Hindawi
Security and Communication Networks
Volume 2023, Article ID 1414530, 13 pages
https://doi.org/10.1155/2023/1414530

Research Article

A Management Specification for Data Sharing Security in the System Construction of Smart Mine

Haitao Wang,[1] Lina Tan ,[1] Yang Zhang,[1] Qiwen Gong ,[2] Shan Zhang,[2] Yanan Ren,[2] and Tao He[3]

[1] *China Coal Energy Research Institute Co. Ltd, No. 66 North Yanta Road, Xi'an, China*
[2] *The School of Computer Science and Technology, Xidian University, Xi'an, China*
[3] *Institute of Intelligent Manufacturing, Wenzhou Polytechnic, WenZhou, China*

Correspondence should be addressed to Lina Tan; sabrina_tan2017@163.com

Received 27 July 2022; Revised 15 January 2023; Accepted 11 April 2023; Published 29 June 2023

Academic Editor: David Megias

With the development of Internet of Things technology and the informatization of the coal industry, various intelligent applications have emerged in the process of the system construction of the smart mine. During this process, data sharing is essential to the effective use of data resources in the smart mine. In order to improve the protection of coal mine data, this study proposes a set of management specifications for data sharing applied to the system construction of a smart mine to unify security standards in data storage and sharing. It standardizes the processes of data collection, transmission, and storage. We design three sub-specifications for these processes, namely, data source specification, data quality specification, and data storage specification. The data source specification specifies the data collection and transmission standards to improve the security and timeliness of data sharing. The data quality specification sets three evaluation criteria of integrity, accuracy, and timeliness according to the characteristics of each business system and data. The system ensures data quality during data sharing by governing and recording the data failing to meet the criteria. The data storage specification specifies the data storage protocol, data label, and data set restrictions. Only authorized platforms and users can share data and make use of data labels to search data efficiently. Finally, we constructed a coal mine data collection and analysis system. It can collect, manage, store, and safely share the real measured data from a certain colliery according to the specifications.

1. Introduction

With the rapid development of intelligence in the coal industry, the Internet of Things (IoT) has become the key technical support for the construction of intelligent mines. The innovative development of coal resources relies on big data technology. Coal mine data reflect the overall production process, production indicators, safety status, and other production information of coal mine. With the application of big data technology in smart mine, data become one of the most important resources. Colliery data have the following characteristics: first, the scale of the data is large; second, the data collection speed is fast; third, the value density of the data is low; and fourth, the data need high accuracy and strong timeliness. Data sharing is essential to the efficient use of data resources in smart mines. One of the biggest challenges of data sharing is to safely transmit the increasing amount of data. Data sharing is often accompanied by extraction, transformation, and loading processes. This means that data quality, data governance, and data security are particularly important. In order to realize the further progress of intelligent coal mine construction, data standardization has become a challenge.

In the operation of a smart mine, there are three different types of data sharing processes in the coal mine, as shown in Figure 1. First, all kinds of automation systems collect data and store it in the data storage platform. Second, the intelligent coal mine application extracts data from the storage platform and analyzes it. Finally, each coal mine aggregates the data collected by the automation system or from the data storage platform and submits it to higher management.

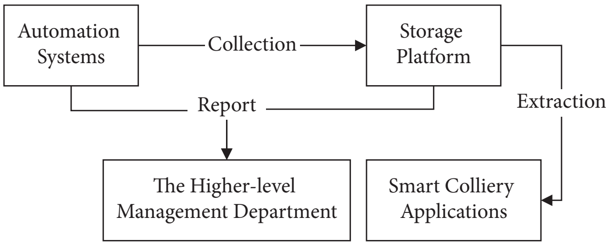

Automation Systems
Collection
Storage Platform
Report
Extraction
The Higher-level Management Department
Smart Colliery Applications

Figure 1: The data transfer processes in smart mine.

Since various automation systems belong to different businesses and have different functions, the type of data it manages and its granularity varies. Therefore, the data formats transmitted to the storage platform by each automation system are different. At the same time, there is a lack of uniform transmission protocols and standards for data sharing at all levels. As a result, storing and using coal mine data can often encounter problems such as low data accuracy, unguaranteed timeliness, data leakage, and difficulties in accountability. Specifically, there are four common problems, shown as follows:

(1) The format of the dataset is piecemeal, reducing data sharing security and timeliness, and the standardization workload is heavy. The data accessed by the intelligent coal mine big data platform is scattered on multiple devices. The file format and data format description of the sent data set are confusing and nonstandard when automation systems transfer data to the storage platform and the upper-level management department. The storage platform and upper management are unable to identify the integrity of the data, which is prone to data inconsistency caused by secondary delivery. Before implementing data from disparate business systems, the receiving unit must restandardize the data according to business demands. This restandardized process has resulted in an enormous workload for those responsible for data collection. It also led to delays in data collection and eventually no one wanted to take on the job.

(2) The lack of clear data quality descriptions makes the storage platform unable to guarantee the correctness of shared data. The coal mine automation system cannot describe data quality very clearly when uploading data. It makes the storage platform unclear about the quality of the data, making it difficult to ensure its correctness and unable to guarantee data standardization. This poses a potential hazard to future data applications. Problems with data quality can lead to duplication of data collection by the system, resulting in the reduction of data sharing, and it is difficult to form a virtuous circle.

(3) Coal mines have diverse production environments and equipment, and there is no uniform reference specification for data governance, especially for specific types of data. Coal mines have different business systems, and each business system corresponds to a variety of equipment. It is common for automated systems to generate abnormal data. The storage platform and data collection unit need to manage the error data submitted by each automatic system and perform different correction and annotation operations according to the type of error data. There is no dedicated governance for characteristic type data and sensitive data, which can easily lead to sensitive information leakage. Desensitized data are difficult to maintain data consistency and business relevance. Data governance is primarily done by people who do not have expertise related to mining. Data governance cannot achieve the expected results due to the lack of reference specifications for specialized governance methods for abnormal or irregular data or sensitive data.

(4) The platform does not have clear storage specifications, erroneous data are difficult to trace, and there are data sharing security issues. From production to the presentation of the final results of intelligent coal mine applications, coal data often go through multiple processes such as extraction, conversion, mapping, and reorganization. The platform does not select the appropriate data storage according to the business characteristics, and the systems easily access different levels of data. In these processes, security risks and errors such as data leakage, data tampering, data loss, data inaccuracy, data redundancy, and data expiration often occur. The lack of storage specification requirements for data storage retention times and record cyclic relationships makes it difficult to investigate errors and improve processes when these problems occur.

Because of the abovementioend problems, this study designs a specification for the data processing process in coal mines to unify security standards in data storage as well as sharing to improve the protection of coal mine data. This specification covers data collection, transmission, and storage in the process of unified data management in intelligent mines. It specifically includes data source specification, data quality specification, and data storage specification. It has the following attributes and functions:

(1) A data source specification: It describes the format of data transmission and file storage in the data collection process. It reduces the workload of coal mine workers, improves efficiency while reducing personnel, and reduces pretreatment work in the subsequent stage. The intention is to improve the security and timeliness of data sharing.

(2) A data governance specification: It helps software developers realize data governance without professional knowledge of coal mining and data mining. Design desensitization rules according to the data needs of different business units to improve the security of sharing special data.

(3) A data storage specification: It defines the data retention period of data storage and the mapping

format between recorded data. This specification makes it easier to track problems and improve the system when errors occur in production. Systems share data securely based on access rights for data.

This paper is arranged as follows. In Section 2, we review some relevant work. In Section 3, we design the top-level structure of the data management specification model. Then, three submodels are proposed, respectively, in Sections 4, 5, and 6. They are data source specification model, data quality specification model, and data storage specification model. In Section 7, we create a coal mine big data system to validate the utility of the specification and demonstrate the implementation of this data management specification. The study is concluded in Section 8.

2. Related Work

The authors in [1] provided a digital construction plan for coal mine big data based on life cycle management, which included technological approaches such as digital data collection, processing, and storage. It can also be used in other industries. The authors in [2, 3] used Internet of Things (IoT) technology to create a smart mining architecture. Their architecture includes data collection, data transfer, data storage, and intelligent applications. The authors in [4] presented a data platform system that combines digital technology, big data, and artificial intelligence. This data platform system can collect, transmit, store, and process smart mining data over the network. However, the main issues encountered in the development of smart mines, such as data transmission and storage efficiency, data quality, and data traceability, cannot be fully addressed in a single system.

Coal businesses employ IoT technology to construct smart mines in order to boost mine production and better manage coal mine big data. However, the problem of transferring huge amounts of data created by end devices has become an important issue that must be addressed. Edge computing is currently a very representative solution for reducing the Internet of Things data transmission delay [5–8]. In the studies of [9, 10], techniques for work assignment in edge computing systems are proposed. They carefully examined the trade-off between data transmission and computer resource allocation. Based on multihop vehicle computation resources, the authors in [11] suggested an adaptive algorithm offloading technique. The aim of these algorithms is to reduce task delays. Another solution to the problem of low-quality intelligent analysis findings produced by data noise in large data sets is to effectively minimize the data set size [12, 13]. The authors in [14] created an edge computing based system to handle data anomaly detection and analysis in underground mining. The edge devices were employed to do anomaly detection jobs, which increased efficiency. The authors in [15] present a study based on edge computing technologies that offered intelligent video surveillance for coal mines. FL-YOLO, a depthwise separable convolution and downsampling inverted residual block algorithm, was used in edge devices to identify security incidents. The authors in [16] developed an unloading task method that took into account network latency, wireless communication air rate, and computer resource consumption. To find the best option, they employed particle swarm optimization. The authors in [17] use federated learning in wireless edge networks to safeguard the privacy of user data, improving the performance of federated learning by jointly optimizing local accuracy and various resource allocation strategies. The authors in [18, 19] provide algorithms for the Internet of Things system's nodes. They evaluate the social relationships between nodes and partition the nodes to increase the Internet of Things' information transmission efficiency and network performance. Our specification is based on IoT devices and edge computing, standardizing data processing, designing anomalous data detection, and governance standards to improve data transmission speed and quality.

Various data standards are used in different fields to describe and manage data storage and transmission. For instance, in the field of geographical information, the International Standards Organization [20] provides a structure that describes the various steps involved in the data description, management, transmission, and sharing. In order to identify the types of errors in the metadata elements, the authors in [21] presented a method that can be used to improve the quality of data based on ISO 19157:2013. In the field of biology, the authors in [22] proposed data specification known as BIND was presented to describe and store the biomolecular information. In the field of medicine, various medical decision-making systems are based on the data collected and stored by multiple sources [23]. To improve the efficiency of telemedicine services, the authors in [24] developed a framework that standardizes the four processes involved in the collection, analysis, transmission, and decision-making of data. Due to the inconsistent nature of the data specifications in materials science, it is difficult to use them in deep learning. For addressing this issue, the authors in [25] created a data specification that is flexible, searchable, and formal. In the smart city, the data collected by the sensors will need to be stored and analyzed to improve the efficiency of the operations. The authors in [26] proposed an attributed-based specification that can be used to find and analyze the data. We proposed a data specification applied to the coal mining industry in the study of [27], but it is still not comprehensive, and this study makes further research on the basis of the study of [27].

3. Data Management Specification Model

This paper mainly discusses the specification of the data processing stage in coal mine, and the general framework is shown in Figure 2. The figure describes the direction of data flow. Data source access is to standardize the data collection behavior of each automation system at the source end, including data source, data format, and equipment information. Data transmission is a standard constraint on the transmission stage between different levels, mainly the specification of data transmission mode, transmission protocol, and data governance. Data storage standardizes data storage locations and storage media.

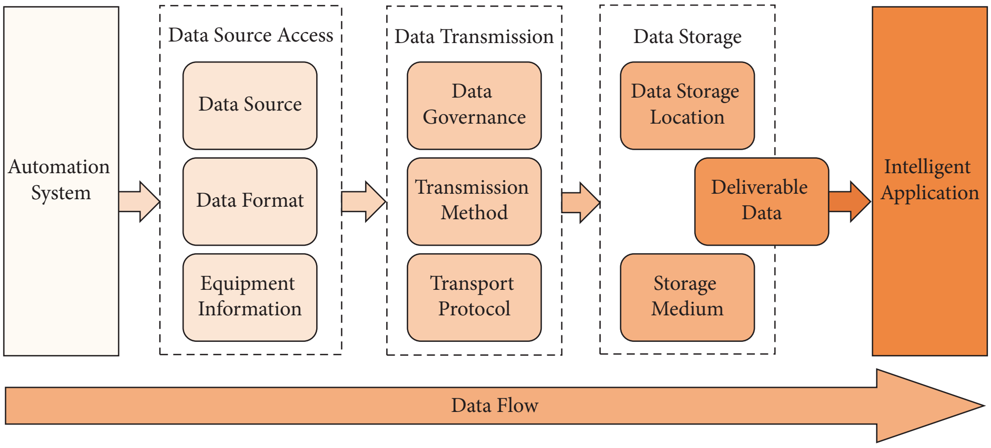


Figure 2: Overall framework of the data specification.

This section defines a data management specification model using unified modeling language (UML) based on the data source access, data transmission, and data storage parts of the abovementioned framework, as shown in Figures 3–6. It covers all data processing stages of smart coal mining, including data characteristics, data transmission, data quality, and data storage. The top structure of the data management specification model is depicted in Figure 3. It consists of three models: data source specification, data quality definition, and data storage specification.

The three models have the following connection. At first, the data source specification governs the data collecting and transmission method. This corresponds to the data source access, transmission method, and transport protocol. Second, the data quality specification outlines the data inspection and data governance procedures to be followed during the transmission process. This corresponds to the data governance module of the framework for data transfer. Third, the processed data is saved following the data storage specification. This requirement corresponds to the data storage module in the framework. Finally, intelligent coal mine applications retrieve and exploit the data.

(1) *Data Source Specification Model.* A full description of the data source is provided by the data source specification model. It specifies a hierarchical classification of the data. Some data must be recorded during the data collection and transmission procedure. The standard mandates documentation of the data source system and associated sources, as well as other pertinent and essential information, to permit traceability of issue data and accountability.

(2) *Data Quality Specification Model.* The data quality specification model is used to establish the data quality standards and assessment criteria for more advanced intelligent applications. Utilizing pertinent details like the data source system and source description, one may examine the data integrity, correctness, and timeliness efficiently. Furthermore, problematic or nonstandard data might be recognized, repaired, and handled by professional experts to raise the level of data quality.

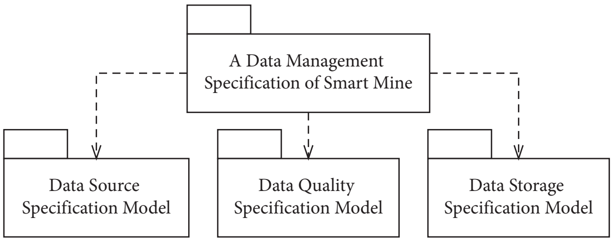

A Data Management Specification of Smart Mine
Data Source Specification Model
Data Quality Specification Model
Data Storage Specification Model

Figure 3: Data management specification model.

(3) *Data Storage Specification Model.* A detailed definition of the record information required to transmit data to the storage platform is provided by the data storage specification model. Data storage location, medium, and life cycle are all specified by the data storage specification model. The placement of the storage aids in making the data flow clear. Applications for coal mine intelligence can discover the information's source. The practitioner can more easily analyze the data lineage with its assistance. The life cycle assists in avoiding data duplication, enhancing the spatial exploitation of data storage, and providing greater support for intelligent applications.

4. Data Source Specification Model

The data source specification model is to standardize the process of data source acquisition and transmission. It unifies the data access process, classifies data hierarchically, and improves data sharing security. Figure 4 shows the data source specification model. During data collecting, the following details must be set at the same time: data source system, data source description, data transfer, identification, contact information, and references.

(1) *Data source system* gives specific information about the data source system, which is used to clarify the scope of business scenario requirements for data sharing and ensure that data usage is not beyond the authorized scope. For systems containing

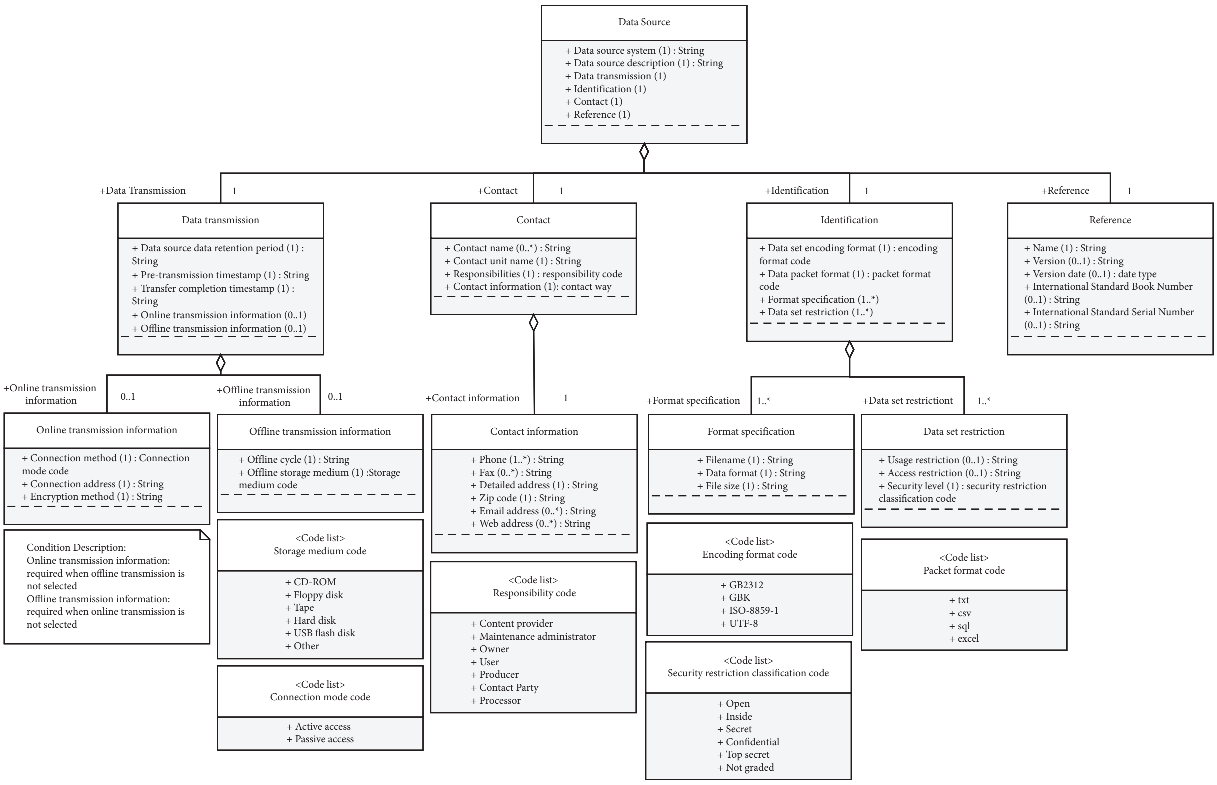

Data Source
+ Data source system (1) : String
+ Data source description (1) : String
+ Data transmission (1)
+ Identification (1)
+ Contact (1)
+ Reference (1)
+Data Transmission
1
+Contact
1
+Identification
1
+Reference
1
Data transmission
+ Data source data retention period (1) : String
+ Pre-transmission timestamp (1) : String
+ Transfer completion timestamp (1) : String
+ Online transmission information (0..1)
+ Offline transmission information (0..1)
Contact
+ Contact name (0..*) : String
+ Contact unit name (1) : String
+ Responsibilities (1) : responsibility code
+ Contact information (1): contact way
Identification
+ Data set encoding format (1) : encoding format code
+ Data packet format (1) : packet format code
+ Format specification (1..*)
+ Data set restriction (1..*)
Reference
+ Name (1) : String
+ Version (0..1) : String
+ Version date (0..1) : date type
+ International Standard Book Number (0..1) : String
+ International Standard Serial Number (0..1) : String
+Online transmission information
0..1
+Offline transmission information
0..1
+Contact information
1
+Format specification
1..*
+Data set restrictiont
1..*
Online transmission information
+ Connection method (1) : Connection mode code
+ Connection address (1) : String
+ Encryption method (1) : String
Offline transmission information
+ Offline cycle (1) : String
+ Offline storage medium (1) :Storage medium code
Contact information
+ Phone (1..*) : String
+ Fax (0..*) : String
+ Detailed address (1) : String
+ Zip code (1) : String
+ Email address (0..*) : String
+ Web address (0..*) : String
Format specification
+ Filename (1) : String
+ Data format (1) : String
+ File size (1) : String
Data set restriction
+ Usage restriction (0..1) : String
+ Access restriction (0..1) : String
+ Security level (1) : security restriction classification code
Condition Description:
Online transmission information: required when offline transmission is not selected
Offline transmission information: required when online transmission is not selected
<Code list>
Storage medium code
+ CD-ROM
+ Floppy disk
+ Tape
+ Hard disk
+ USB flash disk
+ Other
<Code list>
Connection mode code
+ Active access
+ Passive access
<Code list>
Responsibility code
+ Content provider
+ Maintenance administrator
+ Owner
+ User
+ Producer
+ Contact Party
+ Processor
<Code list>
Encoding format code
+ GB2312
+ GBK
+ ISO-8859-1
+ UTF-8
<Code list>
Security restriction classification code
+ Open
+ Inside
+ Secret
+ Confidential
+ Top secret
+ Not graded
<Code list>
Packet format code
+ txt
+ csv
+ sql
+ excel

Figure 4: Data source specification model.

sensitive information, a database encryption system can be deployed. The information stored in the database is encrypted and stored, and an independent permission control system is used to realize the permission control of sensitive data access to ensure the security of its data. Data source system refers to the system from which the data are collected. We define five business layers for coal mine data, which are mine system, subsystem, device, subdevice, and measurement point. The data source system should include two layers: mine system and subsystem. For instance, common coal mine systems are coal mining system, excavation system, drainage system, ventilation system, transportation system, and electromechanical system. Subsystems are divided by area or function. For example, the subsystems of the main drainage system are the central pump house, the pump house below the adit, and the pump house in the 121-panel. In practice, mine systems and subsystems need to be modified according to the characteristics of the colliery. In order to standardize the processing of data, for the data sources of multiple business systems, the unified standard naming management of each business system and its equipment is realized through the master data naming specification. The naming rule of the full name of the mine is the abbreviation of the group company, the full name of the branch (optional), the scope of mining rights-coal mine. The naming rule of working face is working face number-function-working face. A specific example is 123 coal mining face.

(2) *Data source description* is a list of measurement points under the devices and their subdevices to provide the source of the data. As an example, some measurement points for a drive motor. The subdevices of the drive motor are motor, reducer, inverter, and high voltage switchgear. The measurement points of the motor are *A* phase winding temperature, *B* phase winding temperature, *C* phase winding temperature, motor front shaft temperature, motor rear shaft temperature, etc. For data from different automated systems, we use a data access method based on multiple data sources. A mapping relationship is established between the source and target data to achieve a unified naming and standardized description of the data set. Mapping the source data into a standardized format avoids repetitive human standardization work and improves the speed of data standardization. The source of the data can be located by keeping log records while it is being sent. As a result, personnel may inspect the associated data collecting devices and measurement sites to solve issues like data mistakes or inaccuracies when they arise.

(3) *Data transmission* describes the necessary information for data transfer and storage. It consists of the required field data and the retention period. To

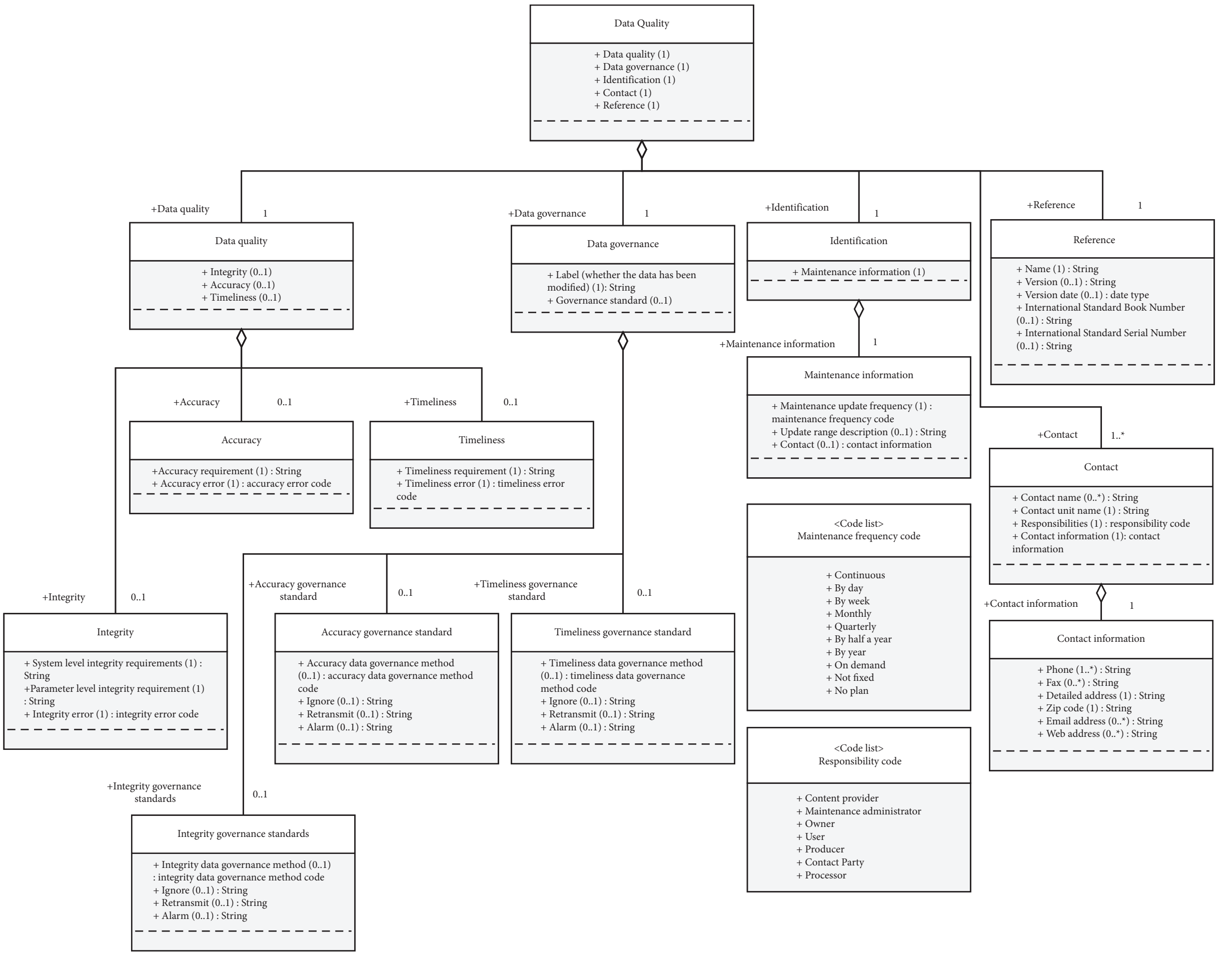

Data Quality
+ Data quality (1)
+ Data governance (1)
+ Identification (1)
+ Contact (1)
+ Reference (1)
+Data quality
1
Data quality
+ Integrity (0..1)
+ Accuracy (0..1)
+ Timeliness (0..1)
+Data governance
1
Data governance
+ Label (whether the data has been modified) (1): String
+ Governance standard (0..1)
+Identification
1
Identification
+ Maintenance information (1)
+Reference
1
Reference
+ Name (1) : String
+ Version (0..1) : String
+ Version date (0..1) : date type
+ International Standard Book Number (0..1) : String
+ International Standard Serial Number (0..1) : String
+Accuracy
0..1
+Timeliness
0..1
Accuracy
+Accuracy requirement (1) : String
+ Accuracy error (1) : accuracy error code
Timeliness
+ Timeliness requirement (1) : String
+ Timeliness error (1) : timeliness error code
+Maintenance information
1
Maintenance information
+ Maintenance update frequency (1) : maintenance frequency code
+ Update range description (0..1) : String
+ Contact (0..1) : contact information
+Contact
1..*
Contact
+ Contact name (0..*) : String
+ Contact unit name (1) : String
+ Responsibilities (1) : responsibility code
+ Contact information (1): contact information
<Code list>
Maintenance frequency code
+ Continuous
+ By day
+ By week
+ Monthly
+ Quarterly
+ By half a year
+ By year
+ On demand
+ Not fixed
+ No plan
+Integrity
0..1
Integrity
+ System level integrity requirements (1) : String
+Parameter level integrity requirement (1) : String
+ Integrity error (1) : integrity error code
+Accuracy governance standard
0..1
Accuracy governance standard
+ Accuracy data governance method (0..1) : accuracy data governance method code
+ Ignore (0..1) : String
+ Retransmit (0..1) : String
+ Alarm (0..1) : String
+Timeliness governance standard
0..1
Timeliness governance standard
+ Timeliness data governance method (0..1) : timeliness data governance method code
+ Ignore (0..1) : String
+ Retransmit (0..1) : String
+ Alarm (0..1) : String
+Contact information
1
Contact information
+ Phone (1..*) : String
+ Fax (0..*) : String
+ Detailed address (1) : String
+ Zip code (1) : String
+ Email address (0..*) : String
+ Web address (0..*) : String
<Code list>
Responsibility code
+ Content provider
+ Maintenance administrator
+ Owner
+ User
+ Producer
+ Contact Party
+ Processor
+Integrity governance standards
0..1
Integrity governance standards
+ Integrity data governance method (0..1) : integrity data governance method code
+ Ignore (0..1) : String
+ Retransmit (0..1) : String
+ Alarm (0..1) : String

Figure 5: Data quality specification model.

prevent data loss and retransmission, the data source system needs to retain the data after successful transmission. Therefore, the retention time of the data source data are the length of time that the data need to be retained in the data source system after transmission. The system needs to record the time before and after data transmission to calculate the transmission delay and verify the data retention time. In the data transmission information, the transmission method must be online or offline information transmission. The online transmission information includes five fields: connection mode, encoding format, transmission protocol, connection address, and encryption mode. From the receiver's point of view, it can be divided into active and passive connection modes. The former model means that the data source opens the query port, and the latter mode means that the data source system sends the data directly to the receiver. If the offline transmission mode is selected, the offline period and offline storage media need to be recorded. That is, the system will record the frequency of offline transmission and the media used, such as weekly or monthly transmission using a hard disk. Encryption methods can be selected from one-way encryption, symmetric encryption, hash function, and digital signature. Users can select reasonable transmission modes and encryption methods according to data characteristics to enhance the security of sharing sensitive data. Recording the whole process of data flow helps to improve the data sharing log.

(4) *Identification* records information about the data format. The specification of data sets and packets prevents the computer from being able to read the code. The data set restrictions to limit the scope and manner in which the dataset can be used. This field ensures that only compliant personnel have access to authorized-mine data, improving the security of data. The identification also requires the data set to follow certain format specifications, reduce data parsing errors, and improve efficiency through a unified file naming format. The file head should be named "coal name; system name; data upload time." Among them, the data upload time refers to the time

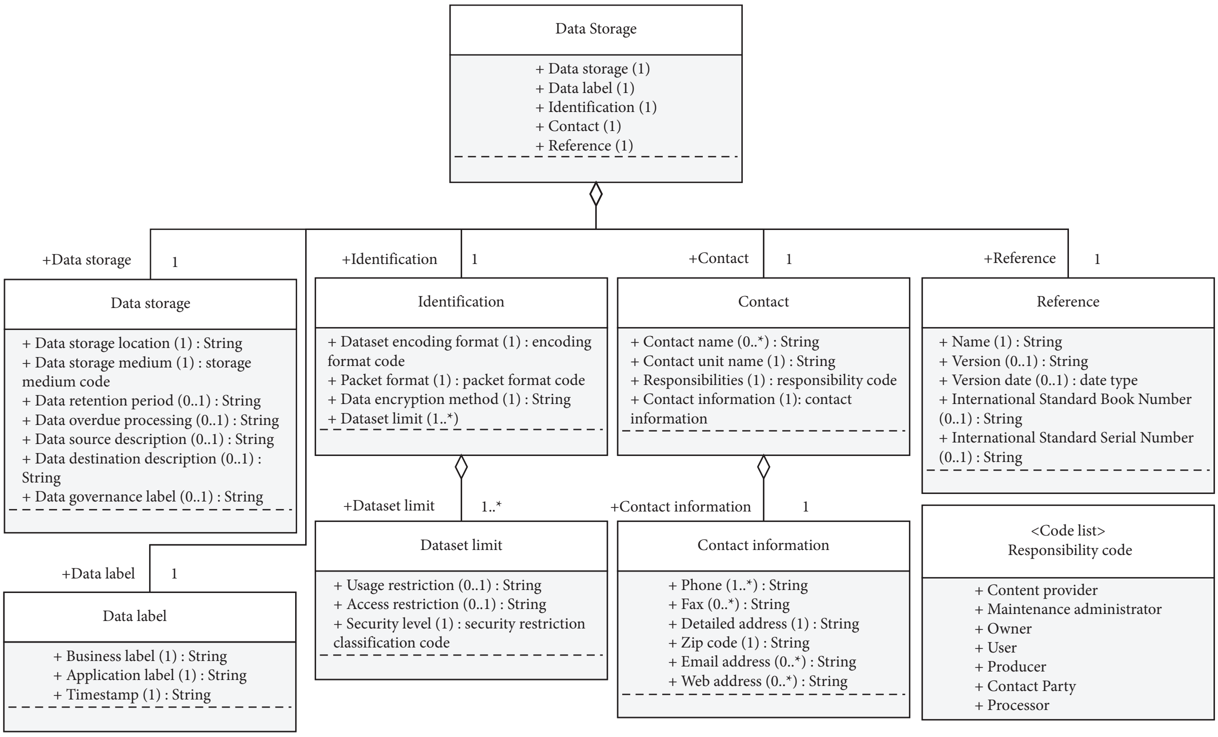

Data Storage
+ Data storage (1)
+ Data label (1)
+ Identification (1)
+ Contact (1)
+ Reference (1)
+Data storage
1
+Identification
1
+Contact
1
+Reference
1
Data storage
+ Data storage location (1) : String
+ Data storage medium (1) : storage medium code
+ Data retention period (0..1) : String
+ Data overdue processing (0..1) : String
+ Data source description (0..1) : String
+ Data destination description (0..1) : String
+ Data governance label (0..1) : String
Identification
+ Dataset encoding format (1) : encoding format code
+ Packet format (1) : packet format code
+ Data encryption method (1) : String
+ Dataset limit (1..*)
Contact
+ Contact name (0..*) : String
+ Contact unit name (1) : String
+ Responsibilities (1) : responsibility code
+ Contact information (1): contact information
Reference
+ Name (1) : String
+ Version (0..1) : String
+ Version date (0..1) : date type
+ International Standard Book Number (0..1) : String
+ International Standard Serial Number (0..1) : String
+Dataset limit
1..*
+Contact information
1
+Data label
1
Data label
+ Business label (1) : String
+ Application label (1) : String
+ Timestamp (1) : String
Dataset limit
+ Usage restriction (0..1) : String
+ Access restriction (0..1) : String
+ Security level (1) : security restriction classification code
Contact information
+ Phone (1..*) : String
+ Fax (0..*) : String
+ Detailed address (1) : String
+ Zip code (1) : String
+ Email address (0..*) : String
+ Web address (0..*) : String
<Code list>
Responsibility code
+ Content provider
+ Maintenance administrator
+ Owner
+ User
+ Producer
+ Contact Party
+ Processor

Figure 6: Data storage specification model.

of generating the data set file. The file body is a collection of measuring point data, and the data format is “measuring point; unit; upper range; lower range; upper alarm range; lower alarm range; data time.”

(5) *Contact* records the contact person and contact unit of the data source system, specifying the institutional departments and responsibilities related to data sharing. If there is a problem with the source data, managers can quickly seek help through contact information.

(6) *Reference* records the industry management methods, professional theoretical knowledge of the coal industry, and relevant technical indicators related to this specification.

Data specification information can be set using XML files. Instance 1 shows a simple example of the data source specification model with XML format. It only gives partial information about the data source specification model.

5. Data Quality Specification Model

The coal data quality standard is described and evaluated using the data quality specification model, which also ensures the accuracy and consistency of the shared data. Figure 5 shows the data quality specification model, including the following modules.

Data quality is the specific criterion for describing and evaluating data. The specific content needs to be developed based on the advice of business experts and combined with the actual situation of the coal mine and the classification of data, taking into account the three attributes of integrity, accuracy, and timeliness.

To examine data integrity at the business level, it is necessary to consider the overall and local aspects separately. System-level integrity requirements describe the specific business information contained in the entire automation system, including system information, subsystem information, and equipment information respectively. For example, the ventilation system includes 3 ventilators, 2 vertical air gates, and 2 fan oil stations. The parameter-level integrity describes all data measurement points in the system. For example, the fan needs to measure the fan blade Angle, the winding temperature and bearing temperature related to the fan motor, the wind speed and efficiency of the fan, etc.

In addition, the accuracy requirements define the corresponding data type, data range, and data length of the measured point data. The data type ensures the accuracy requirements of the data. The data range allows for evaluating the data reasonableness. For example, if the data type is Boolean, then the data have no data range. If the data are of other types with a clear threshold range, then the data range needs to be specified according to the actual situation. The data range can be developed in a variety of ways, including the parameters of the equipment itself and the expert’s estimate of the safety situation in the coal mine. For example, the upper threshold for the pool water level of the gas drainage system is 1.9 m and the lower threshold is 0.8 m. For the data with timeliness characteristics, the timeliness requirements are used to judge the quality of the data. Timeliness means that data will be recorded in chronological order and conform to certain rules of change. It is mainly

```
<?xml version = "1.0" encoding = "UTF-8"?>
<DataSource>
  <dataSourceSystem>WJL-CFTS-2CCB</dataSourceSystem>
  <dataSourceDescription>belt, drive motor and others</dataSourceDescription>
  <dataTransmission>
    <dataSourceDataRetentionPeriod>A Week</dataSourceDataRetentionPeriod>
    <preTransmissionTimestamp>131974608035554296</preTransmissionTimestamp>
    <transferCompletionTimestamp>131974608035554459</transferCompletionTimestamp>
    <onlineTransmissionInformation>
      <connectionMethod>Active Access</connectionMethod>
      <connectionAddress>192.168.100.100:8080</connectionAddress>
    </onlineTransmissionInformation>
  </dataTransmission>
</DataSource>
```

Instance 1: A simple example of the data source specification model.

judged and governed by data over a period of time. Timeliness requirements include time delay requirement and time sequence requirement. The time delay requirement is determined by the frequency of data acquisition. The optional parameters for the time sequence requirement are true and false. True means that the data has timeliness characteristics and needs to be governed using the time series algorithm. Data governance and maintenance are implemented according to integrity error codes. The error code of recorded data has a great impact on system error checking. For example, if the problem is caused by the application software, the data are likely to be skewed. Otherwise, the system should alert the contact to potential security vulnerabilities. At the same time, technicians shall check whether it is time according to the time stamp during data collection.

Data governance consists of a mandatory label and a mandatory governance standard. These mandatory governance standards come from integrity, accuracy, and timeliness. The label indicates whether the data have been modified and what specific modifications have been made. Data governance method, neglect, retransmission, and warning are four common mandatory fields covered by each governance standard. When the data are modified, the data governance method must be recorded. Data governance methods need to refer to expert experience and commonly used methods in related industries. Ignore implies that the inaccuracy is acceptable and should be disregarded. An essential field that contains an unfixable mistake has to be resent. If the data deviate greatly from the normal range, the alarm field will be enabled to inform the responsible person that there are security vulnerabilities in the system.

In the creation of specifications, many methods can be utilized to guarantee the data governance process. For sensitive data, common data desensitization methods are adopted for governance. For ordinary data, we use a density-based outlier identification approach to ensure accuracy. For time-series data, we use a time-series model to ensure timeliness.

The density-based outlier identification method is as follows. First, the data are grouped using quick search and density peak (DPC) [28] clustering algorithms. Then, any points whose distance from each center is more than or equal to the radius in the clustering procedure are picked as outlier candidate sets. Finally, an enhanced local outlier factor (LOF) [29] is employed to find outliers in the candidate outlier collection.

Formula (1) is used to locate cluster centers. Two concepts are defined. One is the sample i local density, which is denoted as ρ_i. Another is the shortest distance between the sample i and the location with a higher local density. It is denoted as δ_i.

$$\rho_i = \sum_{j \neq i} \chi\left(\mathrm{dist}(i, j) - d_c\right). \tag{1}$$

Here, $\mathrm{dist}(i, j)$ is the Euclidean distance between sample i and j. dc is a hyperparameter expressing cutoff distance. $\chi(x)$ is an activation function. $\chi(x) = 1$ when $x < 0$, else $\chi(x) = 0$.

$$\delta_i = \min_{j:\rho_j > \rho_i} (\mathrm{dist}(i, j)). \tag{2}$$

Here, δ_i is the one with the largest local distance among all samples. Those sites with larger rho_i and delta_i are selected as cluster centers.

In the second step, the improved LOF model $f(i)$ is utilized to calculate the degree of an outlier:

$$f(i) = \frac{1}{|N_k(i)|} \cdot \sum_{j \in N_k(i)} \frac{\rho_j}{\rho_i}, \tag{3}$$

$$N_k(i) = \{ j \in S \mid \mathrm{dist}(i, j) \leq \mathrm{dist}_k(i) \}. \tag{4}$$

Here, ρ_i is the local density of sample i. $N_k(i)$ is a set composed of all samples in the k neighborhoods of a sample i. Formula (3) measures the extent of outlying. For example, if $f(i)$ greater than 1, the point i is located in a sparse area. It is an outlier. Otherwise, if $f(i)$ is less than 1, the local density of sample i is higher than its neighbors. This is the normal case. The aforementioned approach may be used to obtain the samples in the outlier candidate set S. Following the sorting of set, data governance may be applied to the first n outliers.

The data governance using time series model is as follows: for common time-series data, the time series algorithm autoregressive moving average model (ARMA) can be utilized. ARMA can assess whether the data include outliers based on a realistic value range. If the data are unreasonable, we will correct it.

For normal, stationary, and zero mean time series $\{x_t\}$, if $\{x_t\}$ is connected to the value and incentive of the preceding n steps, there is a general ARMA model (formula (5) [30]. The ARMA model comprises an autoregressive model (AR) and a moving average model (MA).

$$\begin{aligned}&\alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_n x_{t-n} + x_t \\ &= \beta_1 n_{t-1} + \beta_2 n_{t-2} + \cdots + \beta_m n_{t-m} + n_t.\end{aligned} \tag{5}$$

Here, n and m are the order of autoregressive process and moving average correspondingly. The ARMA model is denoted as ARMA (n, m). If $n = 0$, the ARMA model becomes the MA model. If $m = 0$, this model is an AR model. $\alpha_i \in \mathbb{R}$ is termed the autoregressive coefficient and $\beta_i \in \mathbb{R}$ is the moving average coefficient. The series $\{n_t\}$ is the white noise sequence. Akaike information criterion (AIC) [31] is used to calculate the order of ARMA (n, m) model. The representation of AIC is given by the following formula :

$$\mathrm{AIC}(u) = \ln \hat{\sigma}^2 + \frac{2u}{N}. \tag{6}$$

Here, $u = n + m + 1$ specifies the number of model parameters. N reflects the sample size and $\hat{\sigma}^2$ is the error variance of the model. If the value of AIC is the least, then ARMA (n, m) is the best effective model for forecasting time series.

Identification describes the maintenance information of data quality. The maintenance personnel shall check and update the data quality requirements according to the maintenance and update frequency. When deploying new equipment, they need a set of requirements for timely updating data accuracy and timeliness according to the information of new equipment. *Contact* records the business experts consulted when setting up coal data quality and data governance methods. *Reference* mainly records the international standards referred to when specifying data quality standards and the instructions for equipment-related parameters.

6. Data Storage Specification Model

The data storage specification model describes the storage requirements in colliery data management, as shown in Figure 6. This facilitates the use of technologies such as storage encryption and backup to protect hierarchically classified data. It includes five modules: data storage, data label, identification, contact, and reference.

(1) *Data storage* defines data storage information. It covers the data storage location, medium, and retention time, among other things. The data storage location and the data storage medium are two required parameters that provide the particular URL route and storage media, respectively, by storing the precise URL path for improved traceability. For sensitive data, a good data storage medium needs to be selected. It also provides effective technology and management tools for data storage media to prevent data leakage due to improper use of media and improve the security of data sharing. It assists the system in identifying the precise flow of various business data by capturing the location of data. If it is necessary to retain the data for a certain period, the data retention period field needs to be set. This is an optional field to be set according to the actual requirements. The data overdue processing field is provided when the data are past due. Descriptions of data source and destination show the flow of data. The data source description specifies the data production system. It identifies the department responsible for the data. The data destination description specifies the access rights of each platform to different business data. Each platform should apply for data use to the data management department according to the authority to obtain data use authority and improve data sharing security. Data governance labels highlight the governance mechanism, whereas data labels primarily record the business system to which the data belongs. This is the distinction between the two.

(2) *Data label* includes a business label, an application label, and a timestamp. A typical mine industry business label can be divided into 4 layers: mine system, subsystem, equipment, and subequipment. Application label include worker type labels, device labels, disaster labels, operation labels, region labels, and system labels. Each piece of data can correspond to only one service label but can correspond to multiple application labels and provides a timestamp. Coal industry practitioners and IT industry practitioners can use business label and application label to rapidly query data.

(3) *Identification* shows the coding format, data packet format, and data encryption method for colliery data storage. Appropriate data encryption methods to protect the security of data sharing. To increase data security, it restricts the use of data sets and access personnel through data set constraints. It specifies the scope of data sharing scenarios and the rights holders of data sharing.

(4) *Contact* records the person responsible for storing the data. When the data are lost, we can quickly find the relevant person in charge to follow up on the situation.

(5) *Reference* includes the documents referenced and referenced in the process of formulating colliery storage standards.

7. Experiment

To demonstrate the validity and usefulness of the specification given in this work, we used data from the Wangjialing coal mine to construct a set of coal mine data collection and analysis system. The system requires three identical computers to form a cluster, and the experimental settings are presented in Table 1. We obtained data from the Wangjialing coal mine's IoT devices with the coal company's permission. The data access process in the system is designed based on the data source specification, as illustrated in Figure 7. According to the data source specification model, the data access process must save data source information, data transmission, identification, contact, and reference. This contributes to data traceability and ensures the security of data sharing. In Figure 7, more detailed information on the data source system can be viewed by clicking on the description. For example, if you click on the task with the id is three, you can know that the coal mine is Wangjialing coal mine, the system is main ventilation monitoring system, and the subsystem is mine ventilation room. The data source data retention period is a week. The pretransmission timestamp is 182984608043, and the transfer completion timestamp is 182984608582. The data are transferred online, and the connection method is active access. The contact information of equipment contains the equipment manufacturer and phone number and so on.

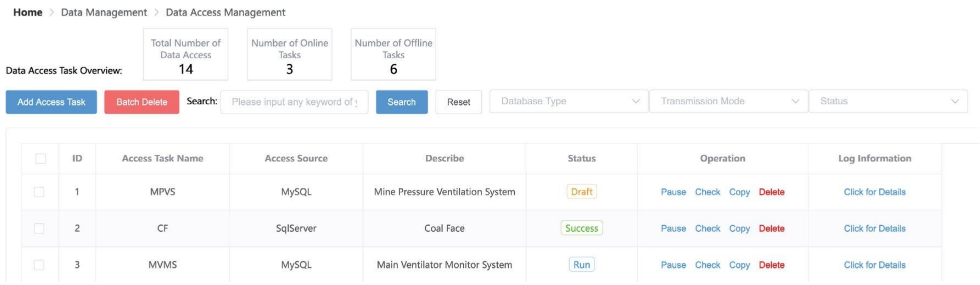
Due to the wide range of data sources, including databases, files, and sensors, the data naming is not uniform. In order to unify the field names of coal mine data and record the data source system according to the data source specification, the data access process needs to implement the data mapping function. The data mapping function of the system is shown in Figure 8. After analyzing the data source file, select the data source file and the corresponding coal mine, system, subsystem, equipment, and field to make them correspond one by one. After completing these tasks, a mapping relationship will be generated between the source data and the target data, that is, a new data access task will be created. The target data can also form a uniform naming standard. The completed mapped data is encrypted by the data encryption standard (DES) algorithm and is then securely transferred to the storage platform.

In the data quality field, data governance is done as the data is being transferred. Data quality criteria for each type of data are defined based on the expertise of the experts and the device specifications. The data quality criteria for the 10 kV incoming cabinet are shown in Figure 9 by taking into account all factor types. This diagram shows the screenshot of the data quality standards of the coal mine data collection and analysis system. This figure shows the data quality specification model for the equipment 10 kV incoming cabinet, covering data quality, identification, and reference information. Completeness indicates this measurement point data for this device is present. Accuracy includes the data type and data range. 10 kV incoming cabinet only has an upper bound for every measurement point. The threshold range and reasonable range of the data determines how the data are processed, specifying whether the data should be governed, ignored, or alarmed. The version in the figure ensures the reference information of the data quality specification model. Each modification by the user will update the version number of the data quality standard and record the updated range description, the modifier, and the date of modification. By clicking on data governance, you can also see the corresponding governance methods and reference information referenced by the setting of the standard. By clicking on data governance, the governance method may also be seen. The data governance process is the next phase, which is determined by the data quality requirements. We have created comparable standards based on distinct parameter features. In order to verify the reliability of the data governance methods mentioned above, we have used the data of 10 kV incoming cabinet and motor as an example for illustration. The size of the dataset is 1000. For generic data (e.g., shaft temperature of the motor), we utilize the aforementioned density based outlier identification approach to find. The results are shown in Figure 10. Outliers that need to be handled are the data points in the red circle in the figure. The data are then adjusted using the mean, median, or other methods. For common time-series data such as current and voltage, the time series algorithm autoregressive moving average model (ARMA) is utilized. Figure 11 illustrates the results of residual analysis on line voltage data. The standardized residuals demonstrate that there is no shifting variation throughout time. The autocorrelation function (ACF) of the residuals suggests no autocorrelations. The Q-Q plot is a normal probability plot that demonstrates that the data conform to a normal distribution. The preceding research reveals that the ARMA model may be utilized to identify voltage data.

Table 1: Experiments parameters setting.

Parameters	Description
System	Windows 10
CPU	Intel core i7
GPU	NVIDIA GeForce GTX 1080
RAM	16 GB

As illustrated in Figure 12, certain data governance outcomes about the exhaust temperature of the pressure fan. It can be shown from the results that the specification can ensure data quality in the big data system. The governed data and associated information will be kept. The database will record the data storage location, retention period, overage processing method, data destination description, identification information, contact information, and reference information. For example, the data storage location is htpps://ip:9000/data/WJH-MVMS. The data retention period is a month. The data destination description is an algorithm platform. During data sharing, only the algorithmic platform and its users are authorized to access and use the data of this subsystem. The system will also tag the data with a data governance label, data label, and application label, recording the governance technique, business system, and data category. In addition, the system includes a security access control function. This module is responsible for authenticating the user's operation rights and only users with login rights can access and operate the data to achieve secure data sharing.



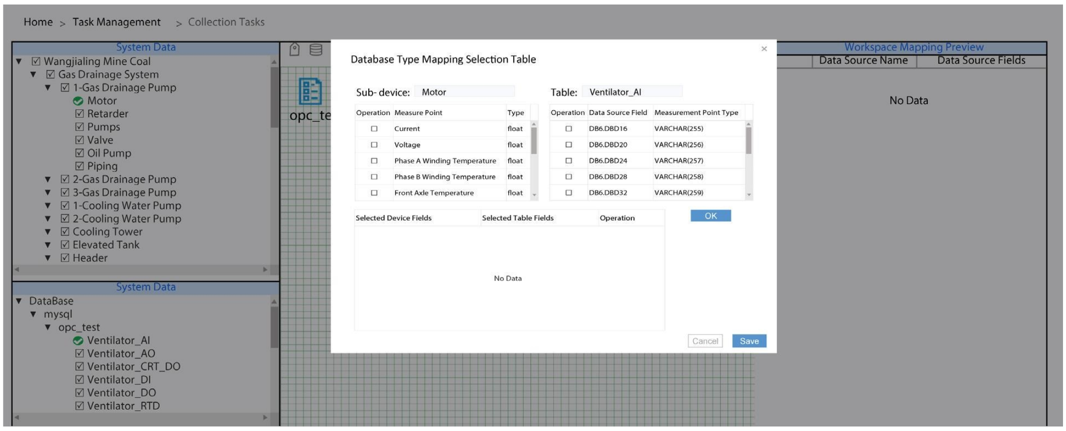
Home > Data Management > Data Access Management

Data Access Task Overview:

Total Number of Data Access	Number of Online Tasks	Number of Offline Tasks
14	3	6

□	ID	Access Task Name	Access Source	Describe	Status	Operation	Log Information
□	1	MPVS	MySQL	Mine Pressure Ventilation System	Draft	Pause Check Copy Delete	Click for Details
□	2	CF	SqlServer	Coal Face	Success	Pause Check Copy Delete	Click for Details
□	3	MVMS	MySQL	Main Ventilator Monitor System	Run	Pause Check Copy Delete	Click for Details

Figure 7: Screenshot of the data access function in the coal mine data collection and analysis system.



Home > Task Management > Collection Tasks

System Data

▾ ☑ Wangjialing Mine Coal
▾ ☑ Gas Drainage System
▾ ☑ 1-Gas Drainage Pump
✓ Motor
☑ Retarder
☑ Pumps
☑ Valve
☑ Oil Pump
☑ Piping
▾ ☑ 2-Gas Drainage Pump
▾ ☑ 3-Gas Drainage Pump
▾ ☑ 1-Cooling Water Pump
▾ ☑ 2-Cooling Water Pump
▾ ☑ Cooling Tower
▾ ☑ Elevated Tank
▾ ☑ Header

System Data

▾ DataBase
▾ mysql
▾ opc_test
✓ Ventilator_AI
☑ Ventilator_AO
☑ Ventilator_CRT_DO
☑ Ventilator_DI
☑ Ventilator_DO
☑ Ventilator_RTD

Database Type Mapping Selection Table

Sub- device: Motor

Operation	Measure Point	Type
□	Current	float
□	Voltage	float
□	Phase A Winding Temperature	float
□	Phase B Winding Temperature	float
□	Front Axle Temperature	float

Table: Ventilator_AI

Operation	Data Source Field	Measurement Point Type
□	DB6.DBD16	VARCHAR(255)
□	DB6.DBD20	VARCHAR(256)
□	DB6.DBD24	VARCHAR(257)
□	DB6.DBD28	VARCHAR(258)
□	DB6.DBD32	VARCHAR(259)

Selected Device Fields	Selected Table Fields	Operation

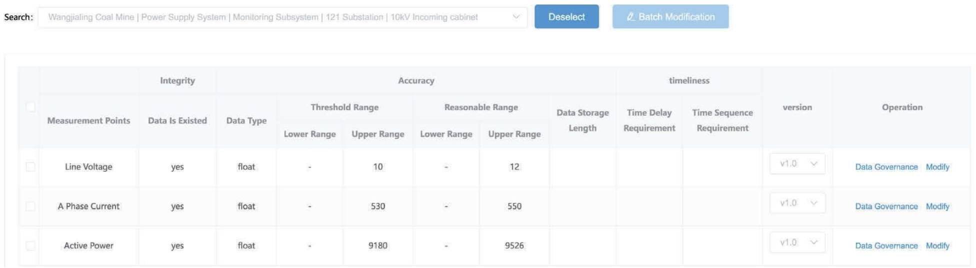
No Data

Workspace Mapping Preview

Data Source Name	Data Source Fields

No Data

Figure 8: Screenshot of the data mapping process in the coal mine data collection and analysis system.



Search: Wangjialing Coal Mine | Power Supply System | Monitoring Subsystem | 121 Substation | 10kV Incoming cabinet

		Integrity	Accuracy					timeliness				
□	Measurement Points	Data Is Existed	Data Type	Threshold Range		Reasonable Range		Data Storage Length	Time Delay Requirement	Time Sequence Requirement	version	Operation
				Lower Range	Upper Range	Lower Range	Upper Range					
□	Line Voltage	yes	float	-	10	-	12				v1.0	Data Governance Modify
□	A Phase Current	yes	float	-	530	-	550				v1.0	Data Governance Modify
□	Active Power	yes	float	-	9180	-	9526				v1.0	Data Governance Modify

Figure 9: Screenshot of the system for 10 kV incoming cabinet data quality requirements.

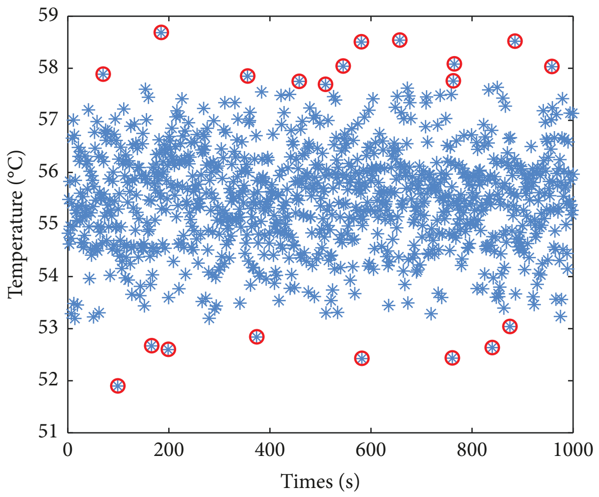


Figure 10: For the measure point of electrical machinery shaft temperature, a density-based outlier identification approach was used to identify the outliers and mark them with red circles to form the scatter diagram of the electrical machinery shaft temperature.

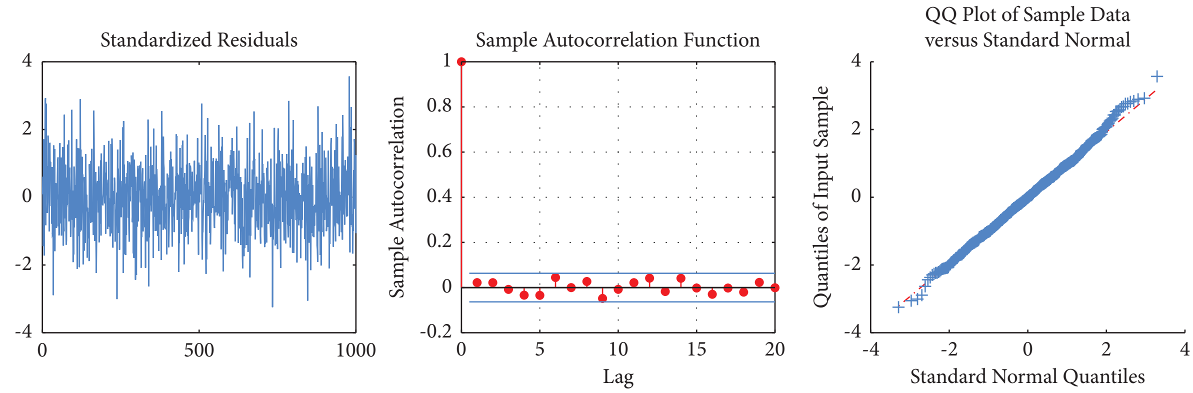
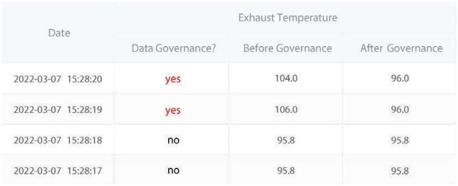


FIGURE 11: For the time sequence data of voltage, we use the ARMA model to judge the timeliness of the data and form a residual analysis diagram. It can be seen in the figure that the data are time series, and the ARMA model can be used to judge and govern the data.



Date	Exhaust Temperature		
	Data Governance?	Before Governance	After Governance
2022-03-07 15:28:20	yes	104.0	96.0
2022-03-07 15:28:19	yes	106.0	96.0
2022-03-07 15:28:18	no	95.8	95.8
2022-03-07 15:28:17	no	95.8	95.8

FIGURE 12: Screenshot of the governance results of the exhaust temperature of pressure fan in the data system.

8. Conclusions

The area of smart coal mining is growing quickly, and the enormous growth in data size creates a great demand for data management and data sharing. The lack of data standards causes inefficient use of computer and human resources and costly costs. To address these challenges, we have carried out the following: first, we offer a set of data specifications for data collection, transmission, and storage for big data practices in the coal mining sector. To improve the generality of data and the security of data sharing, our specification provides a complete data model that fills the gaps in data collection, transmission, governance, sharing, and storage according to the characteristics of the mining sector. Data are divided into business and application categories, and data tags are used to identify the category to which the data belongs, clarifying the scope of data sharing. Both those in the coal mining industry and those in the IT industry can easily find the data they need, allowing them to benefit from the specification. The standard sets up business-level classifications that allow employees to quickly track the source of anomalous data. Special governance rules for sensitive data ensure the security of sensitive data in the sharing process. Appropriate data encryption algorithms and transmission methods are selected according to the data transmission needs of different platforms to ensure the security of data sharing. Second, for the specification, we designed a short XML example for the data source model. All data criteria can be set and imported into the system based on this example. Third, we constructed a coal mine data collection and analysis system based on the standards of the data specification. The access, mapping, governance, sharing, and storage processes of data processing were implemented in the system. Experimental results show that the system verified the validity, usefulness, and security of the specification. Appropriate data encryption algorithms and transmission methods are selected according to the data transmission needs of different platforms to ensure the security of data sharing. In the future, we will try to combine microservices, privacy computing, and other technologies based on this specification to design multi-source data security sharing solutions that can meet cross-industry requirements.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

L.T. and H.W. conceptualized the study; L.T., Y.Z., and T.H. proposed the methodology; H.W., S.Z., and Y.R. managed the software; Y.Z. investigated the study; L.T. gathered the resources; Y.Z. curated the data; H.W. and Q.G. wrote the original draft; Q.G. reviewed and edited the manuscript; and Q.G. visualized the study. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The study was supported by Research and Development of Coal Mine Big Data Platform with Artificial Intelligence Applications' program of China Coal Energy Research Insititute Co. Ltd.

References

[1] L. Xianglan, "Digital construction of coal mine big data for different platforms based on life cycle," in *Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pp. 456–459, IEEE, Beijing, China, March 2017.

[2] Z. Qin, S. Chen, X. Xu, and M. Zhao, "Research on key technologies and system construction of smart mine," in *Proceedings of the 2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, pp. 116–121, IEEE, Singapor, July 2020.

[3] M. Zhao, "Technology of internet of things technology in the construction of smart mine," in *Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 289–292, IEEE, Chengdu, China, May 2020.

[4] X.-D. Hao and H.-W. Yang, "The research of the data platform system for the systemwide supervision and management of the intelligent mine in the opencast coal mine," in *Proceedings of the 2nd International Conference on Electrical and Electronic Engineering (EEE 2019)*, pp. 287–290, Atlantis Press, Penang, Malaysia, May 2019.

[5] L. Dong, M. N. Satpute, J. Shan, B. Liu, Y. Yu, and T. Yan, "Computation offloading for mobile-edge computing with multi-user," in *Proceedings of the 2019 IEEE 39th international conference on distributed computing systems (ICDCS)*, pp. 841–850, IEEE, Dallas, TX, USA, July 2019.

[6] P. Wang, L. Dong, Y. xu, W. Liu, and N. Jing, "Clustering-based emotion recognition micro-service cloud framework for mobile computing," *IEEE Access*, vol. 8, pp. 49695–49704, 2020.

[7] L. Liu, C. Chen, Q. Pei, S. Maharjan, and Y. Zhang, "Vehicular edge computing and networking: a survey," *Mobile Networks and Applications*, vol. 26, no. 3, pp. 1145–1168, 2020.

[8] L. Liu, J. Feng, Q. Pei et al., "Blockchain-enabled secure data sharing scheme in mobile-edge computing: an asynchronous advantage actor–critic learning approach," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2342–2353, 2021.

[9] L. Dong, W. Wu, Q. Guo, M. N. Satpute, T. Znati, and D. Z. Du, "Reliability-aware offloading and allocation in multilevel edge computing system," *IEEE Transactions on Reliability*, vol. 70, no. 1, pp. 200–211, 2021.

[10] L. Dong, Q. Ni, W. Wu, C. Huang, T. Znati, and D. Z. Du, "A proactive reliable mechanism-based vehicular fog computing network," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11895–11907, 2020.

[11] L. Liu, M. Zhao, M. Yu, M. A. Jan, D. Lan, and A. Taherkordi, "Mobility-aware multi-hop task offloading for autonomous driving in vehicular edge computing and networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 1–14, 2022.

[12] L. Dong, Q. Guo, and W. Wu, "Speech corpora subset selection based on time-continuous utterances features," *Journal of Combinatorial Optimization*, vol. 37, no. 4, pp. 1237–1248, 2019.

[13] L. Dong, M. N. Satpute, W. Wu, and D.-Z. Du, "Two-phase multidocument summarization through content-attention-based subtopic detection," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 6, pp. 1379–1392, 2021.

[14] C. Liu, X. Su, and C. Li, "Edge computing for data anomaly detection of multi-sensors in underground mining," *Electronics*, vol. 10, no. 3, p. 302, 2021.

[15] Z. Xu, J. Li, and M. Zhang, "A surveillance video real-time analysis system based on edge-cloud and fl-yolo cooperation in coal mine," *IEEE Access*, vol. 9, p. 68482, 2021.

[16] Y. Meng and J. Li, "Task offloading and resource allocation mechanism of moving edge computing in mining environment," *IEEE Access*, vol. 9, p. 15534, 2021.

[17] J. Feng, L. Liu, Q. Pei, and K. Li, "Min-max cost optimization for efficient hierarchical federated learning in wireless edge networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 56, no. 1-1, p. 1, 2022.

[18] X. Li, H. Qi, and J. Wu, "Node social nature detection osn routing scheme based on iot system," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 14048–14059, 2022.

[19] W. Yang, J. Luo, and J. Wu, "Application of information transmission control strategy based on incremental community division in iot platform," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21968–21978, 2021.

[20] ISO, *2003 Geographic Information-Metadata*, International Organization for Standardization (ISO), Geneva, Switzerland, 2003.

[21] M. A. Ureña-Cámara, J. Nogueras-Iso, J. Lacasta, and F. J. Ariza-López, "A method for checking the quality of geographic metadata based on iso 19157," *International Journal of Geographical Information Science*, vol. 33, no. 1, pp. 1–27, 2019.

[22] G. D. Bader and C. W. Hogue, "Bind—a data specification for storing and describing biomolecular interactions, molecular complexes and pathways," *Bioinformatics*, vol. 16, no. 5, pp. 465–477, 2000.

[23] G. Yu and J. Wu, "Efficacy prediction based on attribute and multi-source data collaborative for auxiliary medical system in developing countries," *Neural Computing and Applications*, vol. 34, no. 7, pp. 5497–5512, 2022.

[24] J. Wu, L. Chang, and G. Yu, "Effective data decision-making and transmission system based on mobile health for chronic disease management in the elderly," *IEEE Systems Journal*, vol. 15, no. 4, pp. 5537–5548, 2021.

[25] Q. Qian, Y. Wang, and S. Zhao, "Materials data specification: methods and use cases," *Computational Materials Science*, vol. 169, p. 11, Article ID 109086, 2019.

[26] H. Moeini, W. Zeng, I.-L. Yen, and F. Bastani, "Toward data discovery in dynamic smart city applications," in *Proceedings of the 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City*, pp. 2572–2579, IEEE, Sydney, Australia, Octber 2019.

[27] H. Wang, L. Tan, Y. Zhang et al., "A management specification for big data sharing in smart mine," in *Proceedings of the 2022 International Conference on Computing, Communication, Perception and Quantum Technology (CCPQT)*, pp. 313–318, IEEE Computer Society, Los Alamitos, CA, USA, August 2022.

[28] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[29] M. Breunig, H.-P. Kriegel, R. Ng, and J. Sander, "Optics-of: identifying local outliers," in *Proceedings of the 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases*, Grenoble, France, September 2002.

[30] S. Lotfan and R. Fathi, "Parametric modeling of carbon nanotubes and estimating nonlocal constant using simulated vibration signals-arma and ann based approach," *Journal of Central South University*, vol. 25, no. 3, pp. 461–472, 03 2018.

[31] L. Ljung, "System identification," *Theory For The User*, vol. 12, 1999.