promote the statistical culture with quality as the core, and promote the scientific, effective, and comprehensive governance of statistical quality.

### 3.2. Huawei

Huawei has started data governance since 2007<sup>[5]</sup>. After two stages of development, it has established a systematic Huawei data management system. The first stage was from 2007 to 2016, through the establishment of relevant data management professional organizations and unified standards, the continuous improvement of data quality and the integration of the entire data process were achieved; The second stage is from 2017 to present. Based on the development of the first stage, Huawei has

aggregated and connected global data, achieving value in business visualization, business automation, and data innovation.

The first is to establish a company-level data quality management policy. Huawei's data quality management policy not only clarifies the data quality management responsibilities and corresponding requirements, but also standardizes the business rules and requirements of data quality management.

The second is to design a data quality management framework based on PDCA. Based on the ISO8000 quality standard system, Huawei has designed a continuously improved PDCA (Plan, Do, Check, Action) data quality management framework. The framework starts from the business needs, under the support of leadership and capacity guarantee, based on PDCA to continuously improve the data quality, and finally achieve the goal of data cleaning.

The third is to monitor business abnormal data based on data quality rules. According to the characteristics of data quality and the types of data quality rules, Huawei designed four types of data quality frameworks, including single-column data quality rules, cross-column data quality rules, cross-line data quality rules, and cross-table data quality rules, including 15 types of rules, covering six characteristics of data quality, including integrity, timeliness, effectiveness, consistency, uniqueness and accuracy. At the same time, the data quality rules will be constantly updated with the deepening of data governance.

The fourth is to drive quality improvement through comprehensive evaluation of data quality. Huawei comprehensively evaluates the overall data quality level of the company through data quality measurement, and measuring it from two aspects: 'design' and 'execution'. The 'design' aspect mainly measures the data asset catalog, data standard, data model and data distribution, and the 'execution' aspect mainly measures the six properties of data quality.

### 3.3. Banking Financial Institutions

In 2018, China Banking and Insurance Regulatory Commission issued the "Guidelines on Data Governance of Banking Financial Institutions Issued" [6], which is used to guide banking financial institutions to strengthen data governance, improve data quality, give full play to data value, and improve business management capabilities. The guidelines include 7 chapters and 55 articles, which clarify the data governance structure, data management, data quality control, data value realization and so on.

In terms of data quality control, it is emphasized that data source management should be strengthened and a data quality monitoring system for the entire data lifecycle should be established. At the same time, a data quality assessment system should be established, and the assessment results should be included in the performance appraisal system of the organization to continuously improve the data quality.

## 4. Data quality management experience summary

The first is to establish and improve the data owner system, deepen the data source and responsibility. Focus on the key business data of the enterprise, clarify the authoritative data source and its management responsibility, and publish a list of authoritative data sources to ensure that 'uniqueness of data source'. Comprehensively promote the construction of 'data owner system', establish a data source control mechanism of 'who produces, who is responsible, who applies, who supervises', and continuously promote full coverage of business data source and responsibility determination.

The second is to clarify the scope of data quality management and establish a closed-loop control mechanism for data quality. By using methods such as master data identification, expert group analysis, and questionnaire survey, the scope and objects of data quality management are defined, and combined with business driving forces, data quality management work is continuously carried out through classification and grading. Based on the entire lifecycle of data, establish a closed-loop control mechanism for data quality, achieve full chain data quality monitoring and auditing capabilities, detect and rectify early, and comprehensively improve data quality.

The third is to standardize data standard management and build a data quality rule verification library. Data standards are an important foundation for achieving high-quality data, and through unified data standards, the quality of source data can be effectively guaranteed. The standardization of data standards should not only consider relevant national and industry standards, but also fully consider the characteristics of enterprise data to support the implementation of data standards. Data quality rules are an important means of determining whether data meets data quality requirements. Through data quality rule verification, data quality issues are discovered and promptly rectified, thereby effectively improving data quality.

The fourth is to build a comprehensive evaluation system of data quality management to promote the overall improvement of data quality. Based on the entire data lifecycle and the practical business needs, a comprehensive evaluation system of data quality is constructed to comprehensively evaluate the overall data quality management level of enterprises, find out the shortcomings of data quality management and improve them, so as to help break through the effect of data quality management and promote the overall improvement of data quality.

## 5. Conclusion and Prospects

Data quality management is the core of data governance, and high-quality data is the foundation for realizing the value of data. On the one hand, this article reviews and compares the DAMA-DMBOK2, the ten steps process, and DCMM in data management theory, and finds that

the three theories have different emphases. DAMA-DMBOK2 focuses on the problems of the theoretical system, the ten steps process focuses on the problems of method practice, and DCMM solves the problems of evaluation. On the other hand, this article summarizes four valuable experiences based on typical practices of the National Bureau of Statistics, Huawei, and banking financial institutions.

When enterprises carry out data quality management, they need to establish a comprehensive data quality evaluation system based on data characteristics and the current situation of enterprise data quality management. When summarizing the current situation of enterprise data quality management, it is necessary to focus on sorting out the types of enterprise data, analyzing the difficulties of enterprise data management, and conducting current research and problem summary on the existing data quality management work of the enterprise, deeply grasping the key points of data quality management, and providing direction for the improvement of subsequent data quality management.

Based on the entire data lifecycle and the workflow of enterprise data quality management, focusing on the core issues of data quality evaluation, drawing on the experience of internal and external data quality management in the system, constructs a comprehensive evaluation system for enterprise data quality audit rules and data quality governance work, comprehensively, systematically, and objectively evaluating enterprise data quality management work. It is very important to construct a comprehensive system of data quality evaluation, which can not only understand the current data quality management ability of enterprises, but also provide a basis for subsequent data quality improvement. In the comprehensive evaluation system of data quality, data quality audit rules are an important means to ensure data quality and a key core of data quality management. The evaluation of data quality governance is the overall evaluation of each stage of the entire data lifecycle. It is an important basis for guiding data quality work and the overall coordination of data quality management work. On the basis of strengthening the design of data quality audit rules, a comprehensive evaluation of data quality governance can effectively promote the improvement and efficiency of data quality management, enhance data utilization efficiency, and realize data value.

In the construction of a comprehensive evaluation system for data quality, the following principles should be adhered to: firstly, combining goal orientation with problem orientation. Adhere to the consistency between the evaluation work objectives and the data quality management objectives and requirements, highlight the difficulties and pain points in data quality management work, and motivate departments to conduct self-evaluation and optimization improvement. The second is to emphasize the combination of process and result guidance. Adhere to the principle of balancing process and results, evaluate data quality management process control while quantitatively evaluating the construction effectiveness of each link, and fully leverage the role of evaluation process control and result guidance. The third is to emphasize the combination of quality improvement

and management optimization. Adhere to the equal emphasis on quality improvement and management optimization, while enhancing data quality, strengthen collaboration and integration between internal business departments. The fourth is to clarify the combination of innovation and implementation. The work of data quality management should not only focus on exploring and innovating based on business needs and data characteristics, but also pay attention to the implementation of innovative results. It should have both the scientificity of theoretical innovation and the feasibility of implementation.

In terms of data quality audit rules, we can refer to the national standard "information technology-Evaluation indicators for data quality" (GB/T 36344-2018) to set the enterprise data quality audit rules differently. In the comprehensive evaluation section of data quality governance, evaluation indicators can be designed based on various aspects of the entire data lifecycle, including but not limited to data collection, data storage, data processing, data services, data sharing, etc. In comprehensive evaluation, it is necessary to first combine expert interviews to form an evaluation index system for enterprise data quality governance work; Then, the weights of each indicator are determined using Analytic Hierarchy Process (AHP) and other methods; Finally, a comprehensive analysis is conducted to obtain the evaluation results, providing a reference basis for enterprise data quality management work.

## References

1. DAMA International. DAMA-DMBOK2[M]. Translated by DAMA China Branch. Beijing : China Machine Press, 2020.
2. Danette McGilvray. Executing Data Quality Projects—Ten Steps to Quality Data and Trusted Information[M]. Translated by Xingchun Diao, Jianjun Cao, Jianmei Zhang et al. Beijing : Publishing House of Electronics Industry, 2010.
3. GB/T36073-2018,Data Management Capability Maturity Assessment Model[S].
4. National Bureau of Statistics. National Statistical Quality Assurance Frameworks(2021)[EB/OL]. [http://www.stats.gov.cn/sj/zxfb/202302/t20230203\\_1901137.html](http://www.stats.gov.cn/sj/zxfb/202302/t20230203_1901137.html), 2021-06-17.
5. Huawei data management department. Enterprise Data at Huawei[M].Beijing: China Machine Press, 2020.
6. China Banking and Insurance Regulatory Commission. ' Guidelines on Data Governance of Banking Financial Institutions Issued'[EB/OL]. [https://www.gov.cn/xinwen/2018-05/23/content\\_5292938.htm](https://www.gov.cn/xinwen/2018-05/23/content_5292938.htm),2018-05-23.