Construction and Practice of Network Data Lifecycle Management System--Based on Metadata-driven Quality Control Strategy

ChenLi

China Science Publishing & Media Ltd. Beijing 100717

Abstract: This paper focuses on the whole life cycle management of network data, and puts forward the quality control strategy based on metadata drive.By dividing the data life cycle into stages, building a multi-dimensional quality indicator system and designing an integrated framework of "strategy-process-tool-supervision," the Bank will promote the transformation of data quality governance from static control to dynamic closed-loop control.After the empirical application to the smart government platform, the data problem discovery rate, response efficiency and automatic alarm coverage rate have been significantly improved, which verifies the feasibility and effectiveness of the system.The research provides the method support and practice path for the construction of standardized and intelligent data governance mode.

Key words: Network data; Life cycle management; Metadata; quality control; Data governance

Introduction

With the accelerated development of digital economy, network data has become a key strategic resource in smart city, intelligent manufacturing and other scenarios.As the core factor of production, the scientific management of data in its whole life cycle has become the key path to ensure the availability and governance quality of data.At present, network data usually flows across platforms and systems, and its life cycle covers multiple stages of generation, storage, processing, transmission, sharing and destruction. The quality of each link is out of control frequently, resulting in serious data fragmentation and limited value release.Especially under the background of increasing trend of data source heterogeneity and structure complexity, the traditional static and segmented data management mode can not meet the high requirements of real-time, consistency and standardization in multi-service scenarios.Therefore, constructing a set of network data management system with dynamic regulation and control ability through the whole process of life cycle becomes the core task of promoting the standardized governance and value promotion of data resources.Based on the practical governance problems, this paper tries to explore and construct a systematic and closed-loop life-cycle management mode from the perspective of metadata-driven, so as to meet the systematic challenges of data quality assurance under the new situation.

1 Literature review and theoretical basis

1. 1 Research status of network data management

With the evolution of information technology, the scale and complexity of network data are growing exponentially. How to realize the scientific and systematic management of network data has gradually become an important research direction in the field of data science [1].The early research mainly focused on the storage and transmission security of data, but in recent years, it turned to the controllability and quality assurance of data in the whole life cycle.Domestic scholars have put forward various framework models in data resource integration, data hierarchical and classified management, data sharing and exchange mechanism, but no unified standard has been formed yet; Abroad, data management models have been explored earlier, such as DAMA Data Management Knowledge Body Guide (DMBOK), which puts forward core knowledge domains such as data governance, data quality, data architecture, etc., which provides theoretical reference for enterprise-level data management.Research shows that the lifecycle data management can effectively solve the problems of quality control in different stages and frequent data islands.With regard to data quality evaluation methods, academic circles have evolved from initial static indicators such as accuracy and integrity to dynamic and visual quality evaluation, such as multi-dimensional data quality evaluation model, interpretable quality quantification method, etc., especially under the background of big data and artificial intelligence, higher requirements are put forward for the management and control of data source, processing path and semantic consistency [2].On the whole, although the theoretical framework of network data management is increasingly rich, the research on quality control in the whole process of life cycle is still weak, and most of them stay at the concept or local application level, so it is urgent to make systematic breakthrough from the methodology and tool platform.

1. 2 Metadata-driven data governance strategies

Metadata, as an important information carrier describing the characteristics, structure, source and use rules of data, plays a fundamental role in improving the efficiency of data management and standardizing data behavior [3].In recent years, more and more studies started to pay attention to the function embedding and policy collaboration of metadata in data governance.In the aspect of data traceability, metadata can record the generation path and circulation track of data, and provide reliable basis for the subsequent traceability of quality problems.In the aspect of standardization, metadata helps enterprises to realize data consistency management in multi-system and multi-platform environment by defining unified data naming convention, data format and data dictionary.In the aspect of consistency maintenance, metadata supports the coordinated update of data in different links by constructing data kinship and dependency mapping, so as to avoid system-level quality degradation caused by local changes.In addition, metadata also has the technical foundation of constructing quality rule base and automatic evaluation mechanism, which becomes the key support of quality closed-loop control.At present, some enterprises have built a data management system based on metadata platform, integrating metadata collection, management and calling capabilities into the data management platform, and realizing the automatic calculation of quality indicators and abnormal alarm functions.However, on the whole, the practice of metadata in our country is still limited to asset registration and data directory management, and its potential in data quality control, policy-driven and life cycle connection has not been fully realized.Therefore, constructing the data governance mechanism driven by metadata as the core becomes an important breakthrough to improve the closed-loop management ability of data quality.

1.3 Theoretical foundation of life cycle management model

Data lifecycle management is an important part of modern data governance, and its theoretical basis comes from multiple authoritative models [4].The DMBOK model proposed by DAMA International Association defines 14 knowledge domains of data management, and regards life cycle as the main line of governance running through each link; the other widely cited DCAM (Data Management Capability Assessment Model) model emphasizes taking capability as the core and constructing governance ecology around four dimensions of data governance strategy, process, organization and technology.A common feature of these models is that data is treated as a dynamic asset, and managed from source to archive by setting management goals and quality standards at each stage.At the level of life cycle management practice, some studies advocate the adoption of a five-stage process of "data discovery-quality assessment-value mining-asset registration-compliance audit," and some scholars propose a minimalist closed-loop structure of "production-use-reduction-deletion."Based on the above theoretical model, this paper proposes a data lifecycle management system with metadata as the main axis, which embeds metadata into the management process of each stage to realize the synchronous driving of policy and process, and the dynamic collaboration of quality and security.The system divides the life cycle into six core stages: generation, integration, application, sharing, maintenance and termination, which are matched with metadata elements and quality control mechanisms respectively, thus constructing a closed-loop management logic of "standard-collection-monitoring-feedback," which not only makes localization transformation on the existing theoretical model, but also provides a feasible technical path and organizational framework for data governance practice.

2 Design of network data lifecycle management system

2.1 Lifecycle phasing definition

Network data life cycle management should be from a global perspective, clear in the actual business and system operation in the experience of the core stage, to ensure that each link of the data quality and governance requirements can be implemented in detail [5].In the network system, the life cycle of data can be divided into seven stages: generation, transmission, storage, processing, sharing, archiving and destruction.The generation stage includes the initial formation of data in terminal equipment, sensors and business systems, which has strong source attribute; the transmission stage refers to the flow process of data among different systems, platforms and nodes, which has extremely high requirements for data integrity and security; the storage stage mainly reflects the landing and retention of data in structured or unstructured databases, which determines the accessibility and persistence of subsequent data; The processing stage involves data cleaning, conversion, analysis and modeling, which is the key node for data value release; The sharing stage emphasizes cross-department and cross-platform data scheduling capability and interface standard; The archiving stage pays attention to the long-term retention of data, historical record integration and query capacity building; The destruction stage marks the end of the data life cycle, and strict mechanisms shall be adopted to ensure that the data is unrecoverable and cancelled in compliance.The above stages are not only the logical sequence of data life cycle operation, but also the important anchor points for implementing data governance measures and embedding quality control strategies, providing a clear path for the establishment of subsequent management system.

2. 2 Overall structure of management system

The construction of the network data lifecycle management system needs to take "strategy + process + tool + supervision" as the core architecture, forming a closed system loop from system design, business implementation to dynamic feedback.In the strategic dimension, formulate a quality strategy system covering data standards, classification and grading, security levels and life cycle rules, and define the management requirements of different data types at different life cycle stages; in terms of process design, construct a standardized operation path covering the whole process to ensure that there are clear operation standards and quality checkpoints for each step from data generation to destruction; At the tool level, digital governance tools including data asset platform, metadata management system, data quality monitoring tools, life cycle process engine, etc. shall be introduced to improve the automation and refinement level of management; the supervision dimension shall ensure the execution force and landing of the whole life cycle management work by establishing normal monitoring mechanism, problem feedback channel and assessment and evaluation system.The system takes data quality as the main line, technology platform as the support, management system as the drive, logically emphasizes the system leader and technical support, and in practice, it is embodied as a continuous optimization process throughout the whole life cycle.

2. 3 Metadata driven mechanism embed

In the life cycle management system, metadata is not only used as the information description tool, but also should be given the driving role.To achieve this goal, a metadata-based management model needs to be built that runs through all phases of the lifecycle.the method comprise that following step of: starting a metadata registration mechanism in a data generation stage to ensure that all new data have a source identifier, a structure description and business semantics; apply a metadata mapping and conversion mechanism in a data transmission and sharing process to realize dynamic adaptation of data formats and semantics among heterogeneous systems; and in a data processing and analysis link, realizing automatic identification of process dependency, accurate configuration of processing rules and visual tracking of data kinship by relying on metadata. In the archiving and destruction phase, metadata is used to define the retention period and destruction method, so as to ensure the compliance and security of data disposal.In addition, it is necessary to build a set of metadata collection and update mechanism, so that it can automatically sense data changes and timely feed back to various management tools to realize policy response and dynamic adjustment.through that embed of the mechanism, the metadata is no longer limit to a carrier of static description information, and becomes a key hub for connecting a strategy system and an execution proces, so that the whole life cycle management is really promoted to be intelligent, closed-loop and sustainable.

3 Metadata-driven quality control strategy design

3. 1 Metadata quality dimension construction

In the lifecycle management of network data, metadata not only undertakes the basic functions of description, identification and tracking, but also drives the implementation of quality control strategies.In order to ensure that metadata has enough governance ability, it is necessary to measure its quality level from multiple angles and construct a scientific and reasonable index system.Based on the actual governance experience and industry standards, this paper constructs a metadata quality evaluation model from the four aspects of technology, business, semantics and security, refines the key indicators, and supports the quality judgment and anomaly monitoring of different links.

Table 1 Metadata Quality Dimensions and Key Indicator System

dimension type	quality index	Key Meanings
technical dimension	Completeness, timeliness, consistency and accuracy	Describe whether the metadata is complete, updated timely, and synchronized and consistent across systems
service dimension	Business Fit, Role Adaptability, Scenario Coverage	Judge whether metadata can effectively map actual business and reflect difference business requirements.
semantic dimensions	Naming conventions, monosemous multi-word rates, terminology consistency	Avoiding ambiguity in terms and improving semantic consistency and readability across multiple systems
security dimension	Access control granularity, sensitive field recognition rate, audit frequency	Enable access control and operational auditing of sensitive metadata to ensure compliance and security

It can be seen from Table 1 that the indicator system emphasizes the comprehensive evaluation from structural attributes, business rules, semantic standards to security control, which not only provides a basis for measuring the quality of metadata ontology, but also provides a standard reference for subsequent monitoring and policy implementation, and realizes the functional transformation of metadata from "descriptive attributes" to "governance engine."

3. 2 Quality monitoring and alarm mechanism

In the process of implementing quality management, we must build a monitoring and warning system which takes metadata as the center and can run continuously.By embedding the blood relationship information, the field dependent path and the processing node identifier in the metadata, a logically complete quality rule network can be constructed, and the automatic detection of abnormal changes can be realized.When the system identifies problems such as field redundancy, primary key missing, inconsistent data caliber, etc., it will automatically compare the rule base and push the abnormal results to relevant responsible persons through the alarm center.The system will also be attached with exception fields, table locations, problem impact paths and suggested repair methods, effectively improving the timeliness of data quality problem response and positioning accuracy.The platform integrates metadata dictionary and standard field template, which can push batch repair suggestions, reduce manual intervention and improve governance efficiency.Through the structure mapping of metadata and the synchronization of semantic rules, the monitoring mechanism can not only cover the conventional field rules, but also monitor the upstream and downstream consistency based on business logic to ensure the stability and controllability of data quality in the whole chain process.

3. 3 Quality closed-loop control process

It is difficult to construct a truly efficient governance system only relying on static rules and alarm push. It is also necessary to introduce a closed-loop control mechanism driven by metadata, embed quality management into each key node of data life cycle, and realize dynamic linkage of five links: standard definition, real-time detection, problem positioning, repair response and audit feedback.In the data generation stage, the system completes the basic configuration such as field structure, naming standard and authority control based on metadata; in the processing and sharing stage, the metadata rule base is continuously called for logic verification and blood relationship confirmation; in the use process, dynamic risk assessment and access behavior monitoring are performed on key fields; in the archiving or destruction stage, the system automatically executes audit and data cleaning operations with reference to metadata life cycle marks to ensure compliance and integrity of historical retained records.The governance platform also supports the automatic generation of quality reports, and counts the number of problems found in each stage, the treatment efficiency, the success rate and the influence scope, so that managers can examine the current strategy effect and adjust the direction.Through this closed-loop mechanism, data quality control has shifted from static review to dynamic adjustment, and from manual follow-up to intelligent response, forming a full-process closed-loop covering "pre-event regulation, in-process monitoring and post-event feedback," realizing visualization, automation and refined management of data governance.

4 Practical cases and application validation

4. 1 Case background

In the current transformation of government data governance, a provincial smart government affairs middle-station project, as a typical case, faces severe challenges such as complex data sources, numerous systems and prominent historical problems.The platform needs to integrate the business data of more than 120 commissions and bureaus in the province, covering population, education, medical care, social security and other fields.Due to heterogeneous source systems, scattered standards and inconsistent calibre, frequent data quality problems, field missing, redundancy, naming conflict, update lag and other situations occur, which seriously affect the credibility of data sharing and the accuracy of intelligent approval.The original quality management system mainly focuses on manual spot check, lacks automatic detection and cross-platform linkage capability, and cannot meet the high-frequency dynamic data governance needs.Therefore, the introduction of metadata-based quality management system for the whole life cycle becomes the key to break the project.

4. 2 Implementation of management system

The project construction is divided into four stages: system design, platform construction, process reconstruction and tool integration.At the level of system design, the data life cycle is divided into six stages and corresponding quality control strategies are set up; in the aspect of platform construction, a unified data asset platform, metadata management system, quality monitoring module and data consanguinity analysis engine are constructed; the process reconstruction stage is reconstructed around standard formulation, quality review, alarm push and closed-loop repair; In terms of tool integration, introduce domestic mainstream data governance platform to open up the governance channels of data collection, monitoring, feedback and audit.

Table 2 Main Function Modules and Functions of the Data Lifecycle Management Platform

function module	core function	main function
metadata collection engine	Automatically identify system fields, tables, structures, and their changes	the system realize unified access of metadata of multi-source data and keep real-time synchronization of that metadata
Data Quality Monitoring Center	Execute quality rules, identify abnormal data, and push alarm information	Improve the efficiency of quality identification, support rapid response and problem location
data kinship analysis tool	Map dependencies between fields	Support for upstream and downstream impact analysis and visualization of data propagation paths
Data Governance Workbench	Summarize monitoring results, generate reports and	Support governance performance evaluation and

	visualize	audit traceability, and improve governance transparency and controllability

It can be seen from Table 2 that the platform functions comprehensively cover the governance requirements of each stage of the data life cycle, especially in the aspect of metadata-driven linkage governance, realizing the automatic closed-loop from problem discovery to feedback processing, effectively improving the governance efficiency and visibility of governance scope.

4. 3 Evaluation of implementation effectiveness

Half a year after the launch of the project, the platform compares the changes of key indicators before and after the implementation, and the system automatically generates periodic quality analysis reports.By comparing indicators such as problem discovery rate, repair response time, and automatic alarm coverage, it is clear that the governance effect brought about by metadata-driven strategy has been significantly improved.

Table 3 Comparative Analysis of Core Quality Indicators Driven by Metadata

quality index	before the implementation	Post-implementation	improvement
Data problem discovery rate (piece/month)	410	1260	↑207%
Average Problem Fix Response Time (hours)	36	12	↓66.7%
Automatic alarm processing coverage	18%	75%	↑317%
high risk field recognition accuracy	62%	91%	↑29%

Table 3 shows that the identification speed, handling efficiency, problem coverage and accuracy of data governance have been significantly improved after the introduction of metadata-driven mechanism.especially in that aspect of quality assurance of high-frequency share fields and key service field.The improvement of system automation governance capability not only reduces the pressure on operation and maintenance, but also strengthens the closed-loop governance of pre-warning, in-process control and post-event feedback, promotes the transformation of government data governance to intelligence, standardization and high efficiency, and provides replicable implementation paradigm and evaluation baseline for projects of the same type.

5 Conclusion

Focusing on the system construction and practical application of network data lifecycle management, this study puts forward a quality control strategy driven by metadata, aiming at solving the outstanding problems existing in current data management, such as fragmented governance, lack of standards, uncontrollable quality and so on.Through the fine division of each stage of the life cycle, the whole process management model covering the generation, processing, sharing, archiving and destruction of data is constructed, and key mechanisms such as metadata collection, parsing, blood tracing and permission control are embedded to realize the high integration of policies, processes and tools.At the policy level, the metadata quality dimension and rule system provide unified basis for data evaluation; at the execution level, the dynamic monitoring and automatic alarm mechanism significantly improves the problem identification efficiency and response ability; on the closed-loop control, a complete governance chain integrating standard formulation, real-time detection, problem repair and report generation is formed.Through the empirical application in the smart government platform, the relevant indicators such as problem discovery rate, repair efficiency and alarm response speed are greatly improved, and the effectiveness and replicability of the system in the actual environment are verified.On the whole, this research provides a feasible technical path and practical paradigm for promoting network data governance to be standardized and intelligent.

References

[1]Dong W ,Sun K ,Ding B , et al.Exploration and practice of high speed EMU digital train based on digital twin[J].High-speed Railway,2025,3(01):54-63.

[2]Xie J ,Sun L ,Zhao F Y .On the Data Quality and Imbalance in Machine Learning-based Design and Manufacturing——A Systematic Review[J].Engineering,2025,45(02):105-131.

[3]Guo P ,Xiao K ,Wang X , et al.Multi-source heterogeneous data access management framework and key technologies for electric power Internet of Things[J].Global Energy Interconnection, 2024,7(01):94-105.

[4]Du B ,Feng B .An Analysis of Policy Ecology of Scientific Data under the Dual Three-dimensional Framework[J].Innovation and Development Policy,2023,5(02):167-194..

[5]Farahani B ,Monsefi K A .Smart and collaborative industrial IoT:A federated learning and data space approach[J].Digital Communications and Networks,2023,9(02):436-447.

Introduction：Li Chen, gender: female, born on December 2, 1984, ethnicity: Han, native place: Hebei Province, professional title: Intermediate Publishing, education: Master's degree, research direction: Computer Science and Technology