

Research on the Management Pattern of Data Governance for the Ground Gathering and Transportation System of Sichuan Natural Gas Well Site

Tianyou Fan*, Wen Cheng, Haohai Wang, Liangchen Lv

Shunan Gas Mine, Southwest Oil & Gas Field Company, PetroChina, Luzhou, 646000, China

Abstract

The Sichuan Natural Gas Well Site is characterized by its large-scale engineering and highly complex business management. How to achieve comprehensive management of gas stations, pipelines, and business chains is the key to data governance for the ground gathering and transportation system. This paper proposes a management pattern based on the expansion of existing information platform modules to establish a comprehensive basic database. It also introduces early warning functions for regular inspection and rolling maintenance of equipment and facilities, which unify the basic data of the ground gathering and transportation project, improve the efficiency of data query and analysis, enhance the level of integrity management, and elevate the development and management level of the ground gathering and transportation system.

Keywords

Natural Gas Well Site Engineering; Ground Gathering and Transportation System; Data Governance; Data Lake; Management Pattern.

1. Introduction

In recent years, data resources have increasingly become a key element in the economic development artery, and the penetration of big data in various fields of society has deepened significantly. It has become a strategic asset of vital importance to China's economic development [1]. The petroleum industry is a strategic and fundamental pillar industry of a nation. The Southwest Oil & Gas Field, as the cradle of China's petroleum industry, has undergone more than 50 years of exploration and development. As new oil and gas reservoirs become smaller in scale and more difficult to exploit, petroleum exploration and development enterprises are facing significant challenges. To better address these challenges, many petroleum exploration and development enterprises have begun digital transformation reforms. During the digital transformation process, with the rapid development of digital oil fields, the petroleum industry has developed a higher demand for data quality. High-quality big data resources can significantly enhance the decision-making level in the oil and gas energy sector, thereby promoting the development and innovation of the national energy sector [2].

At present, the attitudes of many petroleum enterprise executives towards digital transformation are mainly focused on technological innovation, especially data mining and analysis [3]. Therefore, to further deepen digital transformation, it is essential to continuously improve technological innovation to address the numerous challenges currently faced. As petroleum production increases, a large amount of petroleum data is generated, which holds immense information and commercial value. To better extract the value from these data, data mining and analysis are required as support. However, due to the multi-source nature of data collection from different environments and varying data integration methods, it is easier for instance data to contain similar duplicate records, missing values, domain violations, and business rule violations [4]. The generation of these low-quality "dirty data" directly leads to

computational deviations and can have unimaginable consequences. For example, similar duplicate records in the database occupy a large amount of space, directly affecting the efficiency of database usage. If these data, which contain numerous data quality issues, are directly used for data mining experiments, it will inevitably lead to significant deviations in the mined and analyzed information. In severe cases, this may cause substantial economic losses to enterprises. To address these issues, various petroleum enterprises have begun to develop their own data cleaning systems. However, due to the different sources of data, the business data cleaned by each enterprise have different patterns and models, resulting in more limitations in these data cleaning systems, as well as weaker universality and scalability [5].

In addition, another important factor in data quality is the standardization of data cleaning operations. Since all data governance methods are highly dependent on domain experts, these experts often lack certain specific information technology knowledge, which may be required for specific data governance operations [6]. Meanwhile, information technology developers, although possessing the complete technical knowledge required for standardized data governance operations, lack relevant domain knowledge, preventing them from fully developing a universal data governance operation standardization process [7].

Generally, information collected by various ground gathering and transportation systems can be measured. However, these are discrete data, and it is difficult for workers in the oil and gas field gathering and transportation system to intuitively describe the relationships between data at the macro level, ensure the correctness and consistency of data across the entire system, accurately track the flow of data between subsystems, or understand the historical changes of data over time [8]. Faced with these measured data, people inevitably ask: How can these measured data be used to infer their distribution patterns in the region? Even if it is possible to predict the distribution values of various information in the studied area, petroleum engineers still face difficulties in analyzing the large amount of collected data. Moreover, different development ground systems have different standards, and it is necessary to establish a unified underlying database based on different system construction backgrounds to enable data extraction and classification under different standards. At the same time, considering the difficulties and time-consuming nature of manually querying equipment archives, it is essential to address the key engineering and technical problems of data governance, such as ensuring data integrity, improving data utilization and query analysis efficiency, by entering the basic data of the ground gathering and transportation project into an information platform [9].

In response to the above issues, this study proposes to rely on an information platform to create early warning functions for regular inspection and rolling maintenance of equipment and facilities. This can significantly reduce the risk of equipment inspection omissions and overdue inspections caused by manual queries. It is of great significance for risk identification in oil and gas field gathering and transportation systems and for the improvement of the HSE management system. Therefore, data governance has become a key element in the data application of oil companies [10]. The data governance module can standardize data governance according to established rules, thereby greatly improving data standardization and quality [11]. This research is an inevitable trend in the application of computer technology, database technology, and data governance theory in the field of petroleum engineering. As a fundamental project, data governance can break down data silos and business barriers, enhance data governance capabilities in digital transformation, and promote the high-quality development of the petroleum industry.

2. Research Status of Data Governance in Petroleum Engineering

Data governance is a set of management behaviors involving the use of data. It is a series of policies and processes initiated and promoted by the enterprise data governance department,

aimed at how to formulate and implement commercial applications and technical management of data within the entire enterprise. The Data Management Association has provided the following definition: Data governance is a collection of activities that exercise power and control over data asset management. The Data Governance Institute defines it as a system that realizes decision-making power and responsibility division through a series of information-related processes. These processes are executed according to a consensus model that describes who can take what actions, based on what information, under what circumstances, and by what methods ^[12]. Abroad, IBM began exploring data governance theory and putting it into practice as early as 1993. After continuous improvement, it has now formed a relatively complete data governance model ^[13]. Organizations with more mature data governance research include the Data Management Association (DAMA) and the Data Governance Institute (DGI). DAMA has published the Data Management Body of Knowledge Guide-DMBOK, which defines 10 knowledge areas of data management. The European Union, aiming to achieve a "data-driven economic strategy," has adopted the establishment of specialized data governance teams and the formulation of data standardization management methods to carry out data governance work, allowing data to generate value and become a driving force for the economic recovery and growth of the European Union. Chevron Corporation launched a data quality improvement and data governance project in 2009, mainly to address data quality issues such as system data inconsistency caused by multiple data sources and high decision-making analysis costs. By establishing a data governance system, standardizing data standards, clarifying data responsibilities, and building a comprehensive data control capability, the project has improved decision support rates ^[14]. Panian ^[15] believes that classifying all data analysis is a prerequisite for enterprises to establish suitable data governance models. H. Dutta ^[16] proposed a governance structure and process model based on attribute graphs, effectively governing the large amounts of multi-source unstructured data in the enterprise data lake. In summary, data governance research abroad began earlier, mainly because international oil companies are more independent and professional, so the requirements for data professionalism and independence are also higher. Their data construction is relatively more standardized and unified, and the problems are not very prominent. In general, foreign oil companies have their own characteristics in data construction. They pay more attention to the application of oil field data and are more advanced than us in China in terms of timely transforming data assets into value.

Domestic data governance in China started around 2010, mainly focusing on the construction of data warehouses. In April 2016, the Digital Oilfield Research Institute of Chang'an University proposed the concept of oilfield data governance, which is different from previous data organization and data warehouse construction. It advocates ensuring data quality from the source ^[17]. In April 2016, China National Petroleum Corporation (CNPC) made the decision to build a "Shared CNPC" at the Science and Information Technology Innovation Conference, establishing the development direction and target positioning of transforming information technology from centralized integration to shared intelligence, and initiating scientific data governance work. In 2018, Yanchang Petroleum Group completed the formulation of master data standards after four years, built a master data standard platform, and formed a master data management and maintenance system and related norms ^[18-20]. In 2018, Qingxin Oilfield established an oilfield data pool platform based on the concept and system of data governance engineering, using data pool technology and data governance tool technology as the foundation. This platform normalized the oilfield data chain, standardized data storage, and improved data management efficiency, leading to more intelligent data application ^[21]. In addition, companies such as Great Wall Drilling have implemented various governance methods, achieving significant results and greatly improving work efficiency ^[22]. In 2024, Northwest Oilfield comprehensively integrated petroleum exploration data, conducted petroleum resource

evaluation, organized existing petroleum data maps, and combined petroleum data governance technology to form a data management pattern. After detailed exploration data research, design, and review processes, a complete set of exploration data standards was established, and the data governance model achieved good results.

In summary, the production data generated in the petroleum production process of domestic oil companies involves multiple links such as logging, cementing, and drilling, with broad coverage and high professionalism. These links involve many departments, each of which has established several information systems. Although some achievements have been made in business data application analysis and data architecture, there are still many shortcomings. The main manifestations are as follows: (1) Lack of unified data management standards, with some data sources being inconsistent or incomplete; (2) Real-time and quality issues of data, with many data entries containing errors, omissions, and low quality, which restrict data analysis and application; (3) Non-standard data management, with a lack of unified data coding and management standards, resulting in numerous obstacles for data integration applications. These issues lead to low data application efficiency, which fails to provide effective data support for petroleum companies and hinders their transformation towards intelligent and digital operations. Therefore, comprehensive and efficient production data governance work is necessary.

3. Research Framework of Data Governance Project for Sichuan Natural Gas Well Site

The overall design of the research aims to address the question of "how the research should be implemented." Its main task is to identify the composition, connections, and definitions of the physical elements that make up the system. Based on the above user requirement analysis and combined with actual engineering conditions, the research framework and overall structure of the data governance of the ground gathering and transportation system of Sichuan **Natural Gas Well Site** are proposed, as shown in Figure 1, with detailed explanations as follows:

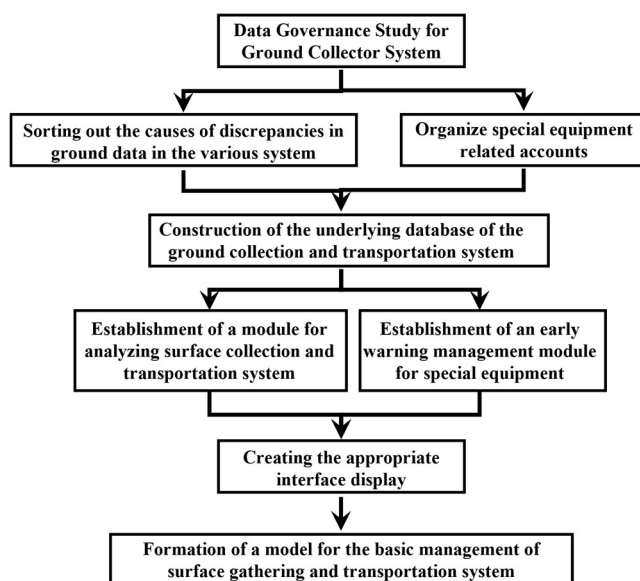


Figure 1. Research framework for data governance of surface gathering and transportation system in Sichuan Natural Gas Well Site

As shown in Figure 1, the research framework can be roughly divided into three major steps: (1) Conducting the cleaning of various development ground basic data of the **natural gas well**

site and investigating the connectivity of data in existing application systems; (2) Building a ground database and early warning reminder mechanism system; (3) Establishing a big data entry and data governance analysis and full-process management mechanism. The quality and completeness of the data collected for the ground gathering and transportation system determine the foundation for the system to be constructed as expected, which requires careful analysis and identification. The construction of the ground gathering and transportation system database is the basis and key for the system to be realized, and all other work is related to it. From both theoretical and technical feasibility perspectives, the establishment of the ground gathering and transportation system database is also a challenge that needs to be overcome. Relying on the existing electronic ID card platform to further develop equipment regular inspection and rolling maintenance early warning function modules is a further improvement and enrichment of data governance research. Finally, based on these efforts, a big data governance analysis and management plan for the ground gathering and transportation project will be formed, providing important basis for effectively guiding decision-making and design.

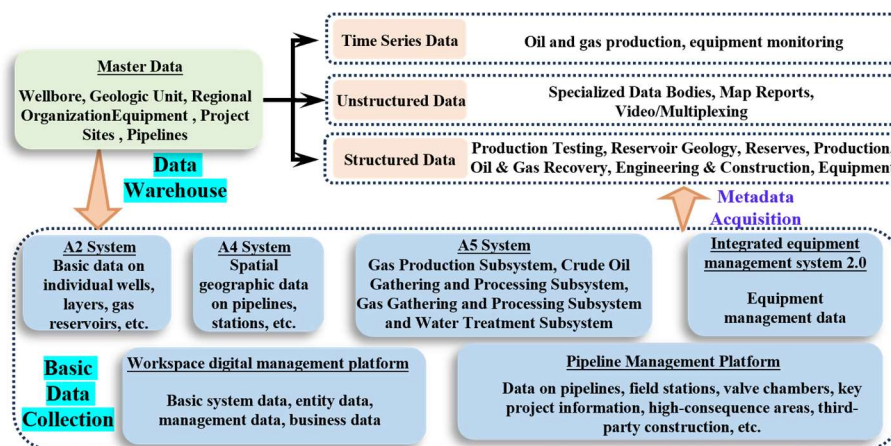


Figure 2. Sources of data collection for data governance of surface gathering system in Sichuan Natural Gas Well Site

As shown in Figure 2, the basic data collection for the data governance of the ground gathering and transportation system of Sichuan Natural Gas Well Site mainly comes from six subsystems:

- (1) The A2 system is a data management application system covering the collection, processing, application display, and hierarchical review and reporting of oil, gas, and water well production data. It mainly includes basic data such as individual wells, stratigraphic positions, and gas reservoirs.
- (2) The A4 system is a fundamental public geographic information service system that provides standardized, unified, and secure geographic spatial information services. Its main business includes the management of declassified oil and gas spatial data and the provision of public map application tools. The system contains spatial geographic data such as pipelines and stations.
- (3) The A5 system is a system for oil production and ground engineering production operation management and data decision support. It includes subsystems for gas production, crude oil gathering and treatment, gas gathering and treatment, and water treatment.
- (4) The Equipment Comprehensive Management System 2.0 covers 22 equipment management business scenarios, including equipment basic management, pre-management, post-management, technical evaluation, safety and environmental protection, energy saving, and comprehensive management data at various stages of equipment management.
- (5) The Pipeline Management Platform system comprehensively controls assets and matters within its jurisdiction through a "one-map" + "business management" approach. The one-map

is based on the A4 system to collect and organize data on pipelines, stations, valve rooms, key project information, high-consequence areas, third-party construction, etc., within the company's jurisdiction. It displays and analyzes data in chart form, optimizes the pipeline business management process, and manages business data throughout its life cycle in the form of "basic attributes + progress supervision."

(6) The Operation Area Digital Management Platform includes modules for standardized job management, task scheduling and organization management, on-site operation process management, integrated application management of gas production operation areas, and QHSE thematic application management. It contains four major categories of data: basic system data, entity data, management data, and business data.

4. Main Measures of Data Governance

4.1. Establishing a Global Management Framework for Data Governance

Data management activities should be combined with "top-level design + localized implementation," jointly carried out by relevant parties to achieve practical results and promote the company's data governance system and enhance data management capabilities. Data governance activities should be implemented in first-line information systems and data warehouses, embedded in the entire system life cycle, and jointly address prominent data issues to improve data resource quality from the source. This enables all parties to share high-quality data governance outcomes and creates a "co-construction, co-governance, and sharing" data governance pattern for the company.

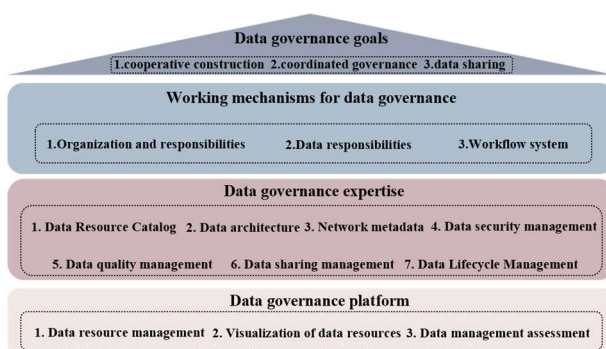


Figure 3. Global management architecture for data governance

As shown in Figure 3, the core of the global management framework for data governance in gas fields is the professional capability of data governance. The connotation of data resource catalog management includes establishing a data resource catalog for gas fields, categorizing it according to business modules such as reservoir, drilling, production, and transportation, and forming a unified data asset list for easy data retrieval, access, and management. The connotation of data architecture management includes constructing a unified data architecture, including data models, data standards, and data interface specifications, to ensure the rationality and compatibility of data structure, and to achieve cross-system data integration and associated analysis. Network soft environment data management refers to the construction of a data network security system to ensure that data is protected from external attacks or internal leaks during transmission, storage, and use. Measures such as data encryption, access control, and log monitoring are adopted to ensure data security.

Data disaster recovery mechanisms include establishing data backup and recovery strategies, using distributed storage, regular backups, and multi-site disaster recovery methods to ensure

that key business data can be quickly recovered in the event of an emergency, ensuring the stable operation of gas field production.

Data quality management involves establishing data quality standards, including indicators such as completeness, accuracy, consistency, and timeliness, and combining manual review with intelligent verification technology to improve data quality and reduce the impact of data errors on production decisions.

Data sharing management refers to the rational allocation of data access permissions based on data classification and permission control mechanisms, enabling data sharing between internal departments and partner units while ensuring the security of sensitive data and preventing unauthorized data leakage.

Data lifecycle management includes formulating data storage, archiving, cleaning, and destruction strategies for different types of data to ensure that data is stored and utilized at the appropriate time, avoiding the accumulation of redundant data, and improving storage resource utilization.

4.2. Standardization of Data Services based on Data Lake

This paper proposes the standardization of data services based on the data lake to ensure the high reliability and security of real-time data, while optimizing the data storage, computing, and service management system to improve overall data processing efficiency and consistency. The Data Information Management (DIM) module plays a key role in this process, mainly used for storing and managing dimensional data, including aggregation rule tables, dimensional tables, and point information registration tables. Through the DIM module, standardized data analysis dimensional tables can be constructed, effectively reducing the risks associated with inconsistent data calculation calibers and non-uniform algorithm standards. This standardization not only ensures data consistency but also significantly improves the implementation efficiency of downstream complex data analysis, making data utilization more efficient and precise.

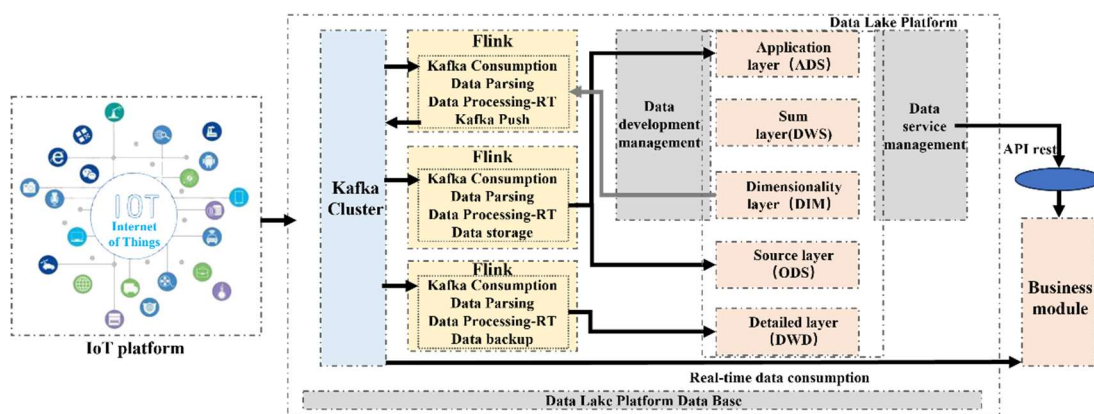


Figure 4. Real-time data lake and service development architecture

The core functions of this module cover the standardized development of real-time data services for oil and gas fields and their entry into the data lake. The entire entry process focuses on the storage and management of standardized real-time data to ensure data integrity, traceability, and efficient access. As shown in Figure 4, the real-time data entry and service development architecture of the data lake platform demonstrates the overall design of the platform. On one hand, based on the exploration and development data lake, a highly available Kafka cluster, Flink computing engine cluster, and MPP real-time data warehouse are built to provide stable computing and storage support, making real-time data processing more efficient

and flexible. On the other hand, the data lake platform provides a comprehensive set of real-time data development, management, and service management systems, enabling developers to more conveniently develop and integrate gas field data services, simplifying data processing workflows, and enhancing system scalability.

In terms of data services, query services are developed through the data lake service management platform and provide efficient data query service interfaces to business modules in the form of API REST. Meanwhile, real-time data push services rely on the Kafka messaging queue system to achieve data push capabilities to business modules, ensuring that data can be transmitted quickly and stably to various application systems. Additionally, the data lake platform relies on high-performance databases and storage devices to achieve efficient data storage and backup management, further enhancing data security and availability.

Overall, this real-time data entry and service development architecture replaces the traditional discrete component-based approach, achieving an integrated process from data collection, storage, management to application. This architecture optimizes real-time data processing efficiency, reduces data management complexity, and lays a solid foundation for the intelligent and efficient development of oil and gas field data services.

4.3. Establishing a Data Governance System Platform

A data management platform is an essential technical support for modern enterprises to implement data governance and enhance the value of data assets. In the field of oil and gas exploration and development, the effectiveness of data governance is directly related to the precision of business decision-making and the level of intelligent production operations. To meet the data governance needs of exploration and development, Sichuan **Natural Gas Well Site** has initiated the construction of an exploration and development data lake platform. After continuous optimization, this platform has evolved into a fully autonomous and controllable big data management system, achieving significant breakthroughs in data access, management, and services.

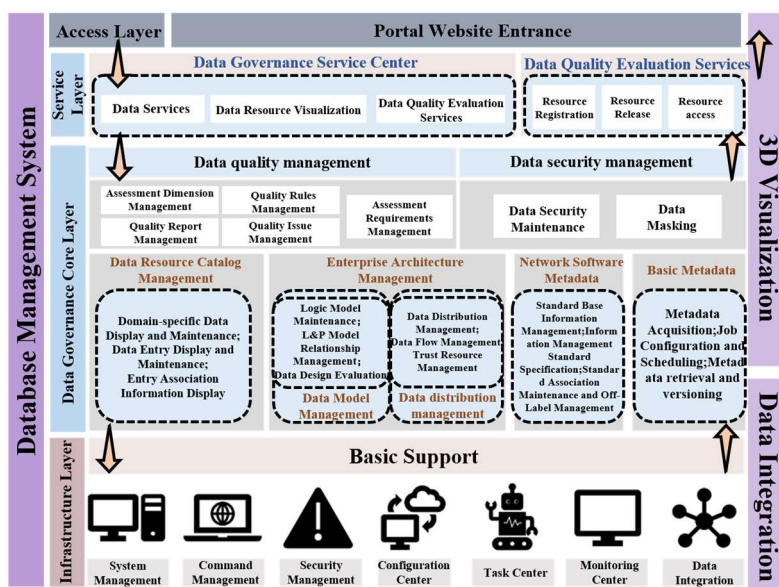


Figure 5. Data governance system platform

As shown in Figure 5, the data governance system platform is depicted. Data access is the first step in big data governance. The platform employs advanced technologies such as high-performance synchronization engines, cloud-native big data architecture, and multi-source

heterogeneous unified storage, enabling data to be accessed and aggregated within seconds. This architecture not only enhances the real-time and reliability of data collection but also accommodates various types of data sources, including complex data from exploration, drilling, geology, reservoir simulation, etc., meeting the needs for multi-source data integration.

In terms of data governance, the platform utilizes key technologies such as model consistency verification, automated secure projection of models, massive entity-relationship (ER) diagram rendering, and full-chain data lineage analysis to build a comprehensive data quality management system. This system ensures data consistency, security, and traceability. Through full-chain data lineage management, change impact analysis can be conducted to ensure smooth data transfer and security management across different business segments. Moreover, real-time linkage between various data management tools further enhances the automation and intelligence of data governance.

5. Conclusion

In the context of the deep integration of digitalization and intelligence, data governance in oil and gas exploration and development has become a key driver for the high-quality development of the industry. Through long-term exploration and practice, Sichuan **Natural Gas Well Site** has developed a digital transformation path that meets its own business needs and gradually established a data governance system that fits the characteristics of oil and gas exploration and development data. This system covers the full life - cycle management of data from collection, storage, management, analysis to application and relies on advanced technologies to build an autonomous and controllable large-scale data management platform. The construction of this platform has provided strong support for core business scenarios such as intelligent oil fields and collaborative research platforms of Sichuan **Natural Gas Well Site**, initially demonstrating the application effectiveness of data governance in improving decision-making efficiency and optimizing business processes.

With the continuous development of cutting-edge technologies such as big data, artificial intelligence, and large models, the strategic value of data governance in oil and gas exploration and development is becoming increasingly prominent. Future data governance systems need to continuously optimize in the following areas: First, strengthen data quality management by establishing strict data monitoring and evaluation mechanisms to ensure the accuracy, consistency, and integrity of data; second, deepen data mining and intelligent analysis capabilities by using machine learning and deep learning models to improve the precision of reservoir description, reserve prediction, and drilling and completion optimization; third, promote data sharing and collaborative applications to break down data silos and achieve intelligent linkage across businesses and platforms, enhancing the reuse value of data.

On this basis, oil and gas companies need to build a more agile and intelligent data governance model to empower exploration and development decisions with precise data insights, improve the efficiency of oil and gas resource discovery, reduce production costs, and ultimately maximize the value of data. Data governance is not only a support for digital transformation but also a key engine for driving the petroleum industry towards efficient, intelligent, and sustainable development.

References

- [1] Sumbal Muhammad Saleem, Tsui Eric, Irfan, et al. Value creation through big data application process management: the case of the oil and gas industry [J]. *Journal of knowledge management*, 2019, 23 (08): 1566-1585.
- [2] Choi Heaok, Park Hwanll. "Oil is the New Data": Energy Technology Innovation in Digital Oil Fields [J]. *Energies*, 2020, 13 (21): 5547.

- [3] Liu Baolei, Li Changxuan. Mining and Analysis of Production Characteristics Data of Tight Gas Reservoirs [J]. 2023, Processes, 11 (11): 3159.
- [4] Diviak Tomas. Key aspects of covert networks data collection: Problems, challenges, and opportunities [J]. Social networks, 2022, 69: 160-169.
- [5] Arena Simone, Manca Giuseppe, Murru Stefano, et al. Data Science Application for Failure Data Management and Failure Prediction in the Oil and Gas Industry: A Case Study [J]. Applied sciences-basel, 2022, 12 (20): 10617.
- [6] Xu Sen. Research and Analysis of Digital Construction in Oil and Gas Field Surface Engineering Based on Internet of Things Environment [J]. Fresenius environmental bulletin, 2020, 29 (11): 9722-9729.
- [7] Sumbal Muhammad Saleem, Tsui Eric. Interrelationship between big data and knowledge management: an exploratory study in the oil and gas sector [J]. Journal of knowledge management, 2017, 21 (01): 180-196.
- [8] Alguliyev Rasim M, Aliguliyev Ramiz M, Abdullayeva Fargana J. Multidisciplinary study of the problems of big data technologies in the oil and gas industry [J]. International journal of oil gas and coal technology, 2020, 23 (01): 92-105
- [9] Zhang Xueqiang, Cheng Qinglin, Sun Wei, et al. Research on a TOPSIS energy efficiency evaluation system for crude oil gathering and transportation systems based on a GA-BP neural network [J]. Petroleum science, 2024, 21 (01): 621-640.
- [10] Almassi Zia, Mirzaei Ramazan, NasrAbadi Mahnaz, et al. HSE (Health, Safety and Environment) culture assessment via HSEMS (Health, Safety and Environmental Management System) (Case study in Arvandan Oil and Gas Company) [J]. Life science journal-acta zhengzhou university overseas edition, 2012, 9 (04): 286-293.
- [11] Su Jian, Yao Shanglin, Liu He. Data Governance Facilitate Digital Transformation of Oil and Gas Industry [J]. Frontiers in earth science, 2022, 10: 861091.
- [12] Lei Ke, Wang Wei, Ren Shengli, et al. Application of Data Governance in Petroleum Enterprises [J]. Petrochemical Technology, 2021, 28 (06): 180-182+16.
- [13] Qin Jiongcong. Data Governance Model under the Background of Big Data [J]. Information and Computer (Theoretical Edition), 2016, (16): 155-156.
- [14] He Yan. Research on the Construction Technology of Oilfield Data Resource Catalog [D]. Northeast Petroleum University, 2021.
- [15] Panian Z. Some Practical Experiences in Data Governance [J]. World Academy of Science Engineering & Technology, 2011.
- [16] Dutta H. Graph-based Data Governance Model for Real-time Data Ingestion [J]. CSI Transactions on ICT, 2016. DOI:10.1007/s40012-016-0079-y.
- [17] Sun Shaobo. Research on Data Governance Methods and Technologies in Oil and Gas Field Exploration and Development Production [D]. Chang'an University, 2018.
- [18] Wang Quan. Research on Digital Oilfield Model and Development Strategy of Daqing Oilfield Co., Ltd. [D]. Tianjin University Thesis, 2003.
- [19] Liu Shunchun. Research on Data Governance System in "Shared CNPC" [J]. Journal of China National Petroleum Management Cadre College, 2019, 26 (6): 21-29.
- [20] Zhang Shiming, Peng Xuefeng, Huang Hexiao. Research on Open University Data Governance Framework [J]. China Electric Education, 2018 (8): 116-126.
- [21] Sun Min. Data Governance Engineering and Its Application in Intelligent Oilfields [J]. China Management Informationization, 2018, 21 (06): 49-50.
- [22] Liu Wenkui, Wang Qian, Yang Yunchun, et al. Application of Data Governance in Petroleum Exploration and Development Collaborative Research [J]. Petrochemical Automation, 2024, 60 (04): 36-39.