Innovation Research in City Public Management Based on a Data Resource Sharing Exchange Platform

Huaihai Hui
School of Economics and Management
Chinese Academy of Sciences
Beijing, P.R.C.
huihuaihai@ucas.ac.cn

Des McLernon
School of Electronic and Electrical Engineering
University of Leeds
Leeds, LS2 9JT, United Kingdom
d.c.mclernon@leeds.ac.uk

Ali Zaidi
School of Electronic and Electrical Engineering
University of Leeds
Leeds, LS2 9JT, United Kingdom
s.a.zaidi@leeds.ac.uk

ABSTRACT

In order to solve the common needs of shared basic data resources in different industries, different departments and different regions in a city, this paper has designed a Data Resource Sharing and Exchange Platform (DRSEP) to solve the demand for comprehensive data of City Public Management (CPM). Firstly, we analyze the basic functional requirements of the platform. Secondly, we study the characteristics of the DRSEP from three aspects: data type, data volume and data transmission/application methods. Thirdly, we provide the overall technical architecture of the DRSEP, which includes: the data resource layer, the resource site layer, the resource integration layer, the resource service layer and the resource application layer. In particular, we devise a Data Organization System (DOS) based on a Data Resource Directory (DRD), a Public Basic Database (PBcDB) and a Public Business Database (PBuDB). The DRD is the metadata standard and resource directory of the platform. Through this directory, the authority, accuracy and unity of the data resources can be ensured. The PBcDB consists of four types of databases: a population database, a corporation database, a macroeconomic database and a geospatial database. The PBuDB includes a video database, an environmental safety supervision database, a building database, a government affairs and emergency database, a credit database and a comprehensive human resources database. Finally, the DRSEP is deployed, tested, applied and evaluated in a city with a population of 2.19 million in western China. Moreover, the test evaluation results show that the platform exhibits an outstanding performance in the integration technology of multi-source heterogeneous data and the reliable transmission technology of massive data.

CCS Concepts
• **Information systems → Data management systems → Information integration → Extraction, transformation and loading**

Keywords
Platform, Data Resource, Management, Innovation.

1. INTRODUCTION

City public data resources are an important foundation for the construction of smart cities and an indispensable information infrastructure [1]. However, due to the limitations of many factors such as management mode and technical means, it has been very difficult for data resources to be shared among different departments and regions. The phenomenon of "Information Silo" [2] is serious. Therefore, this paper designed a Data Resource Sharing and Exchange Platform (DRSEP) of "physically dispersed and logically centralized" mode. Through this platform, the common needs of basic data resources in different industries, different departments, and different regions of a city are solved, and the optimization and utilization of data resources are promoted.

2. NEEDS ANALYSIS

According to the previous research and analysis, the data sources of the DRSEP are scattered among the application systems such as the government affairs network, the Internet and the city IT service center, Alibaba Cloud, the public security special network, and the environmental protection special network. The main characteristics of these data resources are distribution, autonomy, heterogeneity, and repeatability [3].

Distribution refers to the fact that data resources are stored in distributed geographical areas that can communicate with each other. Autonomy means that each department's information system runs its own independent application program. After the system is integrated, local data sources still maintain a certain level of independence. Heterogeneity mainly refers to the heterogeneity of the operating environment and the heterogeneity of the data model. Repeatability refers to several application systems that have been established, but the same data will exist in different application systems.

According to the principle of "large centralized, large system, and large sharing" of the DRSEP, these data resources need to be integrated into the IT service center so that various application systems can access a unified shared data resource. For this purpose, a unified Data Resources Directory (DRD) needs to be established to realize a unified view of different data sources, and then data is obtained from each data source, and a shared service is provided to requester of each data resource. The DRD is the basis for the data resource integration and sharing mechanism. At the same time, it is also the data standard and directory system for the data exchange process. Through the DRD, various business applications can be isolated from the underlying data sources, and various heterogeneous data sources can be accessed through a unified interface mechanism.

Data resources in various application systems of different networks and nodes will be integrated into the DRD. In the process of constructing a DRD, the original data resources need to be merged and reorganized in accordance with semantics, so as to realize the integration of multi-source heterogeneous data. For example, in the household registration system and the people's livelihood application systems, there is "population information", which is the different realization of semantically identical data resources in different application systems. Therefore, the population information of the household registration system and

the people's livelihood application system should be integrated into the "population information" in the basic directory according to the semantics.

After the DRD is established, data needs to be extracted from each application system. For this purpose, it is necessary to develop and deploy a data resource extraction adapter, which is responsible for extracting, filtering, and mapping data resources in the application system. The data center extracts the adapter according to the extraction policy notification or call data resource to form the resource data, and then saves the resource data to the database. Various applications and application systems that need to share data resources may make resource requests to the data center. The data center filters and maps the resource data according to the requested push policy and security settings to form the resource data required by the demand side, and then the resources data is pushed to the demand side through the data push adapter. In the push process, the push adapter can perform mapping and conversion of resource data according to the business logic of the demand side.

Considering the distributed nature of data resources, the DRSEP needs to be distributedly deployed in each region according to the principle of proximity to improve the passability and stability of data resources sharing and exchange. The distributed deployment method inevitably requires strong data exchange capabilities between deployment sites. In the exchange of data between sites, the reliability and security of the communication channel must also be fully considered.

In addition, as the integration and sharing center of the entire city's data resources, the DRSEP must fully consider the performance, reliability, and integrity of the massive quantities of information for data exchange, storage, and service [4].

3. DRSEP CHARACTERISTICS ANALYSIS
3.1 Data Type
At the business level, the data resources shared by the DRSEP are: Public Basic Data (PBcD), Public Business Data (PBuD), and Public Service Data (PSvD). The PBcD consists of four types of data: population data, corporation data, macroeconomic data, and geospatial data. PBcD is the most basic data, and it is a data resource with a relatively low frequency of change. It is the "outline" of city public data. The PBuD is business extended data based on PBcD. PBuD is composed of various indicators that are expanded according to the needs of business applications. It is a dynamic and constantly expanding business data. It is a "catalog" of city public data. The PSvD consists of data of various types of topical applications. PSvD is a service-type data set with specific application scenarios formed by cleaning, mining, and analyzing PBcD and PBuD. It provides integrated application resource services for each application unit. In the data form, the data processed by the DRSEP mainly includes: structured data, semi-structured data, and unstructured data [5].

3.2 Data Volume
By collecting, analyzing, and categorizing, a city with a population of 2.19 million people in western China needs the volume of data that runs on the DRSEP, as shown in Table 1.

3.3 Data Transmission/Application Methods
Data running on the DRSEP needs to exchange data between different networks and different regions. Therefore, in transmission mode, the platform not only needs online data transmission under the condition of network connectivity, but also needs offline data transmission between different networks under physical isolation conditions.

In the application of data resources, because some application systems have autonomous requirements for data sharing, the need to share and exchange data on the platform cannot affect the normal operation of its own application system. Therefore, the required data resources need to be firstly pushed to the application system according to the rules, and then the data is processed by each application system. Finally, the application system pushes the data to the platform for exchange and sharing. This method is called "subscription push".

Other application systems, especially various types of smart application systems deployed on DRSEP, require data sources that are integrated public data resources. Therefore, DRSEP is required to provide various data services and data analysis interfaces. Then, these smart application systems actively acquire data query services and data analysis services from the platform. This method is called "integrated data service".

4. TECHNICAL ARCHITECTURE
The DRSEP, which includes: the data resource layer, the resource site layer, the resource integration layer, the resource service layer and the resource application layer. The overall technical architecture of DRSEP is shown in Figure 1.
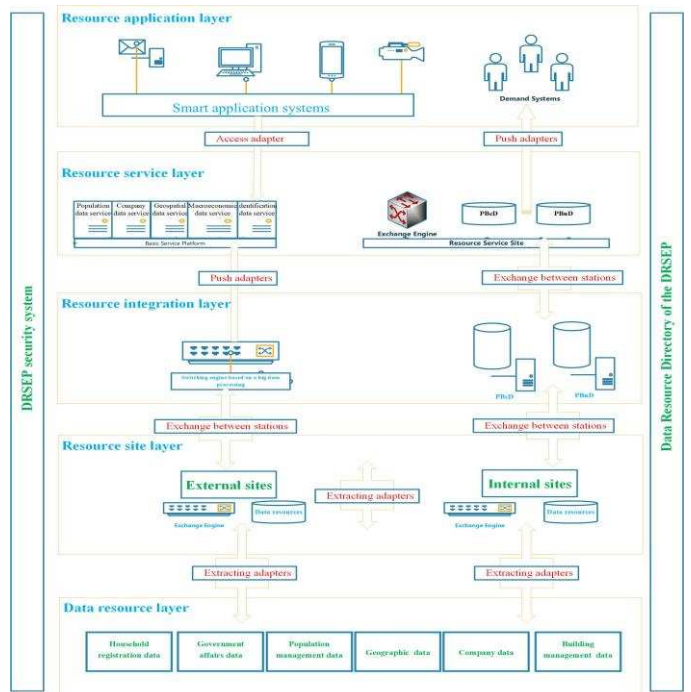
Figure 1: The overall technical architecture of DRSEP

4.1 Data Resource Layer
Through the research and analysis of data resources of various application systems such as household registration management system, government affairs system, population management system, geographic information system, company information system, and building management system, the data resources to be integrated and shared are determined. These original data resources form the data resource layer of the DRSEP.

4.2 Resource Site Layer

The resource site layer consists of external sites, internal sites, and central sites. External sites refer to resource sites set up by police stations, fire stations, taxation agencies, and companies. Internal sites refer to resource sites established outside the IT service center. The central site refers to the resource site established in the IT service center. These sites extract resource data from their original application system by extracting adapters. Then, the necessary filtering and mapping conversions are performed. Finally, these resource data are pushed to the integrated site.

4.3 Resource Integration Layer

The integrated site is set up in the IT service center and can be jointly responsible by the central site. A switching engine based on a big data processing architecture can integrate, analyze, and process it is extracted resource data, and then store them separately in the base database and the subject database.

4.4 Resource Service Layer

The resource service layer includes a basic service platform and a resource service site. The basic service platform can provide integrated query services for basic information such as population information, company information, geographic information, economic information, and personnel identification. The resource service site shares the service pressure of the resource integration site. In the area or network node where the application system deployment is particularly concentrated, the pure service type switching site is deployed.

4.5 Resource Application Layer

Various smart application systems and demand systems can acquire data resources from the resource service layer according to the different security levels established by their service objects.

5. DATA ORGANIZATION SYSTEM

Data Organization System (DOS) is based on a DRD, a Public Basic Database (PBcDB) and a Public Business Database (PBuDB).

5.1 Data Resource Directory

The DRD is the metadata standard and resource directory of the platform. Through this directory, the authority, accuracy and unity of the data resources can be ensured. At the same time, DRD also provides a data foundation when data resources are graded and authorized to use.

The DRD consists of a basic data directory and a business data directory. The basic data directory includes four types of information resources: population data, company data, geographic data, and macroeconomic data. Business data directory form a data resources in various professional fields according to the needs of different data services, including environmental protection, government affairs, taxation, education, and so on.

The DRD defines metadata as three layers: core metadata, business metadata, and application metadata. Core metadata is a basic attribute that all types of data resources must have, such as units and names. The business metadata is an extended attribute defined for the business domain to which the data resource belongs. Application metadata is an extended attribute that favors specific applications. By defining multiple layers of metadata, standards can be provided for hierarchical authorization of data resources.

5.2 Public Basic Database

The PBcDB consists of four types of databases: a population database, a corporation database, a macroeconomic database and a geospatial database. Public basic data is a basic and low-change data resource, and it is the "outline" of city public data. Public basic data can be provided by various management units. As shown in Table 2.

5.3 Public Business Database

The PBuDB includes a video database, an environmental safety supervision database, a building database, a government affairs and emergency database, a credit database and a comprehensive human resources database. Public business data expands various factors based on business application needs. It is a dynamic and expanding business data. It is a "catalog" of city public data. See Table 3.

6. APPLICATION AND PERFORMANCE

The DRSEP is deployed, tested, applied and evaluated in a city with a population of 2.19 million in western China. Moreover, the test evaluation results show that the platform exhibits an outstanding performance in the integration technology of multi-source heterogeneous data and the reliable transmission technology of massive data.

6.1 Application Effects

The platform standardizes the exchange of data resources between different departments and different regions through a unified data standard. The platform enables the integration and management of distributed data resources. It broke the barriers to data sharing and improved the quality and efficiency of the use of data resources. It realizes the interconnection of data resources in a wider range and in more fields. It also realizes the vertical intercommunication and horizontal interconnection between the city's core information systems. The application of the platform has significantly improved the ability of the government to make scientific decisions. The outstanding effects of the application are as follows:

- ➢ By establishing a DRD, the scope of shared data resources for each department is determined;
- ➢ Various data collection methods and methods for providing data services are also given;
- ➢ Carry out unified exchange, cleaning, integration and processing of city public data;
- ➢ Carry out unified organization and management of city public data, and evaluation of application performance;
- ➢ Realize the sharing of public data between city core information systems;
- ➢ Provide daily data services and spatio-temporal data services based on city public databases;
- ➢ Provide value-added services based on data mining.

6.2 Key Performance

Multi-source heterogeneous data integration, the platform is based on a unified public data model. It uses multi-source heterogeneous big data integration technology to store large amounts of data in big data storage platforms in standard data formats through data exchange, cleaning, integration, data mapping, processing, and space-time technology. Multi-source heterogeneous data integration is achieved.

Mass data reliable transmission technology, most of DRSEP's operating environment is an open Internet environment. Its transmission channel is open, and its reliability and security are

low. By establishing a site mutual trust mechanism, the two exchange sites must authenticate each other before data resources exchange and encrypt and decrypt data during data transmission. In addition, digital signatures and verifications of transmitted data ensure the credibility and integrity of the data. At the same time, when data is transmitted, in the event of an application exception, network connection interruption, system crash, etc., the platform can retransmit or retransmit after anomaly elimination, crash recovery, and troubleshooting. The data is reliably transmitted to the receiver.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Suzuki, L CSR (2016) Data as Infrastructure for Smart Cities. PhD Thesis, University College London.

[2] Amalia R. Miller and Catherine Tucker, (2011) Health Information Exchange, System Size and Information Silo, Journal of Health Economics, January 2014, 33(1): 28–42.

[3] Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu, (2008), Cloud Computing and Grid Computing 360-Degree Compared, Grid Computing Environments Workshop, 2008. GCE '08 In Grid Computing Environments Workshop, 2008. GCE '08 (16 November 2008), pp. 1-10.

[4] Boyd D and Crawford K (2012) Critical questions for big data. Information, Communication & Society 15(5): 662–679.

[5] Andrejevic M (2014) Big Data, Big Questions: The Big Data Divide. International Journal of Communication 8: 1673–1689.

[6] Taylor L and Richter C (2015) Big Data and Urban Governance. In: Gupta J, Pfeffer K, Verrest H, and Ros-Tonen M (Eds). Geographies of Urban Governance. Dordrecht: Springer, pp.175–191.

Table 1. Data Volume

Type	Data resources	Structured data/ Semi-structured data	Unstructured data
PBcD	Population data	20 million pieces of data, and the total number is basically stable.	There are a small number of attachment files, including documents, pictures, and short audio and video.
	Corporation data	1.5 million pieces of data, and the total number is basically stable.	There are a small number of attachment files, including documents, pictures, and short audio and video.
	Macroeconomic data	Add 200,000 pieces of data each year.	There are a small number of attachment files, including documents.
	Geospatial data	1 million pieces of data, and the total number is basically stable.	1TB (DLG digital line map, DOM orthoimage data, DEM ground elevation data, etc.)
PBuD	Video data	1000 Road (Live Video)	Add 20TB each year
	Environmental safety supervision data	Add 4 million pieces of data each year.	
	Building data	500,000 pieces of data, and the total number is basically stable.	100,000 items (2D plane graphics data, 3D model data)
	Government affairs data	Add 400,000 pieces of data each year.	Add 400,000 attachment files each year.
	Emergency data	100,000 pieces of data, and the total number is basically stable.	
	Credit data	Add 400,000 pieces of data each year.	
	Corporate governance data	Add 100,000 pieces of data each year.	
	Comprehensive human resources data	200,000 pieces of data, and the total number is basically stable.	Add 100,000 pieces of data each year.
Total		23.3 million pieces of data, and the total number is basically stable. Add 5.1 million pieces of data each year.	About 100,000 pieces of data+1TB, and the total number is basically stable. Add 600,000 pieces/items of data+20TB each year.

Table 2. PBcDB

Public basic data		Instructions	Type
Population database	Population basic information	ID number, name, date of birth, gender and ethnicity	structured data
	Population's residence information	Person's household registration and actual residence information	structured data
	Personnel employment information	Personnel employment status and employment unit information	semi-structured data
	Population family information	Marriage information, child information, family planning information	semi-structured data
	Personnel property information	Houses, vehicles, companies	semi-structured data
	Personnel tax information	Annual tax situation	structured data
	Personnel education information	Academic status and degree	structured data
	Qualification of personnel	Certificates and practicing licenses that the government issues to individuals	unstructured data
	Population Social Insurance Information	Low-income insurance, social insurance payment records	structured data
Corporation database	Company basic information	The unified credit code, organization code, organization name, address, and type.	structured data
	Company financial information	The company's financial data, taxation, and investment.	semi-structured data
	Scientific and technological talents	Information that reflects the company's technology and talent levels	semi-structured data
	Certification of the company	Certificates and practicing licenses that the government issues to company	unstructured data
	Environmental protection information	Quality inspection, safety production, environmental protection	semi-structured data

Public basic data		Instructions	Type
Macroeconomic database	Government-invested large projects	Government-owned assets of large companies. Government-invested public projects	structured data
	Economic indicators	Statistics on government finance income, tax revenue, investment, exports, imports	semi-structured data
	Economic operating data	Product information of various industries, financial information, talent information, scientific and technological information	semi-structured data
Geospatial database	Base map data	DLG digital line map, DOM orthophoto data, DEM ground elevation data	unstructured data
	Geographical object information	Including roads, rivers, lakes, dams, buildings, and vegetation.	structured data
	Place name and address information	companies, schools, hospitals, banks, markets, parks, bus stops, public toilets and their exact address	structured data

Table 3. PBuDB

Public Business Database		Instructions	Type
Video database	Live video	Real-time viewing of monitoring video information	structured data
	Past video recording	Recording data of monitoring videos that need to be reserved for a certain period of time in key areas	unstructured data
environmental safety supervision database	Environmental monitoring data	Statistical data generated from environmental monitoring data	structured data
	Energy consumption data	Energy consumption statistics for companies and public institutions	structured data
	Environmental Information	Information on environmental monitoring, protection and early warning	structured data
	Safety production monitoring data	Data generated after statistics of company safety production monitoring data	structured data
	Dangerous goods transport information	Transportation of hazardous wastes and dangerous chemicals transport vehicles	structured data
Building database	Building basic information	Building code, building name, detailed address, construction age, construction status, service life, main use, structure type, building number, basement number, building height, and total construction area.	structured data
	Building model data	The basic data required for multi-level and visual representation of buildings includes: 2D plane graphic data, 3D model data, etc.	unstructured data
	Building facility information	Water, electricity, high voltage electricity, sewage treatment, communications and other ground and underground network information	structured data
Government affairs and emergency database	Service business	All business related to administrative approvals, public services, intermediary services and other social organization services	semi-structured data
	Company applies for business	Processing status and processing information of the company's requested business	semi-structured data
	Negative list	Negative list of government service clients (individuals companies, etc.)	structured data
	Emergency resources	Can be used for emergency announcement of electronic advertising screens, regional broadcasting, cable television, etc.	structured data
	Emergency refuge information	Emergency refuge information	structured data
Credit database	Personal credit rating	Overall evaluation of personal credit information	structured data
	Personal honor information	Personal positive credit situation	semi-structured data
	Personal negative information	Individual violations, administrative penalties, failure to perform legal obligations	semi-structured data
	Company credit rating	Overall evaluation of company credit information	structured data
	Company honor information	Company positive credit situation	semi-structured data
	Company negative information	Company violations, administrative penalties, failure to perform legal obligations	semi-structured data
human resources Comprehensive database	Basic house information	Housing address, affiliated community, property owner, housing occupants, housing water, electricity, gas and heating payment	structured data
	Digital book	Book titles, authors, publishers, publishing years, keywords, introduction, frequency of reading	structured data
	Book borrowing information	Personal borrowing information in the library	structured data
	Electronic health record	Past medical records and medication status, related inspection attachments	semi-structured data
	Daily medical examination data	Daily blood pressure, blood glucose, weight, height and other test results, related inspection attachments	semi-structured data
	Booking clinic information	Current Medical Clinic Booking Information	structured data
	Medical Information	Preferential medical information, holiday medical institution business hours adjustment information, etc.	structured data
	Sports venue information	Current reservation information	structured data
	Food Safety Information	Food safety policies, regulations, announcements, notices, events, etc.	structured data