AI-Enhanced Government Data Governance in the Era of Big Data: Exploring Sharing Mechanisms and Management Systems

Authors: Xiao Jiong'en 1,2, Wu Yingliang 1

Abstract

This study delves into the literature on management and decision making driven by big data, identifying the sharing mechanisms of big data resources and the interconnection technology of information silos as one of the cutting-edge topics in contemporary big data research. By examining and analyzing the application of government data sharing and exchange domestically and internationally, this paper summarizes issues related to management philosophy and data barriers caused by legacy systems in government data resource sharing and exchange. Based on cloud platforms and in conjunction with the theory of Data as a Service (DaaS), a framework for managing the entire volume of government data resources is proposed. This framework ensures that no alterations are made to existing systems, achieving data non-movement, non-replication, and preservation of the original management model, while defining the rights and obligations of various operating entities regarding data. It addresses the management philosophy issues and systemic barriers encountered in data sharing and exchange.

Keywords Comprehensive Data Resources; Data Exchange Framework; Management Mechanism

1. Introduction

At the 5th Alibaba Technology Forum, Jack Ma proposed that we are currently transitioning from the IT (Information Technology) era to the DT (Data Technology) era [1]. In the Data Technology era, big data is both a category of data and a technology. As a type of data, it exhibits characteristics such as large volume, rapid growth, diversity, and low value density; as a new generation of information system architecture and technology, it can collect, store, and perform associative analysis on data that is massive in quantity, dispersed in source, and varied in format [2]. The sharing mechanisms of big data resources and the interconnection technology of information silos are among the frontier topics in big data research today [3].

The era of big data has brought about a shift from "sample data" to "comprehensive data." With the continuous reduction in the cost of hardware equipment and the improvement in performance, IT architecture has undergone tremendous changes. It has shifted from the traditional design concept of "saving equipment space" to new generations of architectures such as "distributed" and "clustered." Through the new generation of data storage architecture, it achieves the storage and management of comprehensive data, supporting the needs of big data exploration. Comprehensive government data is an important class of data resources in the era of big data, which includes data from fields such as business, taxation, administration, and social security management. By opening up, optimizing, and sharing government big data, administrative costs can be reduced, work efficiency can be improved, and administrative service capabilities can be enhanced. Therefore, the study of government big data opening, sharing, and optimization not only has prominent scientific forefront features but also has significant practical value [4-6].

This article will start with the current management of government data resources, summarize several important government data resource management models from abroad, explore the problems faced by the current management of government data

resources, and on this basis, discuss innovative mechanisms for sharing and exchanging comprehensive government data resources. It aims to provide strong support for the sharing and exchange of comprehensive government data resources at the provincial and municipal levels in the era of big data, and to provide effective practical experience for the sharing, exchange, and optimization of big data technology in the government sector.

2. Research on Government Data Sharing and Exchange and the Challenges Faced

2.1 Application Research on Government Data Sharing and Exchange

Government data sharing and exchange have garnered widespread attention, and various regions have implemented some practices. There are currently three main mechanisms for government data sharing and exchange.

2.1.1 Distributed Service Bus Exchange Mechanism

ESB, which stands for Enterprise Service Bus, is a product of the combination of traditional middleware technology with XML, Web services, and other technologies. It provides the most basic connection hub in the network and is a necessary element in building the enterprise nervous system. Technology based on the service bus is one of the solutions for government data sharing and exchange [7]. The distributed government data resource sharing exchange based on the service bus adopts a service-oriented architecture (SOA) [8], combined with corresponding service semantics, process services, and other technologies to build e-government solutions [9]. For example, Changsha City has adopted the service bus architecture, as shown in Figure 1.

The distributed service bus mechanism has three issues: (1) Publishing services requires partial modification or reconstruction of the original system; (2) Data exchange between entities requires complex configuration services; (3) The work of data cleaning and comparison is handed over to the user end, causing a huge workload on the user side.

2.1.2 Data Centralized Sharing Mechanism

The data center adopts a large-scale centralized model to achieve data sharing [10], collecting data at various levels, collecting data from multiple different institutions, merging and analyzing source data to form a unified resource library. Currently, many national departments, provinces, cities, and universities have adopted the large centralized model of building a data center to achieve data sharing, such as the construction of the National Environmental Data Center, which adopts a large centralized model [11], as shown in Figure 2.

The advantage of this mechanism is that it can unify data resources and ensure the consistency of data provided externally. However, this mechanism leads to a large storage volume of the data center, and the need for collecting, comparing, cleaning, and transforming heterogeneous data from different sources, resulting in a huge workload on the data center server side, and there is an issue of not being able to ensure the timeliness of the data. At the same time, it faces the problem of execution cooperation, whether various departments and units will cooperate with data centralization and sharing, and data sharing will affect the rights and responsibilities of some departments.

2.1.3 Data as a Service (DaaS) Based Sharing Exchange

Overseas applications of SOA and service buses for government data sharing and exchange have also been practiced. With the development of technology and the rapid development of cloud computing and large databases, Olson [12] proposed the concept of Data as a Service (DaaS) in 2009. Higgins et al. [13] have done some more cutting-edge explorations, conducting a lot of research on the acquisition, analysis, and processing of data between public sectors, starting from the problems that hinder the sharing and exchange of information, proposing a more comprehensive information sharing framework, and verifying the effectiveness of the framework through examples. Terzo et al. [14] studied the DaaS framework based on the cloud platform, managing big data through DaaS, and proposed a complete and effective DaaS framework. The basic structure of DaaS is shown in Figure 3.

DaaS is an application based on the IaaS layer, proposing a very advanced data management concept. However, using the interfaces provided by IaaS, it is still necessary to make corresponding modifications to the IaaS system. The research process has found that there are still difficulties in data sharing and processing, the number of databases is constantly increasing, and the interaction between systems is becoming more and more complex. It is still necessary for the development departments of each system to make corresponding modifications to the system. Olivier's research did not propose a good technical method to avoid modifications to the IaaS system.

2.2 Challenges in Government Data Sharing and Exchange

By analyzing the distributed service bus exchange mechanism, centralized data sharing mechanism, and DaaS (Data as a Service) model, the following issues in current government data sharing and exchange are summarized.

2.2.1 Barriers Caused by Legacy Systems

Data resources are distributed across different systems, managed by various entities that control their own resources. Heterogeneous and non-interoperable data create numerous information silos. Adopting a data center or service bus model requires the reconstruction, upgrade, and modification of the original systems, necessitating the cooperation of the original development companies, which entails a significant workload. Even with the cooperation of all units, the effectiveness of data sharing and exchange is still limited. The fundamental reason is that the above architectures require a high degree of manual involvement, resulting in poor real-time performance of the data. This is due to the overly complex implementation process of the systems. Non-real-time data during the exchange process, high system maintenance costs, and low benefits hinder the continued investment in the construction of data sharing and exchange, leading to a vicious cycle.

2.2.2 Obstacles of Data Management Philosophy

In the era of big data, the management philosophy of many organizations still adheres to traditional concepts. The goal of data management has not shifted from data management to data governance, leading to numerous obstacles in the work of data sharing [12]. Data sharing and exchange involve multiple business departments and address issues of rights and responsibilities during the data sharing and exchange process. The issues of rights and responsibilities of various agencies regarding data, as well as concerns about the storage, security, and ownership of data after sharing and exchange, are not clear enough. Some units worry about losing control of the data after

sharing and exchange and fear data leakage. The above-mentioned issues related to management philosophy affect the advancement of data sharing and exchange efforts.

3. Government Comprehensive Data Resource Management Framework

In light of the aforementioned issues, this paper proposes a cloud platform-based management framework for government comprehensive data resources, integrating the theory of Data as a Service (DaaS). The framework is demand-driven, employs advanced technologies, and robust management mechanisms. It ensures that no alterations are made to the existing systems, with data remaining in their original locations without migration. It defines the rights and obligations of various operating entities, addressing the management philosophy issues and systemic barriers in data sharing and exchange. This paper first presents the framework for government comprehensive data resources and then discusses solutions to facilitate a shift in data management philosophy and the dismantling of legacy system barriers. The framework for the management of government comprehensive data resources is illustrated in Figure 4.

3.1 Framework of Government Comprehensive Data Resources

The framework of government comprehensive data resources consists of three levels: the infrastructure layer of the cloud platform, the data layer of the data sharing and exchange platform, and the service layer of the data opening platform.

In the government comprehensive data resource framework, the cloud platform serves as the foundational layer of the entire system. With the development of cloud computing in recent years, various regions have successively established government cloud platforms, and most of the government bureaus and offices have undergone cloud migration. The deployment of government application systems on a unified cloud platform is a favorable condition for the establishment of the government comprehensive data resource management mechanism. For systems that have not yet

migrated to the cloud platform, they must also be connected to the electronic government network to ensure the effective implementation of the government comprehensive data resource management system.

The data layer of the data sharing and exchange platform is based on reflective middleware research [15-17] and system runtime architecture reconstruction technology [18], further deepening research and improvement to achieve the automatic generation of system data interface APIs (Application Programming Interfaces) without any modifications to the original systems and without any changes required from the original system development companies. It employs a microservices-based approach for encapsulation, deployment, and runtime control, management, and supervision. The interface, during operation, acts as a data access bridge outside the original system, reading or writing data by simulating operations on the actual system.

The service layer of the data opening and sharing platform includes unconditional openness to the public, contractual openness to big data industry enterprises, and mutual sharing and exchange of data between government departments.

3.2 Innovative Solutions to Government Data Sharing and Exchange Issues

The government comprehensive data resource management platform innovatively addresses the challenges of system silos and management philosophy barriers.

3.2.1 Innovative Technology to Eliminate Legacy System Barriers

Adopting the DaaS concept of data sharing and exchange, and based on in-memory analysis technology and runtime system architecture reconstruction technology [19-21], a system index pipeline is established between the systems of various units to break down the barriers between legacy systems and eliminate information silos. On this basis, five major thematic databases and a data resource interface pool are established to achieve data openness.

The five basic libraries, as shown in Figure 4, are logical databases, different from the data center model. Their essence is not to store any data but to perform rule validation, data source merging, and rule collation and integration based on multiple application systems providing data interfaces. After integration, multiple data interfaces provide a single data interface externally. This innovative mechanism ensures the real-time nature of the data.

Data interfaces from various units are collected to build a data resource catalog, and a data resource pool is constructed through automatically generated data interfaces. Users obtain data resources provided by relevant units according to the published resource catalog, and the data interface resource pool is continuously optimized according to actual situations.

3.2.2 Innovative Technology to Facilitate the Shift in Data Management Philosophy

The government comprehensive data resource management system does not store data; instead, data remains in the original systems and is exchanged through data interfaces. During the data exchange process, the construction of data interfaces is carried out through the data sharing and exchange platform, but the data is not stored on the platform. Users grasp the status of data openness through the data resource catalog, submit applications for the required data resources through the system, and the department where the data is located has ownership of the data, reviewing applications

to determine whether to provide data and the period of data openness.

Based on software architecture reconstruction technology, it greatly reduces the investment in system transformation funds for various units, improves the ease of system sharing and exchange, and encourages departments to clearly define their rights and obligations regarding data resources. The data management philosophy shifts from a passive to an active approach, achieving collaborative data governance across the organization.

In summary, based on the framework of government comprehensive data sharing and exchange, the application of advanced technology can solve the problems faced by the service bus and data centralization models. The efficient operation of the government comprehensive data sharing and exchange system also requires an efficient operational management mechanism. The operational management mechanism includes two aspects: one is the management of people, specifically including the responsibilities, tasks, and powers of users; the second is process management, which ensures smooth data flow and continuous updating and optimization of data through the establishment of processes.

4. Government Comprehensive Data Sharing and Exchange Management Mechanism

User role management mechanisms and data exchange process mechanisms are two critical aspects of building a government comprehensive data sharing and exchange system, supporting the effective operation of the government comprehensive data sharing and exchange system.

4.1 Innovative Data Sharing and Exchange User Management Mechanism

The government comprehensive data resource sharing and exchange platform is a platform for precise docking of government data sharing supply and demand. The main roles are divided into four major categories: data users, data providers, platform regulators (including security regulators), and platform operators.

Data users are those who have the right to browse and access the data services provided by the platform, including everyone from the public, businesses, and staff of various committees and offices; Data providers are business system administrators of various committees and offices, or individuals authorized by their units to provide data to the data sharing and exchange platform; Platform regulators (including security regulators) provide the sharing and exchange working mechanism, understand the dynamics of sharing and exchange through a visual platform in real-time, and are responsible for supervising and controlling the use and maintenance of the platform. The regulators are the government's big data management agency and the public security bureau; Platform operators provide the technical implementation environment for sharing and exchange. Operators, through backend management, are responsible for providing tools for data collection, cleaning, and processing for the data sharing platform, and provide technical support for data users through the platform. By providing platform tools, they offer services such as data collection, cleaning, processing, and sharing, and charge corresponding operation fees. The government-authorized data management company has exclusive operating rights. The user roles of the government data sharing and exchange are shown in Table 1.

4.2 Innovative Data Sharing and Exchange Process Mechanism

The efficient operation of the government comprehensive data resource management system, in addition to requiring an advanced and rational framework, must also be supported by a comprehensive sharing and exchange process and a data channel publication process.

4.2.1 Data Sharing and Exchange Process

Data users must first log in to the shared exchange platform's data catalog service subsystem and search for the required data resources according to the business responsibilities of various units. Depending on the classification of the data's confidentiality level, which is determined by the unit to which the selected data belongs, if it is unconditionally shared data, the data user can directly obtain the data channel calling method to access the data. If it is conditionally shared data, a data resource acquisition application can be made. The unit to which the data belongs reviews the application through the sharing and exchange platform. After confirming that the basis for the data acquisition application is sufficient, legal, and valid, they will grant approval. The sharing and exchange management submodule controls the permissions for the data channel calling method and authorizes the approved data users. Only then can data users obtain the data channel calling method to access the data. The data exchange process between business units and platform data is shown in Figure 5.

When data users make data channel calls to obtain data, it is driven by the data pipeline operation service engine. Through the system data API interface, data exchange is completed from the business system of the information resource provider to the business system of the information resource user. The sharing and exchange platform will monitor the exchange tasks carried by the data pipeline and record, summarize, and analyze the status of the monitored tasks.

4.2.2 Data Channel Publication Process

After the platform's operational technical team receives applications for new publications from various units' data catalogs, they will initiate the work of activating the generation of data catalogs, creating, and publishing data channels. Using DaaS technology, once the data catalog and data channels are generated, the service configuration administrator needs to set up in the backend management system according to the system information of the new publication applications from various units' data catalogs. They will complete the registration and publication of the catalog and channel, save the published service information in the platform management database, publish the data channel, update and announce the data catalog, and users can obtain new provided fields and related data through the updated catalog. The data channel publication process is shown in Figure 6.

5. Conclusion

This paper studies the current state of data sharing and exchange platforms, analyzes existing problems, and proposes an innovative framework for comprehensive data resource sharing and exchange. It constructs a platform for the management of comprehensive data resources, breaking down barriers between systems and assisting units holding data to transform their management philosophy. The paper elaborates on the innovative management of government comprehensive data resource sharing and exchange from the dimensions of technical practice, management philosophy, and management mechanisms. The design ideas and overall plan of this paper have been

practiced in cities that are vigorously developing and pioneering in big data, achieving good results and providing effective practical experience for other cities to build data sharing and exchange platforms.

References

1. Ma Y., Zeng M., Tu Z. "Internet Plus: From IT to DT": A New Economic and Social Development Strategy that Citizens Need to Understand [J]. Decision and Information, 2015(12): 67-69.

2. Miao Y. Big Data, the Key Resource for Changing the World [J]. China Economic and Trade Guide, 2015(31): 10.

3. Xu Z., Feng Z., Guo X., et al. Frontier Issues in Management and Decision Making Driven by Big Data [J]. Management World, 2014(11): 158-163.

4. Lu P. Construction of Library Information Resource Sharing Management Model [J]. Library Forum, 2010, 30(1): 86-88, 4.

5. Zhang M. Analysis of Information Resource Exchange Organization Based on Resource Sharing [J]. Information and Documentation Work, 2008(4): 77-81.

6. Wang R. Design and Empirical Study of the Anti-Poverty Measurement Index System in Rural Areas: A Case Study of Yunnan [J]. China Rural Economy, 2006(12): 69-76.

7. Li X., Yang Y., Guo W. Data Sharing and Exchange Platform Based on Enterprise Service Bus [J]. Computer Engineering, 2006(21): 217-219, 223.

8. CHAI H., ZHOU X., YANG G., et al. Information Share and Exchange Platform Based on ESB [J]. Microelectronics & Computer, 2008, 25(4): 116-120.

9. Ye Y., Li J., Wei D. Research and Implementation of an E-Government Platform Based on SOA [J]. Microcomputer Information, 2010, 26(12): 153-154, 118.

10. Shi X. Design and Implementation of a Data Integration Platform for Shared Data Centers [D]. Suzhou: Soochow University, 2006.

11. Li S., Xu F., Kong Y., et al. Discussion on the Construction of the National Environmental Data Center [J]. Environmental Protection, 2006(2): 77-81.

12. OLSON J. Data as a Service: Are We in the Clouds? [J]. Journal of Map and Geography Libraries, 2009, 6(1): 76-78.

13. HIGGINS E., TAYLOR M., LIABOA P., et al. Developing a Data Sharing Framework: A Case Study [J]. Transforming Government: People, Process and Policy, 2014, 8(1): 151-164.

14. TERZO O., RUIZ P., BUCCI E., et al. Data as a Service (DaaS) for Sharing and Processing of Large Data Collections in the Cloud [C] // IEEE Computer Society Press. Seventh International Conference on Complex, Intelligent, and Software Intensive Systems. Taichung: IEEE Computer Society, 2013: 475-480.

15. Tong N., Chao L. The Change of Data Management Philosophy in the Era of Big Data: From Result-oriented to Process-oriented [J]. Information Theory and Practice, 2017, 40(2): 60-65.

16. Huang G., Wang Q., Mei H., et al. Research on Reflective Middleware Based on Software Architecture [J]. Journal of Software, 2003(11): 1819-1826.

17. Huang G., Mei H., Yang F. Runtime Software Architecture Based on Reflective Middleware [J]. Science in China Series E: Technological Sciences, 2004(2): 121-138.

18. Song H., Huang G., Wu Y., et al. Modeling and Maintenance of Runtime Software Architecture [J]. Journal of Software, 2013, 24(8): 1731-1745.

19. Yang J. Research on the Assembly Technology of Network-oriented Software Based on Software Architecture [D]. Beijing: Peking University, 2007.

20. Mei H., Huang G., Lan L., et al. An Adaptive Method for Service-oriented Architecture Software Based on Architecture [J]. Science in China Series E: Information Sciences, 2008(6): 901-920.

21. Mei H., Huang G., Zhao H., et al. A Service-oriented Software Development Method Centered on Software Architecture [J]. Science in China Series E: Information Sciences, 2006(10): 1100-1126.

Author Biographies

Xiao Jiong'en (1984 —), male, from Shantou, Guangdong, doctoral candidate, main research directions include big data application analysis, data resource sharing and exchange, cloud computing and applications;

Wu Yingliang (1963 —), corresponding author, male, from Taihu, Anhui, professor, doctoral supervisor, main research directions include information resource management and information systems, knowledge engineering and knowledge management, e-commerce.

1. Department of E-commerce, School of Economics and Trade, South China University of Technology, Guangzhou, Guangdong 510006;

2. Guiyang Municipal Big Data Development Management Committee, Guiyang, Guizhou 550081