

## *Retraction*

# **Retracted: Data Integration Method Design of Decision Spatial Information System**

### **Security and Communication Networks**

Received 30 January 2023; Accepted 30 January 2023; Published 2 February 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Security and Communication Networks* has retracted the article titled “Data Integration Method Design of Decision Spatial Information System” [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The author does not agree to the retraction.

### **References**

- [1] D. Li, “Data Integration Method Design of Decision Spatial Information System,” *Security and Communication Networks*, vol. 2022, Article ID 8114454, 12 pages, 2022.
- [2] L. Ferguson, “Advancing Research Integrity Collaboratively and with Vigour,” 2022, <https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/>.

## Research Article

# Data Integration Method Design of Decision Spatial Information System

Daqian Li 

State Key Laboratory of Eco-Hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an 710048, Shaanxi, China

Correspondence should be addressed to Daqian Li; 1170410017@stu.xaut.edu.cn

Received 18 April 2022; Revised 11 May 2022; Accepted 20 May 2022; Published 21 June 2022

Academic Editor: Mohammad Ayoub Khan

Copyright © 2022 Daqian Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of technology, more and more enterprises have begun to use data integration analysis to make decisions. The decision spatial information system is an information system that provides decision support after analyzing and processing the spatial information. Data integration refers to a method of reasonably centralizing different or multisource data for data sharing, which can assist in decision management. This paper aims to study a more efficient data integration method for decision space information system. In this paper, an emergency information decision-making system for digital water network is designed, and a specific data integration algorithm is proposed. And the performance test of the early warning capability and power consumption of the system is carried out. The results show that the early warning time of the system designed in this paper is within 1 to 4 seconds, while the early warning time of the traditional system is within 1.5 seconds to 8 seconds. The power consumption of the system designed in this paper is not more than 25 W, while the traditional system power consumption is up to 50 W. The power consumption of the system in this paper is about half of the traditional system. Compared with the traditional system early warning capability, the system designed in this paper is significantly faster, the system power consumption is lower, and the system is more stable.

## 1. Introduction

In recent years, the availability of vast amounts of information, generated at high speed in every sector in various forms and formats, has been unprecedented. Its ability to leverage big data is an opportunity to gain more accurate analysis and improve decision-making in industry, government, and many other organizations. However, dealing with big data can be challenging, and proper data integration is a key factor in achieving high information quality.

According to the World Economic Forum report, about 70% of the data generated is not used. This usage limitation is mainly due to lack of interoperability and data links in isolated silos. In fact, large volumes of data have exacerbated the heterogeneity problem, as have the source types that generate data in heterogeneous formats and different semantics. These data-related problems are common in the Earth observation (EO) field. Earth observation data use

different terms that are difficult to reconcile because they reflect overlapping disciplines. These issues lead to misunderstandings in access, pricing, and data rights, as well as inefficiencies in data exchange and management, impeding the understanding of environmental phenomena. Data mining is a technique of data integration from available data documents. Pattern targeting from the data and document associations of documents is a well-known problem in data mining. The analysis of data content and the classification of documents are comprehensive tasks of data mining. Some of them are supervised, while others are unsupervised ways of compiling files. The term “federated database” refers to the sequential integration of distributed, autonomous, and heterogeneous databases. However, federations can also include information systems, not just databases. When integrating data, there are more than a few issues that must be addressed. Here, one needs to focus on the issue of heterogeneity, more specifically semantic heterogeneity, that is,

issues related to semantically equivalent concepts or semantically related or unrelated concepts. Its classification solves this problem.

Guaranteeing the quality of integrated data is undoubtedly one of the main problems of integrated data systems. When looking at multicountry and historical data integration systems, it is important to build an integration layer where the “space” and “time” dimensions play a relevant role. The layers accessible by the end-user are as complete as possible “by design.”

The innovation of this paper is to design a digital water network emergency decision information system. In this paper, the performance test of early warning capability and power consumption of the system is carried out, and the results obtained have reference value and significance.

## 2. Related Work

Regarding the method of data integration, many scholars have carried out related research on it. Among them, Shibuo et al. combined seamlessly integrated river, urban, and coastal hydraulic models with flood control facility information and finely resolved rainfall synthesis [1]. Mk A proposed an integration profile. Aiming to design and develop an intelligent approach, he integrated heterogeneous data and knowledge sources into personalized healthcare. The integrated profile proposed by Khovrichev et al. mainly focused on remote monitoring of patients with arterial hypertension and chronic heart failure [2]. Park et al. conducted two surveys independently for certain regions and proposed a measurement error model approach to integrate mean estimates obtained from the two surveys. The predicted values of the counterfactual results are used to create a composite estimate for the overlapping area. They provided the application of the technology in the project [3]. Masmoudi et al. proposed a knowledge hypergraph-based data integration and query method and applied it to Earth observation data. Masmoudi et al. proposed to proceed in two stages, namely, virtual data integration based on a knowledge hypergraph and query processing based on a hypergraph. The obtained results show that the proposal enhances query processing in terms of accuracy, completeness, and semantic richness of responses [4]. Mekala used the idea of ontology as a tool for data integration. Mekala clarified this concept and briefly explained a technique for building an ontology using a hybrid ontology approach [5]. Daraio et al. presented a method for accessing data in multipurpose data infrastructures such as data integration systems. This approach has the following properties: (i) It eliminates the need for end-users to access a single source of data. At the same time, (ii) it ensures that the amount of information available to users is maximized at the integration layer. Its method is based on the integrity-aware ensemble method. This approach allows the user to prepare all the maximum information available from an integrated data system without the need for preliminary data quality analysis of each database contained within it. A case study by Daraio C on research infrastructure for scientific and innovative research demonstrated the usefulness of the proposed approach [6]. Dalla Valle and Kenett proposed a new data integration

method that can calibrate online generated big data using interview-based customer survey data. Dalla Valle and Kenett illustrated how this integration improves the information quality of data analysis work in four InfoQ dimensions (i.e., data structure, data integration, temporal dependencies, and temporal ordering of data and targets) [7]. Although the above methods of data integration have achieved some results in some fields, they have limitations. And most of them are theoretical studies, with fewer actual tests and applications. And this paper will design a decision space information system through data integration. By testing the system, it can see how it actually works. Therefore, it is of great significance and practical value.

## 3. Data Integration Method of Decision Space Information System

### 3.1. Decision Space Information System

*3.1.1. Spatial Information.* Spatial information mainly includes geographic information, GIS information, and geospatial information, as shown in Figure 1. Geospatial information is information with geographic coordinates. It has a wider range of data on the Earth, including cloud, atmosphere, soil, and other data. GIS information mainly refers to the data involved in the field of GIS research, which is usually generated and managed by the geographic information system. It includes remote sensing image data, vector data, and UAV data. Geospatial data mainly refers to data with a spatial reference relationship with the geospace [8, 9].

*3.1.2. Decision-Making System.* The water conservancy information decision-making system is an indispensable part of the water conservancy work, and it is an important means for the managers and decision-makers of the water conservancy department to obtain the basic situation of the current water conservancy in the area. It can provide the basis for orders issued in peacetime and emergency situations. By establishing a water conservancy information decision-making system, users can access the geographic information services provided in any area, expanding the scope of use. The traditional decision-making method is generally the combination of manual and office automation software, which is inefficient and causes feedback lag. Compared with the traditional decision-making method, the information-based decision-making system is of great significance because of its outstanding advantages [10].

Informatization data integration improves the efficiency of data collection. The implementation process of the traditional collection method is to make data forms, issue forms, organize relevant personnel to fill in the forms, input the collected data, and perform statistical analysis and processing. The whole decision-making process is cumbersome, time-consuming, and expensive. The use of an information-based decision-making system omits many complicated processes and can effectively improve work efficiency. Informatization data integration improves data security. Traditional data collection methods have security

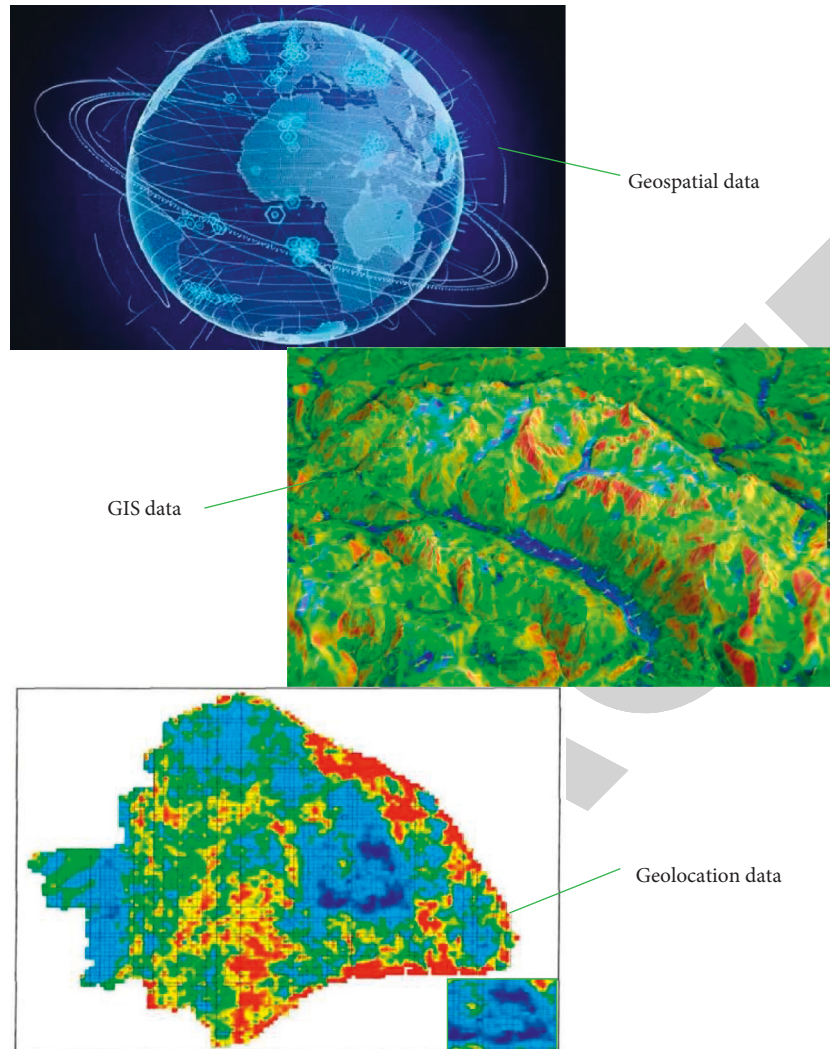


FIGURE 1: Spatial information.

risks. If the staff can see the data, modify the data, or leak the data when entering the data, the whole process involves many people, and the responsibility is difficult to identify. Information collection improves the accuracy of data. There are many errors in the process of data entry and processing in the traditional acquisition method. For example, staff are prone to problems when entering data, resulting in inaccurate results. Although this problem should not arise, it is difficult to control during operation. In the network acquisition, the whole process is directly connected with the sensor by the computer network, which shortens the period of data collection. The collected relevant data is managed by a unified and powerful database, which has high reliability and validity. It makes saving, finding, updating, and maintaining data more convenient [11, 12].

### 3.2. Data Integration Methods

**3.2.1. The Concept and Characteristics of Data Integration.** Data integration refers to the reasonable concentration of different or multisource data for data sharing. It can assist

enterprises in decision-making management. As for the source of data, it is mainly through the big data window obtained by the computer. Among them, IoT data is one of the most important data sources today. The structure of the Internet of Things can be shown in Figure 2. IoT data has the main characteristics of wide distribution, large data scale, high data rate, multiheterogeneity of data, high data authenticity, and nonspatiality [13].

Multisource means that the data comes from different sources. The data generated based on different platforms, different data production units, and different data formulation standards become multisource data. For spatial geographic data, its multisource is more obvious. In China, there is no unified department and standard for spatial geographic data collection. These data will be very different in scale, projection, spatial reference, data format, and so on. This often results in data integration between ministries and ports in different regions according to their own business standards. Moreover, because the hardware of data acquisition and the software of data processing are different, the content and method of data expression will also be very different [14].



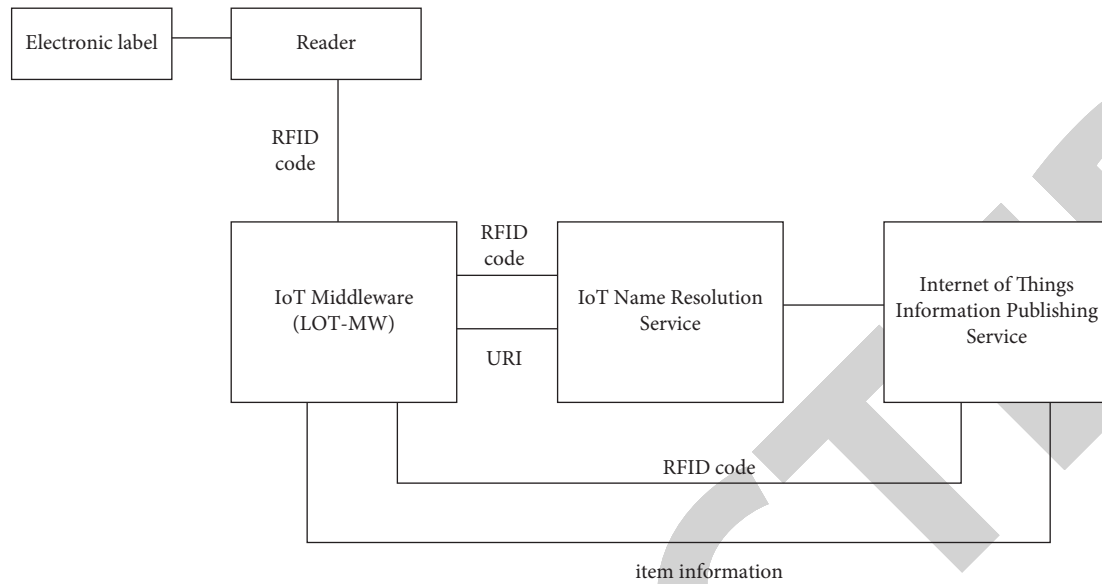


FIGURE 2: Schematic diagram of the structure of the Internet of Things.

The multisource of data is mainly manifested in the following: (1) Multisemantics: for anything on the Earth, in addition to the spatial location attribute information, it also has other physical characteristics, geometric characteristics, humanistic characteristics, and economic characteristics. In this way, the expression of the same thing in spatial data will be different based on its own application field, and the language expression will also be different. (2) Multispace-time and multiscale: geospatial data has multispace-time and multiscale properties. It is mainly manifested in the dynamic changes of geospatial objects over time. In order to continuously and completely express this dynamic change, it is required to endow the geospatial data with time characteristics under a certain time series. (3) Diverse acquisition methods: with the rapid development of measurement technology, remote sensing technology, UAV technology, GPS technology, and IoT sensor technology, data collection methods are diverse. (4) Diverse data processing and storage methods: based on their own business needs, each software manufacturer has great differences in data model, presentation content, and storage format. It also results in the characteristics of multisource data [15].

### 3.2.2. Types of Data Integration.

- (1) Geospatial data integration: Geospatial data is the data source and data basis of the entire digital water network decision-making spatial information system. In particular, remote sensing image data can accurately and timely reflect the temporal and spatial dynamic change information of crops in the planting area. And remote sensing image data has the characteristics of different resolutions, different time scales, and different formats due to different sensor types. Over time, it will constitute a massive dataset. It needs to be integrated and scientifically managed.

It has the characteristics of massive, real-time, and multisource geospatial data.

- (2) Integration of UAV data: UAV technology has also been widely used in the field of digital water network emergency management. For water conservancy conditions, field investigation and other aspects have a unique role. In the integration of UAV data, it mainly involves the preprocessing of UAV data, such as sag change processing, air belt deviation correction, and image stitching. The preprocessed UAV data, like other types of data, can be directly accessed by users through the same data management platform [16].
- (3) Integration of spatial data and IoT data: since IoT data itself does not have geospatial information, it is objectively spatially distributed. Therefore, the integration of spatial data and IoT data plays an important role in reflecting the spatiotemporal dynamic information of regional crops. This paper organically associates geospatial data with IoT data and endows IoT data with spatial geographic information to realize the integration of spatial data and physical network data. The integration of IoT data itself mainly utilizes Web Service technology (Web Service is an independent, low-coupling platform. It contains programmable web applications. It can use open XML standards to describe, publish, discover, coordinate, and configure these applications. It is used to develop distributed interoperable applications. It enables different applications to run on different machines without the need for additional, specialized third-party software or hardware. They can exchange data with each other or integrate) and SOAP technology (SOAP is a platform-independent, and it does not depend on the programming language. It is simple, flexible, and easy to extend simple

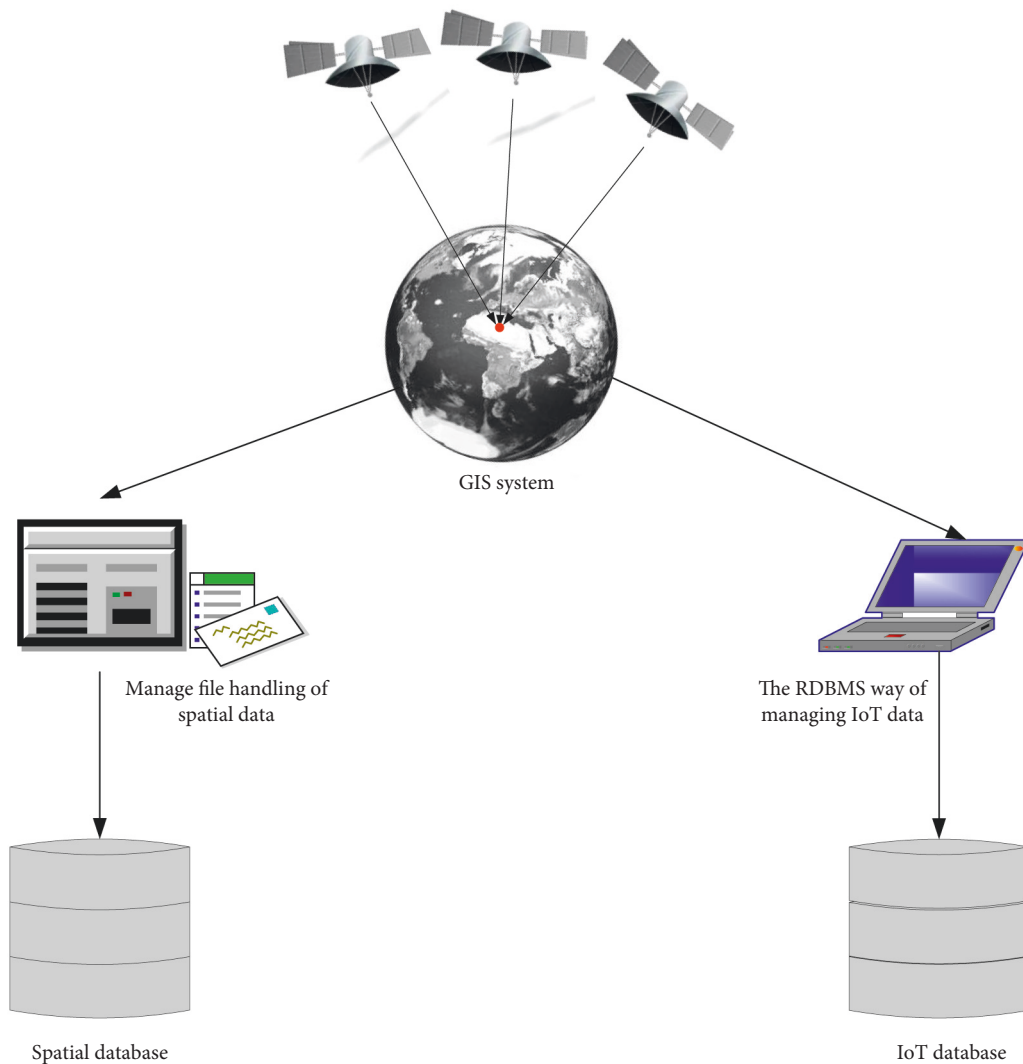


FIGURE 3: Framework structure of spatial data and IoT data integration.

object access protocol). It realizes data transmission through wireless network, 3 G/4 G network, GPRS network, Ethernet, and other transmission channels. And on the data platform, through the corresponding data access interface, the IoT data is centrally and unifiedly managed in the relational database. Figure 3 shows the framework structure of spatial data and IoT data integration [17,18].

- (4) Digital water network spatial data service center system: a digital water network or “smart” water network refers to a water supply network that uses sensors and IoT technology. It can provide more additional functions, which enable managers and users to maintain, operate, and use the water supply network more efficiently. This paper applies the three related technical methods of geospatial data integration, spatial data and IoT data integration, and UAV data integration to the digital water network

spatial data service through system development. It is also in the development process of the system. It organizes and manages these multisource and heterogeneous data scientifically and effectively. It provides a stable and reliable data source for the digital water network emergency decision-making spatial information system [19].

### 3.2.3. Common Methods of Data Integration.

- (1) Data interoperability mode: As shown in Figure 4, the data interoperability mode is committed by the Open Geospatial Information Alliance to provide the standardization of software, data, and services in the geographic information industry. Many commercial GIS software companies and data service providers are currently advocating to follow the OGC model for data interoperability. The main problem with this

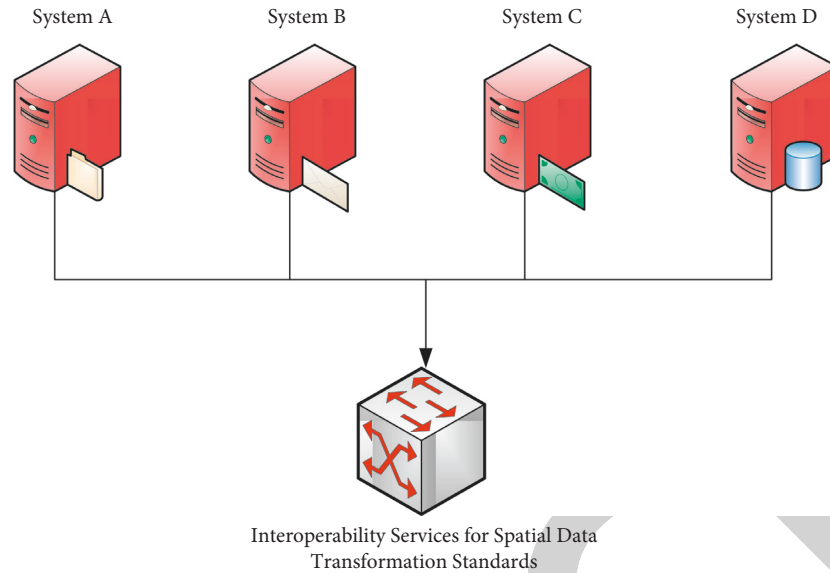


FIGURE 4: Data interoperability services.

approach is that the OGC standard is not widely used around the world [20, 21].

- (2) Multisource data adapter mode: as shown in Figure 5, this method provides access interfaces to other data in different formats through software. It enables users to directly access different types of data. There is also a big problem with this approach. By writing a driver program that accesses data to parse the data, it achieves the purpose of adapting to the unified data access interface. It greatly increases the development cost and maintenance cost of the software [22, 23].
- (3) Mutual conversion between formats: as shown in Figure 6, this method mainly lies in the direct format conversion of spatial data of two different formats. Direct format conversion will inevitably result in the loss of data and information due to the different emphases of content expression between data. And this method requires frequent format conversion, which is very cumbersome. This is also the shortcoming of this integration method [24, 25].

In terms of spatial data integration, the shortcomings of several mainstream multisource heterogeneous data integration methods are shown in Table 1.

**3.2.4. Research Status of Data Integration.** Compared with foreign countries, China's GIS technology started relatively late, and there is no unified data format specification. However, in recent years, with the vigorous promotion of the Ministry of Science and Technology, the Ministry of Industry and Information Technology, and a large number of enterprises, institutions, and universities in the field of GIS development, China's GIS has developed rapidly. In terms of the integration of multisource spatial data, Chinese researchers have also done a lot of research, realizing the seamless integration of spatial data. It uses metadata-driven, semantic transformation, and other methods to solve

middleware problems and improve the efficiency of data integration [26].

However, most of China's research on spatial data integration technology starts from its own practical application. In terms of practical applications, many mature data integration platforms have emerged, such as Baidu Maps, Google Maps, and Bing Maps. The data sources that can be aggregated include SuperMap RSET data service and WFS service. Although it can solve the existing data integration problems, it still cannot achieve a general data integration solution. Once application scenarios change or new requirements change, these integration methods often require a lot of changes. It cannot well meet the requirements of data integration [27].

Research status has become more and more mature, but it is rare to actually form a relatively mature and complete data integration platform. The main performance is as follows: it does not carry out systematic integrated development but only realizes some functions of data integration, and it does not carry out application-level integration [28]. For example, only a graphical application interface has been developed, which realizes simple data reading and processing, data format conversion, and data display. It is inconvenient to expand and reuse the system. The technical route it adopts is relatively old. Technology in the computer field is changing with each passing day, and new technology replaces old technology. Whether it is for user experience or for the improvement of system performance, system robustness is very necessary. There is no established system for the integration of IoT and UAV data [29].

## 4. Design and Test of Emergency Information Decision-Making System Based on Digital Water Network

**4.1. Design Process.** In this paper, the water emergency information decision-making system is designed according to the design principles of system practicability, expansibility,

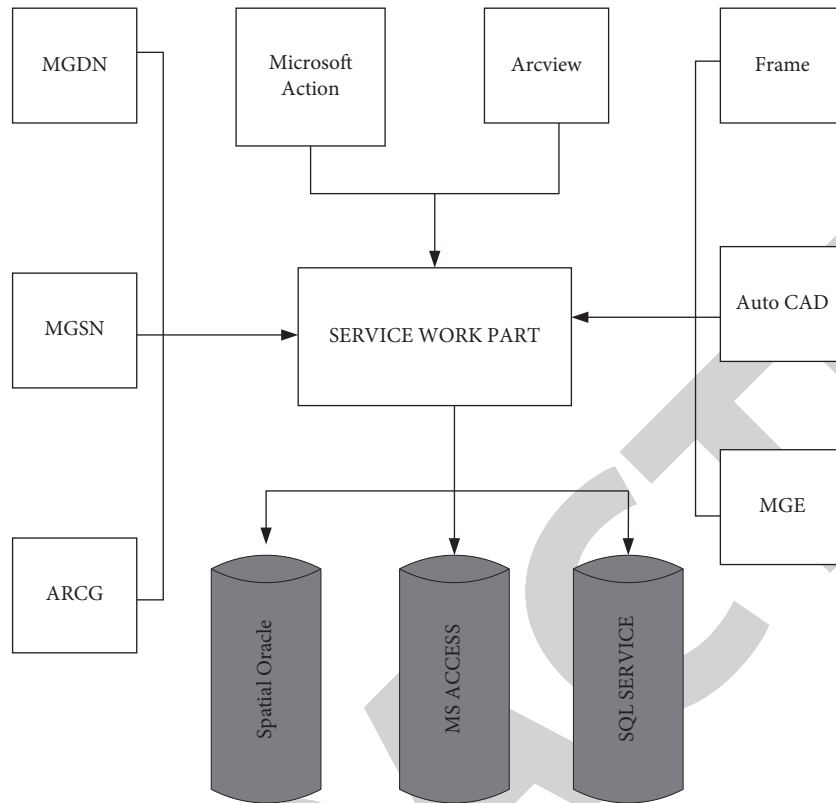


FIGURE 5: Multisource data adapter pattern.

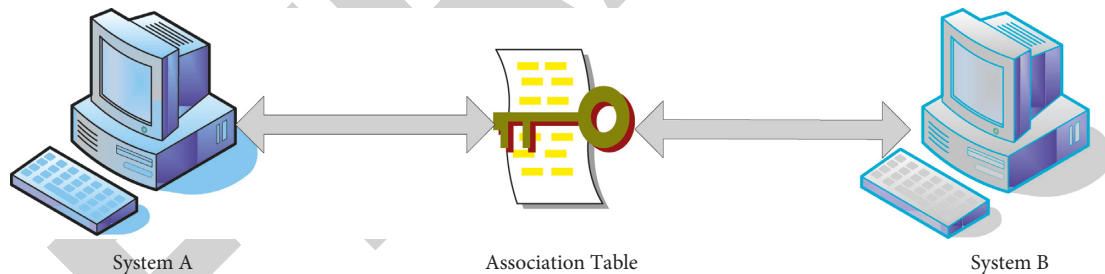


FIGURE 6: Mutual conversion between formats.

openness, system reliability, security, completeness, system standardization, and reasonable database structure. The flowchart of the research route of system design is shown in Figure 7.

The technical route of the system is shown in Figure 8.

The system designed in this paper mainly includes water conservancy WebGIS management module, comprehensive information management module, water level and rain monitoring module, and video data management module. The water conservancy WebGIS management module includes functions such as data production, data release, and data editing; the comprehensive information management module includes functions such as user login, authority control, and log management. The integrated module of water level and rain monitoring includes functions such as network connection, data capture, and data display. The video data management module includes functions such as

video browsing, video control, and video management, as shown in Figure 9.

Data integration method and establishment of database: the database is the background of the system and has a large storage function. It helps businesses store all the information they need. The most important thing in the establishment of the database is data integration. The operating speed of the system is closely related to the designed database. If the designed database is better, the system will run faster. If the designed database is not ideal, the system will run relatively slowly and slowly. The data types of data integration are shown in Table 2.

The database of the water conservancy information decision-making system is mainly composed of a login table, a sluice station information table, a river information table, and other tables. The login table includes user name, user password, and user type. The gate station information table



TABLE 1: Disadvantages of multisource heterogeneous data integration methods.

Conversion method	Shortcoming	Advantage
Format conversion mode	Loss of information; no direct access	Simple and convenient; suitable for conversion of any system data format
Adapter mode	Poor scalability. Expensive and bulky to write	No information loss, direct access, and universality
Interoperability mode	Difficult to achieve in a short period of time	Unified data access specification; strong scalability

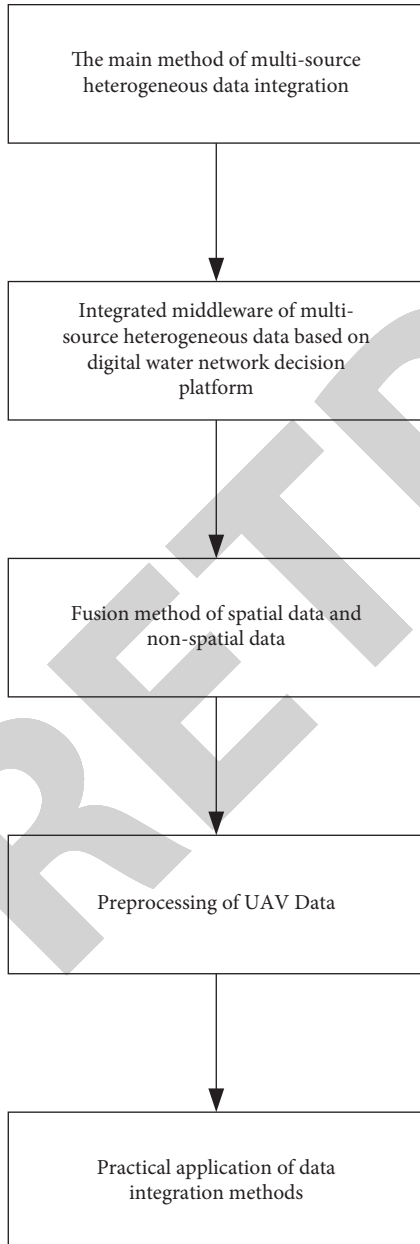


FIGURE 7: Flowchart of the research route.

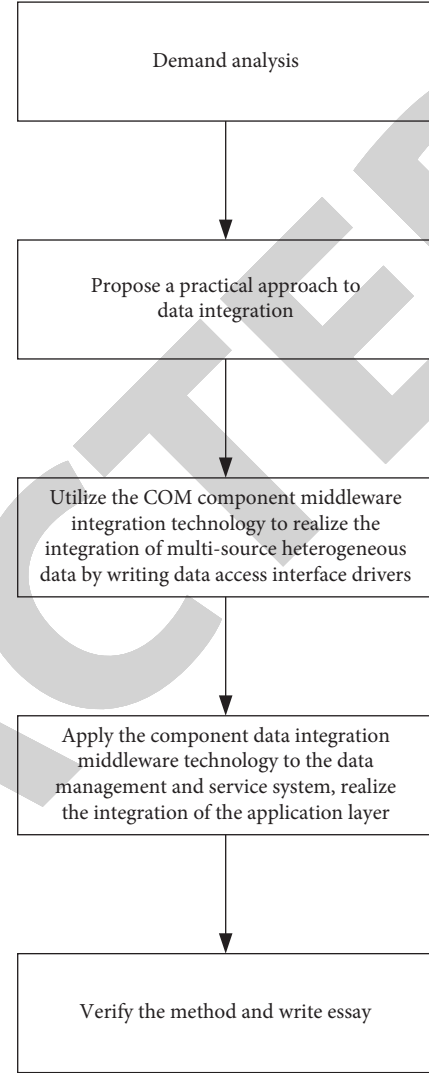


FIGURE 8: Flowchart of technical route.

stores the specific information of the gate station, including address, construction time, and name. The river information table mainly saves the specific information of the river, including the length, the flow area, and the constant water level, as shown in Tables 3–5.

4.2. *Feature Data Integration Algorithms.* A data can be regarded as a set of words, and its features can be represented by a vector space model (VSM); that is, a data  $d$  can be represented by its feature vector, which can be expressed as

$$V_{(\bar{d})} = (w_1 t_1, \dots, w_i t_i, \dots, w_n t_n). \quad (1)$$

Among them,  $\bar{d}$  represents the data,  $t_i$  represents the  $i$ -th feature data,  $w$  represents the weight of the feature data, and its weight can be expressed as

$$w_i(t, \bar{d}) = \frac{t f(t, \bar{d}) * \log(N/n + 0.01)}{\sqrt{\sum_{t, \bar{d}} [t f(t, \bar{d}) * \log(N/n + 0.01)]^2}} \quad (2)$$

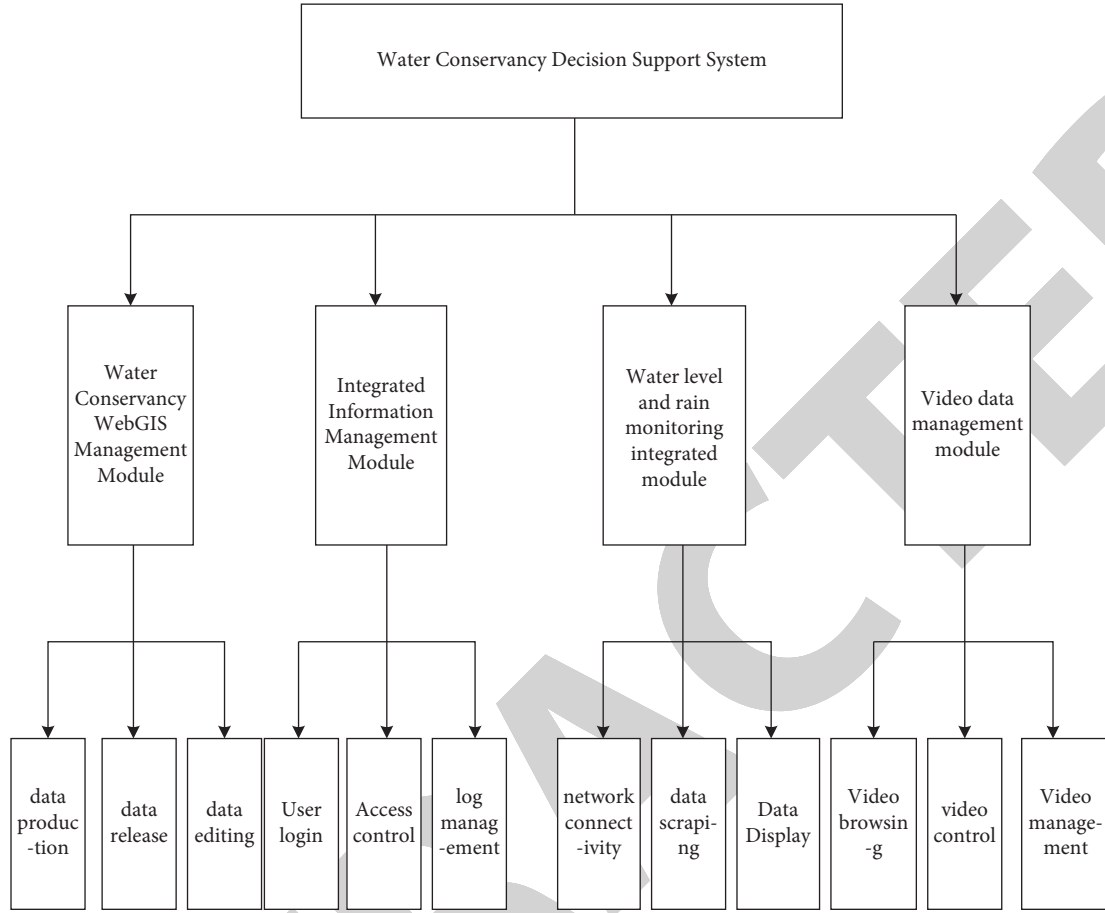


FIGURE 9: System frame design diagram.

TABLE 2: Data types required for system integration.

System management data	Product data	IoT monitoring data	Basic geographic data	Image data
Log management	Preprocessed images	Water condition monitoring	Administrative regions	Remote sensing image
User management	Surface parameters	Weather monitoring	Land use	Drone data
Monitoring point management	Water conservancy parameters	Raw weather data	Basic geographic data	Image data
Collection template management satellite sensor	Monitoring report	IoT monitoring data		

TABLE 3: System administrator table.

Field name	Type of data	Primary key	Use
ID	Autonumbering	Yes	User ID
Yhm	Text	No	Username
Mm	Text	No	Password
Lx	Text	No	User type

TABLE 4: Gate information table.

Field name	Type of data	Primary key	Use
ID	Autonumbering	Yes	Gate ID
Jd	Number	No	Longitude
wd	Number	No	Latitude
Zksl	Number	No	Number of gates
jcsj	Date/time	No	Build time
Gzll	Number	No	Gate flow
Name	Word	No	Sluice name

Among them,  $tf$  represents the word frequency of the data,  $N$  is the total number of data categories, and  $n$  is the number of data in one category.

As for a data sequence, if its horizontal eigenvector is  $d_i$  and its vertical eigenvector is  $d_j$ , then the similarity between the two is

$$\text{Sim}(d_i, d_j) = \frac{\sum_{k=1}^M w_{ik} * w_{jk}}{\sqrt{(\sum_{k=1}^M w_{ik}^2)(\sum_{k=1}^M w_{jk}^2)}} \quad (3)$$

TABLE 5: River information table.

Field name	Type of data	Primary key	Use
ID	Autonumbering	Yes	River ID
Length	Number	No	Length
Ljdy	Text	No	Flow through the region
Csw	Number	No	Normal water level

Among them,  $M$  is the dimension of the feature vector. The link weights for feature data can be expressed as

$$w_{(t, \overline{\text{link}})} = \frac{tf(t, \overline{\text{link}})}{\sqrt{\sum_{t \in \bar{d}} [tf(t, \overline{\text{link}})]^2}} \quad (4)$$

Among them, the denominator represents the normalization factor.

Let the similarity between the feature vector and  $w$  be  $S$  and the maximum value of  $S$  be  $\text{dis}$ , and then

$$\text{dis} = f\left(\max_{i=1-3}\{S_i\}\right). \quad (5)$$

The link score for feature data can be expressed as

$$\text{Score}(\overline{\text{link}}) = \lambda \text{dis} + \mu \text{Score}(\overline{\text{page}}). \quad (6)$$

Among them,  $\lambda$  and  $\mu$  are weight coefficients.

The weight scores for different data categories can be expressed as

$$D_i = \text{Cov}_{ij}. \quad (7)$$

Among them,  $\text{Cov}_{ij}$  represents the coverage ratio of this data category.

That is, for a random variable  $X$  with a normal distribution, its confidence rate is  $1 - \alpha$ , and

$$P = \left[ \bar{X}_n - \frac{S}{\sqrt{n}} t_{\alpha/2} < \mu < \bar{X}_n + \frac{S}{\sqrt{n}} t_{\alpha/2} \right] = 1 - \alpha. \quad (8)$$

For data subject to normal distribution, the regional acquisition rate is represented by  $h$ , the coverage ratio is  $c$ , and then the overall acquisition rate  $H$  and coverage ratio  $C$  are

$$\begin{aligned} H &= \frac{n * m * c}{n * m * c / h} = h, \\ C &= \frac{n * m * c}{n * m} = c. \end{aligned} \quad (9)$$

Among them,  $m$  represents the number of data serial ports.

However, for practical statistical situations, it is usually

$$H = \frac{n' * m * c}{n * m * c / h} \leq h, \quad (10)$$

$$C = \frac{n' * m * c}{n * m} \leq c.$$

Let  $C_i$  and  $H_i$  be the acquisition rate and coverage ratio after the  $i$ -th data learning and  $E$  be the expectation, and then the objective function can be expressed as

$$f_1 = \frac{1}{n} \sqrt{\sum_i (H_i - E(H))^2}, \quad (11)$$

$$f_2 = \frac{1}{n} \sqrt{\sum_i (C_i - E(C))^2}.$$

According to the matching mode of the data serial port, the negative correlation measurement method can be used to obtain

$$m_n(A_p, A_q) = \sqrt{P(A|\neg B) * P(B|\neg A)} = \sqrt{\frac{f_{01}}{f_{0+}} * \frac{f_{10}}{f_{+0}}} \quad (12)$$

Among them,  $A_p$  represents the attribute of the A dataset,  $A_q$  represents the attribute of the dataset B,  $\neg B$  is the inverse attribute of the B dataset, and  $\neg A$  is the inverse attribute of the A dataset.  $f$  is the number of times the data attribute appears in the serial port, that is, the target value;  $P$  represents the probability.

And it can be calculated according to the data series of the previous overall coverage ratio  $C$  as follows:

$$m_n(A_p, A_q) = \sqrt{\frac{C[j][j] - C[i][j]}{n - C[j][i]} * \frac{C[i][i] - C[i][j]}{n - C[j][j]}}. \quad (13)$$

Among them, the larger  $m_n$ , the stronger the negative correlation.

In addition,  $A_i$  and  $A_j$  represent data attributes, and then the identification mode of the data is

$$\Theta_{A_i} = \{A_i, = A_j A_i, = \text{null}\}. \quad (14)$$

Let the number of layers of a data node Node from the root node root be recorded as  $\text{Depth}(\text{Node})$  and the number of child nodes of Node node be recorded as  $\text{Width}(\text{Node})$ , and the weight that should be assigned to each edge in the decision tree is

$$\begin{aligned} \text{Weight}(\text{Node}) &= k_1 * \frac{\alpha^{\text{depth}(\text{Node})}}{\text{width}(\text{Node})} \\ &+ k_2 * \frac{1}{\text{depth}(\text{Node}) + 1}. \end{aligned} \quad (15)$$

Among them,  $k_1$  and  $k_2$  are constants,  $k_1 + k_2 = 1$ ,  $\alpha$  are controllable parameters, and

$$\alpha = \min\left(\frac{\text{width}(\text{Node})}{\text{Weight}(\text{Parent}(\text{Node}))}\right). \quad (16)$$

For the final semantic similarity expressed by Sim, we can get

$$\text{Sim}(A_p, A_q) = \frac{1}{[\text{dis}(A_p, A_q) + 1]}. \quad (17)$$

**4.3. Simulation Experiment.** This paper tests the performance of the designed digital water network emergency information decision-making system and tests its early warning capability and system energy consumption, respectively. The test results are shown in Figure 10.

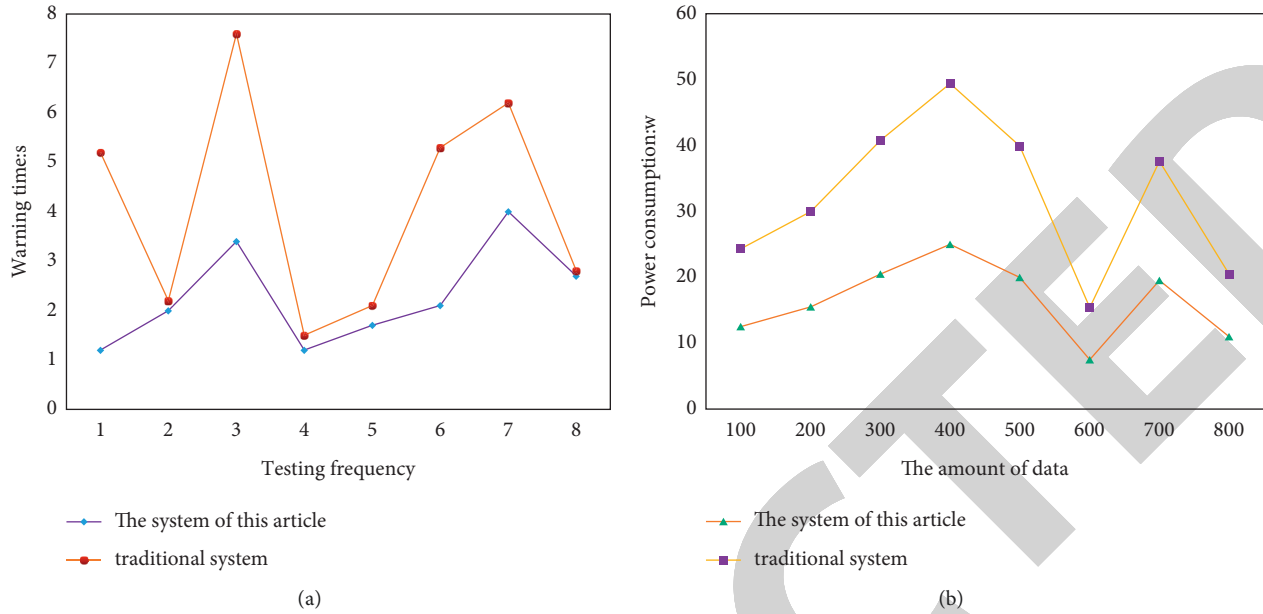


FIGURE 10: System performance test results. (a) Comparison of emergency early warning capabilities. (b) Comparison of system power consumption.

As can be seen from Figure 10, the early warning time of the system designed in this paper is within 1 to 4 seconds, while the early warning time of the traditional system is within 1.5 seconds to 8 seconds. The early warning time of the system designed in this paper is significantly faster, and the maximum power consumption of the system designed in this paper does not exceed 25 W, while the traditional system power consumption is up to 50W. The power consumption of the system in this paper is about half of the traditional system, and the power consumption is significantly lower. Moreover, the curve is more stable, and the upper and lower amplitudes are significantly lower, so it can be concluded that the system designed in this paper has a better stability.

## 5. Conclusion

This paper firstly summarizes the overall content of the full text in the abstract. Secondly, the era background of big data is introduced in the introduction. This paper introduces the concept of decision space information system and data integration. This paper summarizes the innovations of this paper. In the related work part, this paper exemplifies some related researches in order to understand the current situation of the related content researched in this paper. Then in the theoretical research part, the paper first introduces the decision space information system, including its characteristics and related content. Secondly, this paper introduces the relevant content of data integration methods, including its development status and integration methods. At the end of this paper, the relevant calculation methods and the design of the system are explained in the experimental part, and the early warning ability of the system and the energy consumption of the system are tested. The results show that, compared with the traditional system, the early warning

capability of the system is significantly faster, the system power consumption is lower, and the system is more stable.

## Data Availability

This paper does not cover data research. No data were used to support this study.

## Conflicts of Interest

The author declares no conflicts of interest.

## Acknowledgments

This study was supported by the Shaanxi Water Conservancy Science and Technology Project, Research on Socialized Water Conservation Knowledge Service and Personalized Water Conservation Behavior Regulation, 2019slkj-13 (Water Conservancy Science and Technology Project); Shaanxi Provincial Natural Science Basic Research Program (Key Project), Research on Modeling and Control Technology of Huangchigou Water Distribution Hub Phase of Hanjiang to Weihe River II Project, 2019JLZ-16 (Provincial Fund); and Shaanxi Water Conservancy Science and Technology Project, Research on the Connection, Control and Joint Commissioning Mode of Rivers, Reservoirs and Canals Based on Digital Water Network, 2020slkj-16 (Water Conservancy Science and Technology Project)

## References

- [1] Y. Shibuo, H. Sanuki, S. Lee, K. Yoshimura, and S. Sato, "Advanced countermeasures for urban inundation: an example case of data integration approach to solve social issues,"

- Journal of Information Processing and Management*, vol. 60, no. 2, pp. 100–109, 2017.
- [2] M. Khovrichiev, L. Elkhovskaya, V. Fonin, and M. Balakhontceva, “Intelligent approach for heterogeneous data integration: information processes analysis engine in clinical remote monitoring systems - ScienceDirect,” *Procedia Computer Science*, vol. 156, pp. 134–141, 2019.
  - [3] S. Park, J. K. Kim, and D. Stukel, “A measurement error model approach to survey data integration: combining information from two surveys,” *Metron*, vol. 75, no. 3, pp. 345–357, 2017.
  - [4] M. Masmoudi, S. B. A. B. Lamine, H. B. Zghal, B. Archimede, and M. H. Karray, “Knowledge hypergraph-based approach for data integration and querying: application to Earth Observation,” *Future Generation Computer Systems*, vol. 115, no. 1, pp. 720–740, 2021.
  - [5] D. A. Mekala, “An ontology approach to data integration using mapping method,” *International Journal for Modern Trends in Science and Technology*, vol. 6, no. 12, pp. 28–32, 2020.
  - [6] C. Daraio, S. Di Leo, and M. Scannapieco, “Accounting for quality in data integration systems: a completeness-aware integration approach,” *Scientometrics*, vol. 127, no. 3, pp. 1465–1490, 2022.
  - [7] L. Dalla Valle and R. Kenett, “Social media big data integration: a new approach based on calibration,” *Expert Systems with Applications*, vol. 111, pp. 76–90, 2018.
  - [8] Q. Li, H. Tjong, X. Li et al., “The three-dimensional genome organization of *Drosophila melanogaster* through data integration,” *Genome Biology*, vol. 18, no. 1, p. 145, 2017.
  - [9] N. M. Raj, D. R. Jegadeesan, and S. Teja, “A survey ON an ontology development ON data integration,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 6, no. 3, pp. 595–600, 2019.
  - [10] A. C. Kakouri, C. C. Christodoulou, M. Zachariou et al., “Revealing clusters of connected pathways through multi-source data integration in huntington’s disease and spastic ataxia,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 26–37, 2019.
  - [11] P. A. Zadeh, L. Wei, and A. Dee, “BIM-CITYGML data integration for modern urban challenges,” *Electronic Journal of Information Technology in Construction*, vol. 24, no. 17, pp. 318–340, 2019.
  - [12] A. K. Maity, R. J. Carroll, and B. K. Mallick, “Integration of survival and binary data for variable selection and prediction: a Bayesian approach,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 68, no. 5, pp. 1577–1595, 2019.
  - [13] M. Wiedau, L. Von Wedel, H. Temmen, and R. N. Welke, “ENPRO data integration: extending DEXPI towards the asset lifecycle,” *Chemie Ingenieur Technik*, vol. 91, no. 3, pp. 240–255, 2019.
  - [14] F. Economou, Y. Panagopoulos, and E. Tsouma, “Uncovering asymmetries in the relationship between fear and the stock market using a hidden co-integration approach,” *Research in International Business and Finance*, vol. 44, pp. 459–470, 2018.
  - [15] F. Azzalini, S. Jin, M. Renzi, and L. Tanca, “Blocking techniques for entity linkage: a semantics-based approach,” *Data Science and Engineering*, vol. 6, no. 1, pp. 20–38, 2021.
  - [16] N. Fikri, M. Rida, N. Abghour, and K. A. Moussaid, “An adaptive and real-time based architecture for financial data integration,” *Journal of Big Data*, vol. 6, no. 1, p. 97, 2019.
  - [17] S. A. Giroux, I. Kouper, L. D. Estes et al., “A high-frequency mobile phone data collection approach for research in social-environmental systems: applications in climate variability and food security in sub-Saharan Africa,” *Environmental Modelling & Software*, vol. 119, pp. 57–69, 2019.
  - [18] M. Kettouch, C. Luca, and M. Hobbs, “SemiLD: mediator-based framework for keyword search over semi-structured and linked data,” *Journal of Intelligent Information Systems*, vol. 52, no. 2, pp. 311–335, 2018.
  - [19] O. Hajoui, R. Dehbi, and M. Talea, “An approach for big data interoperability,” *Journal of Engineering and Applied Sciences*, vol. 13, no. 17, pp. 7323–7328, 2018.
  - [20] A.-S. Thelisson, A. Missonier, G. Guieu, and L. S. Luscher, “A paradoxical approach symbiotic to postmerger integration: a French longitudinal case study,” *European Business Review*, vol. 31, no. 2, pp. 232–259, 2019.
  - [21] M. V. García, A. Armentia, F. Pérez, and E. M. Estévez, “Vertical integration approach for the intelligent Oil & Gas field,” *At - Automatisierungstechnik*, vol. 66, no. 10, pp. 859–874, 2018.
  - [22] M. Bhattacharya, J. N. Inekwe, and M. R. Valenzuela, “Financial integration in Africa: new evidence using network approach,” *Economic Modelling*, vol. 72, pp. 379–390, 2018.
  - [23] X. Zhang, H. Chen, Y. Ruan, and D. C. Pan, “Matviz: a semantic query and visualization approach for metallic materials data,” *International Journal of Web Information Systems*, vol. 13, no. 3, pp. 260–280, 2017.
  - [24] H. Chen, D. L. Fan, L. Fang et al., “Particle swarm optimization algorithm with mutation operator for particle filter noise reduction in mechanical fault diagnosis,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 10, 2020.
  - [25] H. Chen, Y. Lu, and L. Tu, “Fault identification of gearbox degradation with optimized wavelet neural network,” *Shock and Vibration*, vol. 20, no. 2, pp. 247–262, 2013.
  - [26] Y. Zeng, G. Chen, K. Li, Y. Zhou, X. Zhou, and K. Li, “M-skyline: taking sunk cost and alternative recommendation in consideration for skyline query on uncertain data,” *Knowledge-Based Systems*, vol. 163, pp. 204–213, 2019.
  - [27] S. Rajendran, O. I. Khalaf, Y. Alotaibi, and S. Alghamdi, “MapReduce-based big data classification model using feature subset selection and hyperparameter tuned deep belief network,” *Scientific Reports*, vol. 11, no. 1, Article ID 24138, 2021.
  - [28] M. Rajalakshmi, V. Saravanan, V. Arunprasad, C. A. T. Romero, and O. I. Khalafi Khalaf, “Machine learning for modeling and control of industrial clarifier process,” *Intelligent Automation & Soft Computing*, vol. 32, no. 1, pp. 339–359, 2022.
  - [29] O. I. Khalaf, G. M. Abdulsahib, and B. M. Sabbar, “Optimization of wireless sensor network coverage using the bee algorithm,” *Journal of Information Science and Engineering*, vol. 36, no. 2, pp. 377–386, 2020.