Construction of Social Governance Data Integration Model and Decision Analysis Method in Big Data Era

Yuchun Li [1,2 a]

[1] Department of Public Policy and Management, Shih Hsin University, Muzha Road, Taipei, China
[2] Department of Public Administration, Dongguan City College, Wenchang Road, Dongguan, China

[a]e-mail: liyuchun@dgcu.edu.cn

Abstract. In order to solve the problems of complex distributed data integration and weak data analysis visibility in social governance, this paper proposes a distributed data integration and visualization application method. Based on the big data processing mode, the method accesses, extracts and integrates the database data scattered in different network routes, conducts mining analysis, enhances the ability of data dynamic description and Web visualization, and provides service-oriented intelligent social governance decision analysis and application. On this basis, a prototype system is designed and developed and tested. The experimental results show that the method improves the real-time, security, accuracy and visualization of data analysis of data integration, and is more simplified than the traditional mode.

Keywords: social governance; decision analysis; big data;

1 Introduction

The social governance system is a complex system based on big data, involving political and legal affairs, comprehensive governance, stability maintenance, letters and visits, public security, justice, urban management, civil affairs, human resources, family planning, safety supervision, emergency response and other functional departments as well as municipal, district, and street departments. It is a management service resource of multi-level system of office (township) and community (village), which has the characteristics of large amount of data, many levels of variables, large differences in unit composition and high dynamics. Traditional relational databases lose performance, functions and performance when storing large data sets. Cost advantages, and even more powerless when processing and querying large data sets, it is urgent to design and optimize data storage, management and query platforms for big data.

Presently, the data integration of multi-service application systems is mainly based on the two-way conversion model between relational database and

© The Author(s) 2022
H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 252–263, 2022.
https://doi.org/10.2991/978-94-6463-108-1_29

extensible markup language (XML), which lacks the characteristics and attributes of big data. Based on demand analysis, it combines with the social governance business process, integrate the concept of big data, build a data integration and decision

analysis platform that connects top and bottom, horizontal collaboration, fast response, and strong

support. Under the condition of maintaining data consistency and good scalability, we integrate heterogeneous data distributed in different systems, and transparently access the data resources of the original system for comprehensive application of decision analysis.

2 Design principles and integration strategies

The large amount of structured and unstructured data involved in social governance data is characterized by large data volume, large type and value potential.

Compress, sample, and mine the basic data of "people, places, objects, situations, affairs, and organizations" scattered in different fields, we build a social governance spatial database, public database, business database and population database, and establish normalized updated data synchronization mechanism, which based on the B/S structure collaborative work application environment. It realizes the information interconnection management of the integration of graphs, texts, tables and workflows, and provides a big data analysis and management platform with resource sharing and rapid response, which is compatible with traditional data warehouse applications. In comparison, it has the characteristics of large amount of data, complex query and analysis, etc. It uses metadata, database, data warehouse, geographic information system (GIS), network, Web, etc. to integrate and integrate social governance technology framework [1]. To meet the different needs of data, the platform considers the functional combination, data combination and data scope of different user interfaces in an all-round way, and can ensure the independent operation of each system when calling business systems, using data and accessing video, without affecting each other.

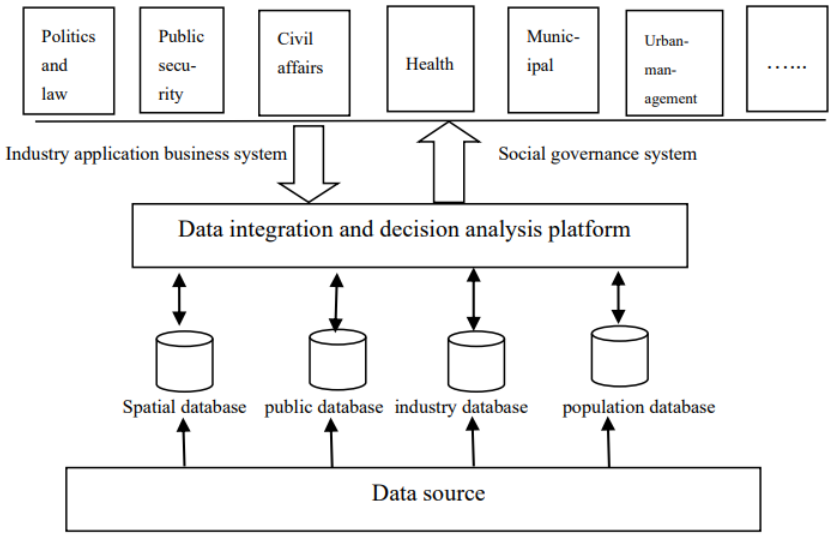

Fig. 1. Design model structure

The platform is built on the decentralized, heterogeneous and professional application systems of different departments, integrates static data, real-time monitoring data and multimedia streaming data, and provides real-time monitoring data for social governance on the basis of ensuring the independent operation of each business system, video resource access, display services and other intelligent decision analysis support means. Figure 1 shows the design model structure, it can be seen that the platform integrates a variety of data source data, and extracts to form spatial database, public database, industry database and population data. Spatial data includes data such as electronic maps, remote sensing images, and virtual three-dimensional data. Public data includes urban objects, urban facilities, landmark locations, management organizations and other data. Business data includes management event data, mainly including system operation to generate business-related data, involving letters and visits, floating population, public safety, corrections, assistance and education, security prevention, comprehensive law enforcement and other events, as well as the distribution of events defined by the process, reported cases, evaluation indicators and other data. Population data includes actual population, monitoring population, service population, etc. The data of key populations and aided populations are unified and integrated, and a dynamic population information resource database is established to realize cross-department, cross-level and cross-platform population data information sharing and sharing.

3 Technical Architecture

Based on big data, we have comprehensive use of geographic information, human-computer interaction, data mining, information retrieval, data visualization and other

information technologies, and using the integration method of data extraction, while integrating multi-database heterogeneous data. It does not affect the independence of each special application system Operation, not only ensures the instant demand of professional applications for data access, but also ensures the instant update and unification of each data version, and always grasps the latest status of social governance from a global perspective, and can rationally make scientific judgments on urban development and adjustment [2]. Figure 2 shows the overall technical architecture, including data extraction from special business application system database groups, data storage and management, accelerators, and big data analysis applications.

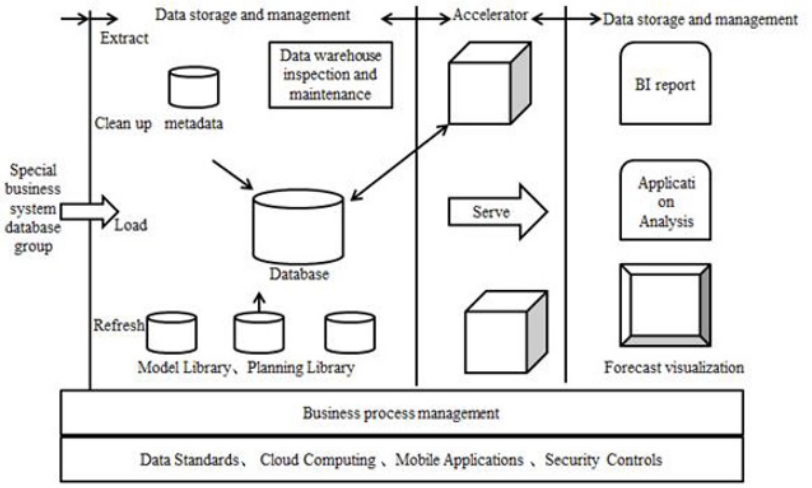

Fig. 2. Platform technical architecture

3.1 Access and acquisition of different data sources

Data source is the basis of data integration. Extract, clean and load data from various types of data sources in each special business system group, and load the data into the platform according to certain rules and standards. The platform access part is composed of a series of services and interfaces which are formed. Each service and interface corresponds to a specific data source. The data source provider needs to fill in the connection information of the data source on the page and register with the service center. Only the registered data source can become valid data Source. Heterogeneous database access refers to connecting data sources of heterogeneous database systems through interfaces and classes, and accessing metadata of data sources. Web data source access refers to reading data from XML documents and inserting them into the platform database. When entering a Web data resource, the system queries the metadata database to store the style sheet field registered with the Web resource, reads the predefined schema in the file, and forms a Web data file data resource description file.

3.2 Data extraction, loading, transformation and output services

Data extraction, loading, transformation and output services are the main contents. It is necessary to establish a theme-oriented, integrated, relatively stable data integration management platform that reflects historical changes to support management decision-making and realize the operation and operation of special social governance businesses data management such as data storage, query and extraction. Integrated data management is used to support decision-making and is oriented to analytical data processing. It is different from the existing operational databases of business units. It is an effective integration of multiple heterogeneous data sources. Afterwards, it is reorganized according to the theme, and includes historical data, and the data stored in the data integration platform is generally not modified. Figure 3 is the processing structure diagram of data integration, including data acquisition, storage transformation, output service and application service.

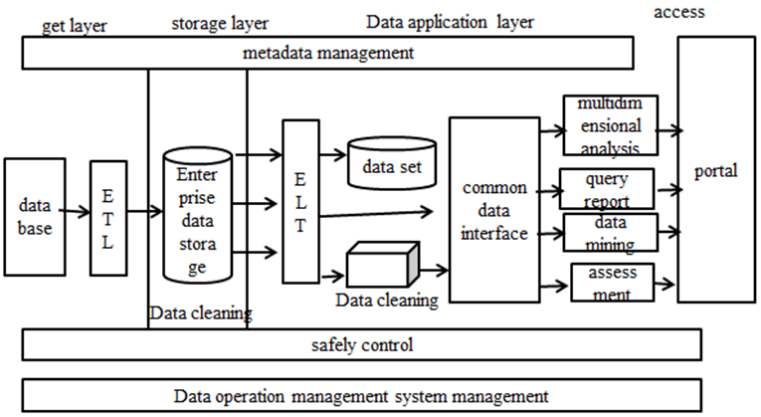

Fig. 3. Data integration processing structure

Data extraction is to import data from online transaction processing systems, external data sources and offline data storage media to the data integration platform, mainly involving interconnection, replication, increment, transformation, scheduling and monitoring. Storage and management is the key to data integration. It extracts, cleans and integrates data effectively, reorganizes it according to the theme, finally determines the physical storage structure of the data, and organizes the storage of metadata. The organization and management of data is different from traditional databases, including data security, archiving, backup, maintenance, recovery, etc.

The performance of data focuses on multi-dimensional analysis, mathematical statistics and data mining. Through the interoperability of data with different architectures, in-depth social governance business, multi-dimensional data analysis, flexible combination and mining of information, it provides online analysis capabilities and realizes decision-making results: diversified outputs which form a complete smart social governance ecosystem.

The compression algorithm of data compression plays a vital role in the real-time

optimization of the system. The incremental swing door trending (SDT) uses the structured query language (SQL) database to store the compressed data, and uses the LZW (lempel-ziv-welch encoding) compression algorithm performs secondary lossless compression, which improves the compression efficiency of storage. The compression module in this paper adopts the compression algorithm and strategy.

3.3 Rule-driven decision analysis

The platform integrates data from different thematic databases through certain rules, conducts intelligent in-depth mining and quantitative analysis of global social governance data, and displays it on a unified operating platform and provides special decision-making analysis services for industry management. The XML rule-based query language (XML -based query language, XML- RL) is extended to define the rules of data conversion in the system, and the XML data generated by the multi- source data integration module is passed through the rule engine to generate data that conforms to the rules. The specified operations are processed accordingly. The platform integrates report management, indicator customization, model construction, multi-dimensional display and GIS graphic display, etc. It provides a comprehensive and multi-angle analysis of the development status and trend of social governance, improves the ability to respond quickly to events and planning capabilities, including social governance monitoring and early warning, emergency command management, auxiliary decision support, etc. Monitoring and early warning is through multi-scale mining of correlated data to discover hidden, potential, and frequent correlations and patterns behind the data [3]. It realizes the analysis of the data of social governance signs and indicators, performs sub-item, regional and comprehensive display, real-time and dynamically grasps the status of urban social governance, makes statistics on the operation of social governance, predicts the operation of a period of time in the future, and realizes the overall monitoring of social governance.

Emergency command and management is a method of analyzing, planning, organizing, coordinating and managing public emergencies, providing information services, resource scheduling, order implementation, deployment and supervision, mobilizing resources in a timely and effective manner, responding to emergencies, and mitigating hazards. Decision analysis support is to provide decision makers with human-computer interaction methods through data, models and knowledge to make semi-structured or unstructured decisions. The platform is based on the accuracy and functional diversity of information retrieval, and integrates intelligent human-machine interface. Channel interaction mode and web page, audio, video and other carrier interaction display forms, input or output information in a way that is closer to the way people perceive the external world, improve the ability to choose and control information expression forms, and improve the logic of information expression forms and people. The degree of combination with creative ability expands human information processing ability in sequence, symbolic information and parallel and associative information, making decision analysis support more humanized.

Social governance integrated data has the characteristics of big data: the massiveness reaches the level of PB data; the diversity of types covers semi-structured and

unstructured data information such as form data, videos, images and location coordinates; the rapid data flow makes decision analysis need to be fast, continuous real-time processing, and integration of artificial intelligence interaction; low-level data mining needs to mine small-scale value information from massive data [4]. RES (rule execution server) service is an effective way to build large-scale scalable systems. It provides constraints on the technical architecture, which can help achieve loose coupling, maintainability, evolvability and scalability [5], and build a social governance that integrates data management, graph management, multi-dimensional display, and network publishing. Interactive platform, is providing multi-professional, multi-level, multi-objective comprehensive decision analysis services and virtual experience.

4 Data Integration Model

At present, there are many technologies that can realize the seamless integration of multi- source data, including data middleware technology, SMS (Short messaging service) technology, data standardization mode, etc. initial integration.

4.1 Establish a data processing model

Due to the complexity of the social governance business, it is impossible to directly establish a data model from the real social governance business. Instead, the scattered business data is abstracted into an information model, and a conceptual model of the social governance information business data is established, and the business data conceptual model is further developed. It is transformed into a business data model that can be implemented in a computer and ultimately supports the database system. Figure 4 shows the construction structure of the data processing model.

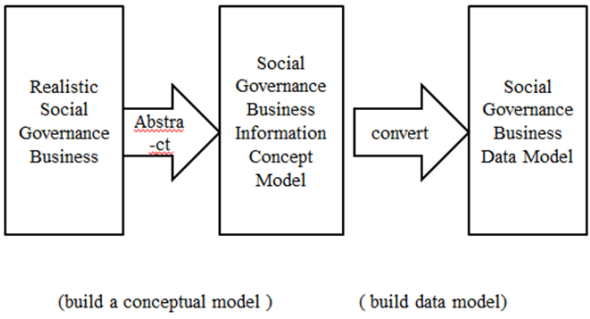

Fig. 4. Data processing model construction structure

In the model preparation stage, we firstly clarify the modeling purpose, collect the necessary information, and clarify the characteristics of the object; model assumptions refer to the necessary and reasonable simplification of the problem according to the characteristics of the object and the purpose of modeling, and make it in precise lan-

guage. Hypothesisly, the problem should be linearized and homogenized as much as possible; in the model construction stage, the internal laws of the object and appropriate mathematical tools are used to construct the equation relationship or other mathematical structure between various quantities; prove theorems, logical operations, numerical operations and modern computer mathematical methods; in the model analysis stage, mathematical analysis, error analysis, data stability analysis, are carried out on the model.

4.2 Multi- source data composite model based on document type definition

The use of a unified model will facilitate the unified processing of data and generate data with consistent models, so that the subsequent data conversion modules can use the same method to process these complex data sources and reduce the complexity of parsing data in different formats.

The emergence of XML technical standards enables the unified description of the data of each subsystem. The data model after the integration of multi-source data is stored in the form of XML, including three dimensions of time, geography and industry, among which geographic information and thematic data are all composite objects, which are defined by the actual business system, and will not be repeated here. The XML document type definition (document type definition, DTD) is defined as follows:

<?xml version=”””’ encoding=”UTF-8”?>
<! ELEMENT time (#PCDATA) >
<! ELEMENT areainfo (#PCDATA) >
<! ELEMENT business Info (#PCDATA) >
<! ELEMENT date (time, areainfo, business Info) >

5 Decision analysis method based on XML-RML

After the transformation of the unified model, the system generates a large amount of XML data including geographic information and thematic data composite information. These data cannot be directly displayed as final data, nor have the ability to be used for decision analysis. In order to solve this problem, in this paper, the rule-based XML manipulation language (extensible markup language rule-based manipulation language, XML-RML) extends a manipulative rule language.

5.1 XML-RML

XML-RM regards XML document as a complex object data model, and is a rule-based XML query language [6]. The query statement of XML-RL language consists of two parts: query clause and construction clause. The query clause is a rule-based path expression that is used to extract data from XML documents. The construct clause is used to construct query results and is defined as follows:

Querying qexp1, ….., qexpn.

Constructing cexp.

5.2 XML-RML Rule Language Definition

Based on the idea of rules, the XML-RML language is extended and named as the rule-based XML manipulation language.

Combined with actual requirements, a rule may produce multiple results, and different results correspond to different operations. Therefore, the structure of the query statement is extended. The set of operations after a predefined set is included in the result set in its construction clause. Add corresponding operations. Operations are a set of sets, corresponding to a series of different interfaces provided by the system. The expanded structure is described as follows:

Operation: opl, op2, …, opm

Querying qexpl, …, qexpn

Constructing [cexpl, opl], [cexp2, op2], ..., [cexpk, opk]

The operation makes the data rule conversion module have good flexibility and expansibility, thereby saving costs and reducing the difficulty of application development. In the case of an increase in the resource business platform, as long as the corresponding rules and interfaces are added accordingly, it can be realized.

5.3 XML-RML Implementation

The implementation of XML -RML is based on the XML-RML query language, obtains the result of each constructive clause expression, and calls different interfaces according to different operations, and finally completes the corresponding operations. The algorithm implementation is shown in Figure 5.

```
Algorithm do Work (Data, Σ)
Input: DTD - compliant XML data set data XML-RML
rule set Σ
Goal: Form the collected data into business data accord-
ing to the rules and perform related operations.
Method:
For reach rule r in Σ {
// Record the query clause of each rule
Q = r. querying
For each constructing expression cexpi inr
{
// Construct an XML-RL query statement using the
XML-RML query clause and each result expression of
the construct clause
rml = new XML-RL (Q, cexp);
// Query the input and record the result
        result = do Query (data, rml);
// According to the definition of XML-RML language,
perform related operations on the results
Do Operation (result, opi);
    }
} .
```

Fig. 5. Algorithm implementation

5.4 Efficiency Analysis

Assuming that the size of the XML-RML rule set Σ is m, the size of the operation set is n, and the time complexity of the XML-RL query language is O(X), then the time complexity of the above algorithm is O. The complexity of the XML-RML algorithm depends on the time complexity of the XML-RL query language, so it can be seen that the complexity of the algorithm is at the polynomial level.

6 Analysis of test results

On this basis, a prototype system is designed and developed, and the performance test is carried out on the system. The test environment consists of a central platform server IBM System p6 550 Q and 10 database servers. IBM System p5520, and 10 database servers are distributed in network nodes. Among them, one of them is a geographic information database, one is a real-time database of the central platform, and the remaining 8 are resource-specific business databases. The operating system is Linux, and the database software is Oracle. Figure 6 is the test result of analyzing the average response time, from which it can be seen that, the system can fully support the needs of social governance in data integration and decision analysis.

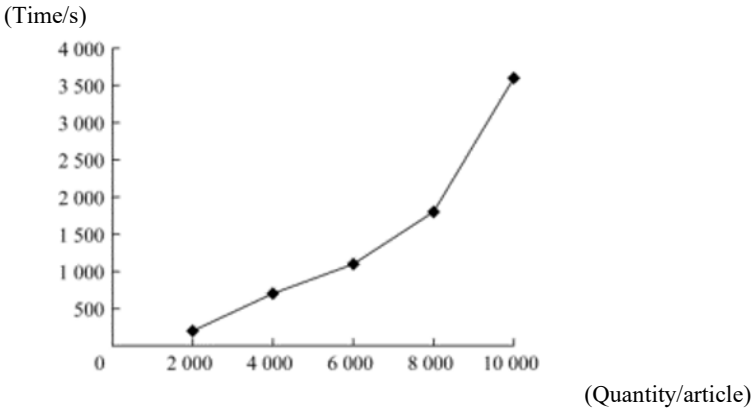

Fig. 6. Average response time test results

The test content is tested with the collected data as a sample. The data from the geographic information system and the resource-specific business system has an average data size of 5MB. In the system, the defined rules should not be too many, otherwise the business processing efficiency will be too low. The number is defined as 2000, the average response time of the statistical processing of the system in the case of different amounts of data, the unit is ms.

7 Conclusion

Based on the current situation of high complexity of social governance distributed structure data integration and weak data analysis visibility, this paper proposes XML-RML language and data integration visualization application model, which provides a standard reference for multi- source data composite, and conducts a trial, and solicited the opinions of experts in the field. The results show that the use of this system to assist in social governance can improve the integration efficiency of distributed data, provide an effective means to support social governance decision-making analysis, and solve the problem of multi-system social governance. They solve the problem of collaborative work. The system has good scalability, thereby saving costs and reducing the difficulty of application development.

Acknowledgments.

Yuchun Li (1984 -), male, from Fuyang, Anhui Province, lecturer of Dongguan City College, doctoral candidate in the Department of Public Policy and Management of Shih hsin University, research direction: social governance and public policy.

References

1. SONG Yue, ZUO Qunchao, NIU Haibo, National geological database consolidation and integration of the basic technical framework [J]. China Mining Mageazine, 2016, 25 (6): 154-157
2. CHEN Ruming. Challenge, value and coping strategie of big data era [J]. Mobile Communications, 2012, 36 (17), 14-15
3. LIU Mengmeng, ZHAO Shuliang, CHEN Min, et al. Scaling-up mining algorithm of multi-scale association rules mining [J]. Application Research of Computers, 2015, 32 (10.): 24-29
4. SUN Zuhan, OKApi Visualization RESTful Service Composition Modeling and Implementation [D]. Hangzhou: Zhejiang University, 2016
5. He Yanrun. Wisdom city information architecture research of data centric [J]. Mobile Communications, 2013, 37 (21): 13-17
6. Liu M, Ling T W. Towards declarative XML querying [C]// Proc 3rdInternational Conference on Web Information Systems Engineering. Singapore: IEEE Press, 2020: 127-136

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.