Research on Railway metadata management information system based on CWM

Chenyu Bi[1, a], Lei Huang[2, b]

[1] School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

[2] School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

[a]bcy1108@163.com, [b]13241157@bjtu.edu.cn

Keywords: metadata, railway, CWM.

Abstract: China's railway system is very complicated. The data in these systems can be quite useful to promote railway's development. But the data resources are scattered, heterogeneous due to the different system architectures. And data integration between systems is insufficient. Besides, the level of data standardization and quality is poor which make the value of the railway bigdata is hard to play effectively. The paper will start from this situation, try to make better use of railway bigdata through building a metadata management system. Promoting the standardization of data, making the data play its due value.

1. Introduction

At present, the research about big data is being launched around the world. Both academics and industry have widely recognized that the effective use of the value contained in big data will have a huge impetus for industrial upgrading and social-economic development.

With the development of China, railway is shouldering an important historical mission in the process of building a well-off society. Railway informatization plays an important role in supporting and ensuring the reform and development of China Railway Corporation.

Liu Yifei thought, due to the weak concept of master data management in the initial stage of railway information system construction, the railway data is not standard and hard to analyze. And the

Bigdata era has high requirements for data quality. She proposed to build railway master data platform to try to solve the problems [1]. Zheng Qiaoran has studied how to integrate railway freight data and how to make use of the data [2]. Liu Yanjun et al has summarized the status and needs, and proposed the application architecture and technical architecture [3]. There is still a lot of study on railway big data. However, research on railway metadata is relatively little. Ran Youhua discussed applicability of Metadata crosswalks in the geographic information metadata database system of Qinghai-Tibet Railway and carried out its preliminary design [4]. Li Yang et al proposed the main functions of the railway geographic information system (RGIS) metadata in RGIS, provides an idea for the construction of metadata in RGIS [5]. Li Hongyue's paper aimed at the lack of unified classification and coding schemes and metadata standards for RGIS, started discussion from the spatial information of railway engineering [6].

In summary, though China Railway Corporation has made some major decisions and strategic plans for the reform of organization and so on. But objectively, there is still a clear gap between informatization construction and the reform requirements. There are problems such as information island and dispersed resources. The data quality is poor. And the current research which about railway data focuses on master data and bigdata application, pays little attention to manage metadata. Metadata can facilitate the efficient use of data sets and help data production units maintain and manage data effectively. This paper will try to take this as a start point, and try to build a railway metadata management system, clarify relationships between systems, and manage data effectively.

  DOI: 10.25236/msied.2018.006

2. Theoretical basis

2.1 Metadata

The definition of metadata has many expressions. Metadata is not a new concept, but there are innovations in the different environment.

Metadata is defined as data describing data, descriptive information about data and information resources. Metadata is data about other data, or structured data for providing information about a certain resource. Its purpose is to: identify resources; evaluate resources; track changes in resources during use, and achieve simple and efficient management of large amounts of networked data [7].

2.2 CWM

CWM (Common Warehouse Metamodel) is a complete metamodel architecture defined by the OMG organization in the data warehouse system for metadata modeling of data warehouse construction and applications [8].

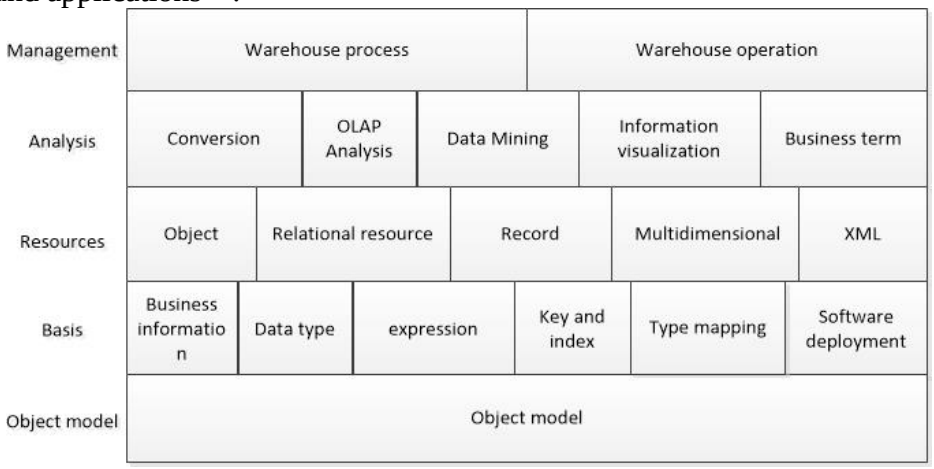

Fig.1. CWM architecture

3. Railway metadata management system

3.1 Metadata standard

From the above, we can know that one important reason why the railway data assets are not be exploited well is that different systems may be heterogeneous, the data integration between systems is insufficient, and the data itself is not clear. So, we import the metadata to solve the problem. First, let's define some concepts. For example, the account information comes from accounting processing subsystem, and is used in fixed asset accounting subsystem, then we call account information as a data entity. And what account information contains, such as receivable balance, and advance balance. We call them data item. The data entity refers to the various information contained in the system and is main part of our research. The railway metadata must contain the information includes where the data entity come from, where it will go and other information. See the table below for details.

Table 1 Metadata information

Classification	Description
Basic information	Data entity unique flag/ source/direction/etc.
Maintenance information	Maintenance department/people in charge etc.
Provider information	Provider department/contact way etc.
Classification	Belong to which industry, which system etc.
Open condition	Unconditional/In the same system/etc.
Sharing condition	Unconditional/ No sharing/Sharing in the same system/ etc.

3.2 System architecture

Based on CWM model and actual requirements, we divide the system into the following sections. First is the data asset catalog, according to the above, we divide this section into four parts, they are system and subsystem, data entity, data item, data interaction. Second is the data collection, there are three data sources, physical database, IT migration process and ETL process. And the third is metadata management, user can publish new data entity, and enter the information in the above table. User can also publish data item, and enter similar related information. Besides, user can modify, delete the metadata. The fourth part is about metadata standardization. There are two way to ensure the standard, one is by CWM and the other relies on human. The final one is authority management, this part is to make sure the data is secure. We can receive information through different sources, perform verification processing automatically or manually, sort according to certain standards, obtain detailed information of each data entity, and understand the interaction between different systems, thus prepare to carry out further data value mining.

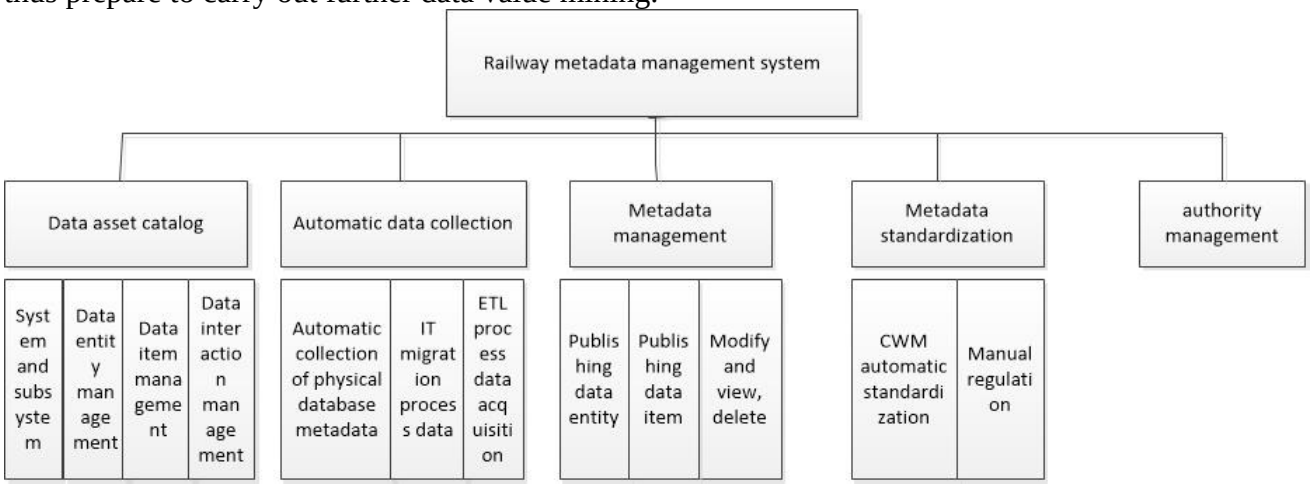

Fig. 2. Railway metadata management system functional architecture

3.3 Some implementation examples

Through the above analysis ang design, we can initially realize a prototype system for railway metadata management. The following is a concrete implementation.

1) Data entity flow among systems

The picture displays the data entity interaction. Which system does the data entity belong to and which system it will interact with. Through this picture, we can know the relationship among different systems clearly, and it's also convenient for the managers to perform data review and identify the problem.

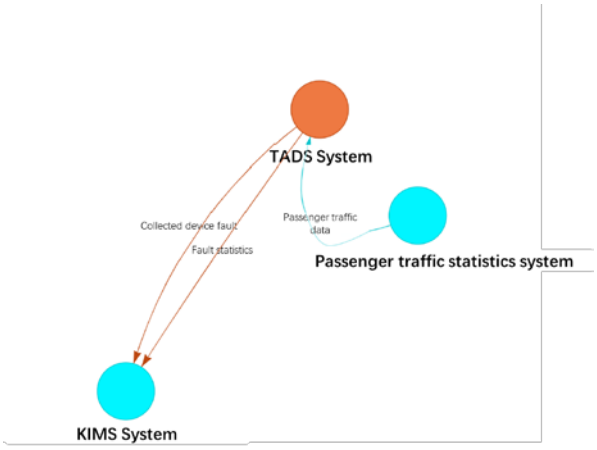

Fig. 3. Data interaction among systems

2) Data entity publish

The picture displays the data entity publish. The user needs to enter relatively comprehensive information about the entity before publishing it.

This page guarantees that every published data entity has clear and standardized metadata for future unified management and data mining.

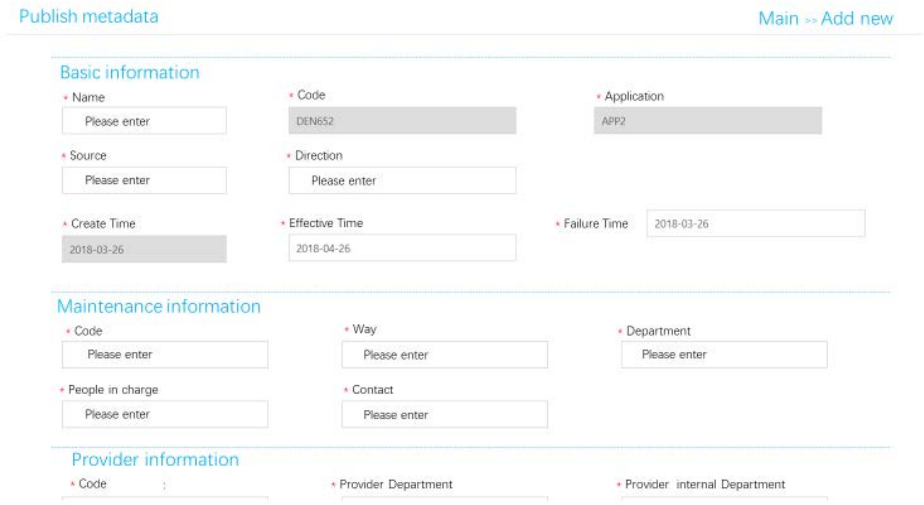

Fig. 4. Data entity publish

4. Conclusions

This paper starts with the current status of data management in railway system, sums up the difficulties faced by railway data governance, and clarifies the importance of metadata management. And based on the CWM specification, the railway metadata management system was designed and implemented. This system can provide ideas for the data governance of railway. In order for the railway to make full use of data assets, it is necessary to improve the quality of the data. Through managing the metadata, can effectively grasp the information of each data entity and gradually standardize each data. At the same time, combined with the management of railway master data, it can better support the top-level decision-making and future development of the railway.

References

[1] Liu Yifei. The Key Technologies research of Railway Master Data Management for Big Data Applications [D]. China Academy of Railway Sciences, 2018.

[2] Zheng Qiaoran. Research on data integration technology solution and application cases of railway 95306 website [D]. Southwest Jiaotong University, 2018.

[3] Liu Yanjun. Research and Design of Railway Big Data Asset Management Platform [A]. China Intelligent Transportation Association. Proceedings of the 12th China Intelligent Transportation Annual Conference[C]. China Intelligent Transportation Association: China Intelligent Transportation Association, 2017:6.

[4] Ran Youhua. Designing the Metadata System of Qinghai-Tibet Railway Geographical Database Based on Metadata Crosswalks [J]. Remote Sensing Technology and Application, 2004(05):379-385.

[5] Li Yang, Liu Ying, Liu Rengkui. Study on Metadata Problem for Railway Geographic Information System [J]. China Railway Science, 2003(05):76-80.

[6] Li Hongyue. Study on the metadata standard for spatial information of railway engineering affairs [D]. Southwest Jiaotong University, 2003.

[7] Baidu.Metadata [DB/OL].

https://baike.baidu.com/item/%E5%85%83%E6%95%B0%E6%8D%AE/1946090?fr=aladdin

[8] Zhang Mingzhi. Metadata Management System Based on CWM Specification Design [J]. Computer Knowledge and Technology, 2014, 10 (02): 254-258.