Research on Logistics Data Governance System

Shenrui Li, Deqian Fu[(✉)], and Jianlong Qiu[(✉)]

Linyi University, Linyi, China
{fudeqian,qiujianlong}@lyu.edu.cn

Abstract. With the rapid development of logistics, this paper expounds the data governance, introduces the basic status of logistics data, combs the problems faced by the logistics industry data, and proposes to establish a logistics data governance system to provide standardized theoretical support for China's logistics data governance.

Keywords: data governance · logistics data · Standardized theoretical support

1 Introduction

With the continuous development of information, digitalization and intelligence, the important role of data has become more and more widely recognized by the whole society [1]. With the network becoming a new infrastructure, China's data began to grow rapidly. At the first plenary meeting of the IT Governance and IT Service Management Technical Committee held in Australia in 2014, the Chinese representative first proposed the concept of data governance [2]. In 2020, the central document "Opinions on Building a More Perfect System and Mechanism for Market-Oriented Allocation of elements" clearly proposed that we should accelerate the cultivation of the data element market, promote the open sharing of government data, enhance the value of social data resources, and establish a unified and standardized data management system, which fully demonstrates the importance of the Party Central Committee and the State Council on data governance [3]. In January 2022, the "Fourteenth Five-Year Plan" for the Construction of Modern Circulation System was released. The plan points out that the modernization of China's circulation system is still not high, and the free flow of goods and resource elements still faces hidden barriers. As an important part of the circulation field, the logistics industry will further improve the rules for the confirmation, transaction and application of data resources, and promote the compliance transaction, safe and efficient use of logistics resources. In April 2022, the Opinions of the Central Committee of the Communist Party of China and the State Council on Accelerating the Construction of a National Unified Market was issued, which further clarified the need to build a modern circulation network, speed up the digital construction, promote the innovation of the technology and business model of the third-party logistics industry, and establish and improve the basic systems and standards for data security, rights protection, cross-border transmission management, transaction circulation, open sharing, security certification,

etc. The establishment of a unified national market will further promote the investigation of data resources in various industries and strengthen the development and utilization of data resources such as logistics data [2, 4]. At present, many government departments, enterprises and institutions have paid full attention to their own data assets and carried out relevant work on data governance through various advanced technical means [5–8].

With the transformation of the logistics industry from the traditional industry to the digital and intelligent newmodel, the logistics industry is also actively carrying out the digital transformation in order to maintain the leading position in the highly competitive market environment. In this process, the dependence on data will become higher and higher. These logistics data are multi-source, multimodal, heterogeneous and distributed in different logistics systems, and have great potential value. At present, China's logistics data generally has problems such as low quality, lack of uniform standards, data privacy security, and so on. It is urgent to carry out research work on data governance.

2 Problems

Logistics business personnel still have some pain points in the daily use of data, mainly including:

First, the construction of data standardization is insufficient. Data standards will run through the whole process of data management. Although we have developed a series of normative documents, system documents, process documents, etc., the existence of standards does not mean that data standardization has been implemented. For example, the standardization of business data and master data needs further improvement.

Second, data quality. Data quality is the lifeline of data. Poor data quality seriously affects the conclusions of data analysis. Some may mislead decisions, such as dirty data, missing or changing dimension data, and a series of problems need to be addressed. For example, the lack of scanning information leads to inaccurate waybill routing; The data dimension value changes, and the business volume of a certain channel increases or decreases sharply. Therefore, a correct evaluation of data quality is required.

Third, data security. Compared with other industries, logistics data has its own characteristics. Business subjects have high requirements for data security and privacy protection, so it is necessary to hierarchical control sensitive and private logistics data to ensure data reliability and security to the greatest extent.

3 Logistics Data Governance System

Although the construction of the logistics system is improving, there is insufficient data sharing between the systems, and the degree of integration is relatively poor. Railway enterprises need to further do a good job in data maintenance to facilitate integrated management. At present, the intensity of data management is backward, lack of sufficient standardization, data inconsistency, data inaccuracy and other problems, data quality needs to be continuously improved in the later operation, and data needs to be controllable. Based on the analysis of the above status quo, this study will propose a logistics data governance system, refer to Fig. 1 for details.
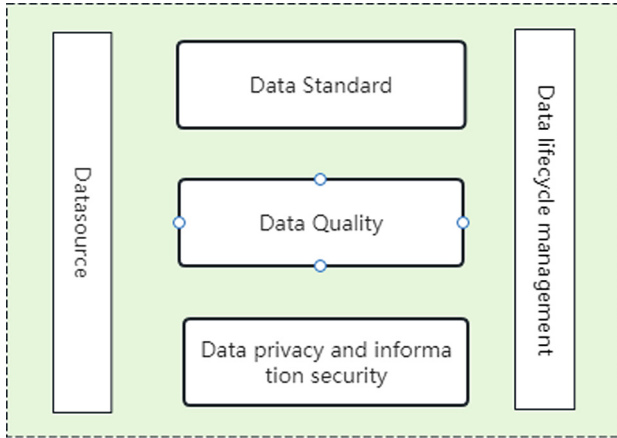
Fig. 1. Logistics data governance system

3.1 Establish Data Specifications

The purpose of data governance is to manage the massive information distributed in various systems. However, due to the differences among various information systems, such as data code specifications, formats and identifiers, these differences may affect the accuracy of data. Therefore, it is necessary to establish a unified standard system to ensure the reliability of data analysis results, and in line with industry norms. For the attribute value of the database, we can use the unity principle, persistence principle and null value principle to test and constrain. The unity principle refers to adding unique constraint to the primary key or any attribute, so that the value of each attribute is different from that of another attribute. The persistence principle refers to that there is no missing value between the maximum value and the minimum value of an attribute, and each value is unique, which is usually used to check the integrity of the data. The null value principle refers to the use of some special symbols to replace the null value, in order to verify the integrity and accuracy of data. Therefore, industrial or national data standards should be followed when formulating data specifications to ensure the reliability and validity of information. Aiming at the problem of data standards, this paper combines the national standards to design and standardize the logistics data. As shown in the chart Table 1.

3.2 Data Quality Grading Evaluation

First of all, the logistics data quality evaluation index is formulated to evaluate the data quality of the logistics business data to be processed. At the same time, data quality uses ETL tools to clean business data and convert it into correct data in a unified format to ensure the integrity and accuracy of data after comparison between data sources and data dictionaries, improve data quality, and meet the needs of subsequent data applications.

Table 1. Data specifications.

Number	Attribute name	Explain
1	English name	English description of data.
2	Chinese name	Chinese description of data.
3	Constraints/Occurrences	Number of occurrences in data: a) 0.. 1 -data can be filled in and appears once; b) 0.. N -- data can be filled in and can appear multiple times; c) 1. 1 -data is required and appears once; d) 1.. N -data is required and can appear multiple times.

4	data format	The meaning of the characters used in the data format is as follows:
		a) A = alphabetic character;
		b) N = numeric character;
		c) An = alphanumeric Chinese character;
		d).. = separator from the minimum length to the maximum length of character data;
		e), = separator to distinguish the number of numeric characters from the number of decimal places after the decimal point, that is, the number of numeric characters before "," and the number of decimal places after ",";
		f) YYYYMMDDhhmmss = "YYYY" represent year, "MM" represent month, "DD" represent date, "hh" represent hour, "mm" represent minute, "ss" represent seconds. It can be combined according to the actual situation.

5	explain	Brief explanation and application description of data.

3.3 Data Hierarchical Control

Formulate corresponding data sensitivity rules for logistics data according to the actual logistics business scenario. The rule table is as follows Table 2, Table 3, Table 4:

Table 2. Logistics data sensitivity level rule.

Data sensitivity level	Data level identification	Sensitive data type (example)
Grade M1	Sensitive	Customer identity-related data, customer service content data, internal core management data
Grade M2	More sensitive	Enterprise operation management data
Grade M3	Low sensitivity	Open and share data

Table 3. Data classification rule.

Business data 1	Business data 2
Order business data	User data

Table 4. Data grading rule.

M1 sensitive	M2 more sensitive	M3 low sensitive
Customer name, mobile phone number and address		Customer's net name and avatar
Order freight, insured rate, total number of packages	Order number and outlet credit value	Outlet name

3.4 Data Lifecycle Management

To realize data lifecycle management, a data lifecycle management system must be established. Logistics data should be divided into cold data, hot data and hot data according to the generation time and use frequency of data. In addition, it is also necessary to develop a unified data lifecycle management for the whole road. Establish integrated management standards, carry out more detailed management of current data, clarify management standards, management schemes and management systems, and maintain scientific, systematic and unified data management. Life cycle management also needs to specify the rules of data cleaning, the time of data cleaning and the specific process of standard implementation to ensure the correct operation of data life cycle management.

4   Summary

This paper establishes a logistics data governance system based on the actual logistics business scenario, and solves the problems in the actual scenario. Thus, it can greatly improve the standardization and information management level of logistics data governance, protect the privacy and security of the data to a greater extent, provide a certain reference for the data governance of the logistics industry, and maximize the value of logistics data.

References

1. Sun Yue. The logic and value of data rule of law [D]. Jinan: Shandong University School of Law, 2020:12-13
2. Wu Tao. Research on "precision" decision-making of smart government from the perspective of big data governance [J]. Journal of Yunnan University of Administration, 2017, 19 (6): 110-115
3. Zhang Huiping. Building a more open urban data governance system [J]. National Governance, 2021 (Supplement 1): 52–55
4. Zhang Chi. Research on the core value of data center in responding to public emergencies [J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2021, 35 (1): 77-84
5. Yang Xinzhu. Research on key technology and application of provincial government big data center construction [D]. Lanzhou: School of Transportation, Lanzhou Jiaotong University, 2019:1–2
6. Huang Ruhua. Construction of China's government data open sharing standard system [J]. Books and Information, 2020 (3): 17-19
7. Cao Huimin, Deng Tingting. Research on government data governance risk and its resolution mechanism [J]. E-government, 2021 (1): 81-91
8. Wang Yi, Wang Zhiwei, He Xin. Construction and application of smart power station data center [J]. China Mobile, 2021, 54 (3): 61-67176