IP for Sharing and Opening Government Data Based on Big Data

Muchuan Yang[1, *], Shaoliang Yang[2]

[1]Coventry University, Coventry, UK

[2]Baoding Preschool Teachers College, Baoding, China

*Corresponding author: 393314939@qq.com

Abstract. With the continuous application and popularization of big data technology, all government departments are aware of the importance of data as a national basic strategic resource, and all over the country have been promoting the sharing and opening (SOA) of local government data (GA) one after another, improving the organization, formulating government regulations, and building a GA SOA platform. At present, GA are kept separately in each department, while there is no good DS scheme, resulting in low data sharing (DS), lack of authenticity of shared data, easy leakage of data, and difficulty in data tracking. Through the analysis of the above problems, an integrated platform(IP) of GA SOA based on big data is designed, aiming to realize safe and controlled sharing of GA. The application of this platform helps to efficiently integrate data resources and maximize the realistic value of GA for people's services in order to achieve sustainable social development.

Keywords: Government Departments, Big Data, Government Data, Sharing and Opening Integration.

1. Introduction

Under the rapid development of the big data era, the government is also in the process of developing innovative administrative methods and improving e-government construction, so the construction of GA sharing and open platform has become a necessary development trend for government information disclosure and improving digital governance ability. In this context, the research of this paper has a great theoretical and practical reference value for the further construction and improvement of the GA open platform.

With the rapid development of the big data industry, domestic research on the open sharing of GA has been emerging, and many scholars have been exploring the construction and application of the open platform for GA sharing and practice. Some scholars believe that the construction of GA open platform can create economic benefits for governmental organizations, non-governmental organizations and individuals, which is conducive to government departments to play their economic functions reasonably and efficiently, improves the reliability of the strategic policy of enterprises, and the effective use of data can increase the overall economic benefits of society [1]. Some scholars suggest that in the era of big data development, the government open data can stimulate new market vitality, provide data guidelines for the development of various industries, and improve the economic benefits of enterprises, in addition, in the era of big data, open data can create new jobs and optimize the allocation of resources in the talent market [2]. Some scholars point out that the government should improve the national open GA policy and laws and regulations, make appropriate policies and regulations for different situations in different places, and supervise the policy formulation, policy implementation and policy improvement to effectively guarantee the security of open data [3]. It can be seen that the research on the economic mechanism and construction rules of GA sharing and open platform has achieved good research results.

This paper firstly introduces the technology and platform construction features for building an IP for GA SOA, then designs the platform deployment architecture and functional architecture, and carries out the metadata design of the platform system, then carries out the platform response performance and data synchronization delay test, and finally analyzes the DS application effect of the platform.

2. Basic Overview

2.1 Relevant Technologies for Building an IP for SOA Government Affairs Data

(1) Front-end technology

The front-end technology is a good technical solution for data transmission across network security domains. The shared data is copied or automatically synchronized to the designated front-end storage location for private network systems or socialized data. Then, according to the actual situation, application services can be deployed in the front machine to provide standard API interfaces to the outside world, or the data can be exported and synchronized to the shared open platform by the cache database to provide services to the outside world in a unified manner [4].

(2) Data cleaning and processing technology

Data resource cleaning is the processing of GA, checking data consistency, finding invalid and missing values in the data, and processing these useless data information.

(3) Data desensitization and decryption technology

Data pre-processing is required for sensitive, confidential or sensitive data in the process of GA sharing, and the desensitization and declassification process mainly adopts generalization, suppression, scrambling and lossy techniques to achieve protection by sensitive data, and at the same time, through machine learning, data association, knowledge base and other technologies, sensitivity analysis is performed on the data after multi-dimensional aggregation to achieve desensitization and desensitization of data datasets with multiple and and data security [5].

2.2 Features of Building an IP for SOA GA

First, the pluralistic cooperation of construction subjects. Data openness is a product of the times further developed under information disclosure, and is an upgrade of government information disclosure. The GA SOA platform allows the public to freely download, share and use the data resources shared by the platform within the scope allowed by laws and regulations. In addition, the platform sets up an interactive communication window to facilitate the public or individuals to raise questions, applications and opinions to the platform manager [6].

Second, the relative uniformity of the construction rules. The data shared by the platform has relatively unified -ness, and there are unified specification standards or reference methods in terms of data fields, data catalogs, metadata descriptions, download formats, etc. For example, in terms of data fields, they are basically concentrated in the fields of government, economy, education, medical care, transportation, etc.; download formats are also carried out in a variety of formats for users to choose; the DS methods of the platform are improved on the basis of domestic practices combined with regional situations [7].

Thirdly, the transparency and originality of the construction data. The open form of data in the platform emphasizes the originality of data and pays attention to the transparency of data. With the rapid development of the era of big data, enterprises or individuals have the technical ability to analyze and utilize the original data of the government by themselves, which is conducive to adding value to the open data of the government and improving the social benefits created by the open government [8].

Fourthly, the dynamic development of the construction content. The data collection of the platform needs to coordinate and manage the data of government departments at all levels, and the departments at all levels should actively update and upload the latest data under the guidance of open data regulations, so the data of the platform has dynamic development. Whether it is the government or enterprises or individuals, the data for reference and analysis in handling affairs need the latest data in relevant aspects, and the update frequency of the data directly affects the value of the data used by the public. Only by ensuring the continuous updating of data can the platform have a constant vitality of use and create a constant value of open data. In addition, the construction of the platform should be continuously improved with the changing needs of the public [9-10].

3. Design of an IP for SOA GA Based on Big Data

3.1 Deployment Architecture

For the unified management of the GA SOA platform, guaranteeing data security, solving the lack of information technology operation and maintenance forces, saving hardware resources and reducing information technology construction costs, this platform is based on cloud computing and big data new generation information technology, adopting the idea of unified deployment and hierarchical management to realize the flexible configuration and use of IT resources, and the overall deployment architecture is shown in Figure 1.

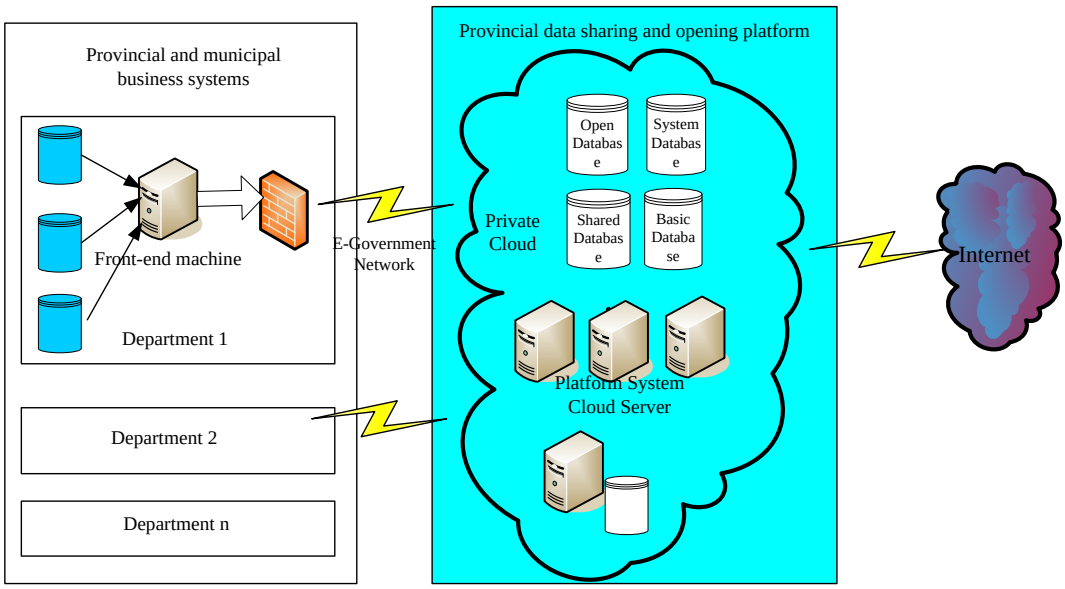

Figure 1. Deployment architecture of the open platform system for GA sharing

In the deployment architecture diagram, the front-end of the business systems at the provincial and municipal levels are deployed on the front-end, and the communication links are opened through the e-government network configuration, and the data collection system is installed on the front-end to interface the business system database with the database system of the GA SOA platform; due to the special nature of the GA resources and the data security level requirements, the platform is deployed on the private cloud, which can ensure the security and privacy of the GA. resources security and privacy, and also allows resource storage and calculation when accessing data from business departments; finally, the data that can be shared and opened are shared to the whole society through the data SOA portal of the platform, and data resource services are provided for all walks of life according to the subject classification, so as to finally realize the open sharing of GA resources [11-12].

3.2 Functional Level of GA SOA Platform System

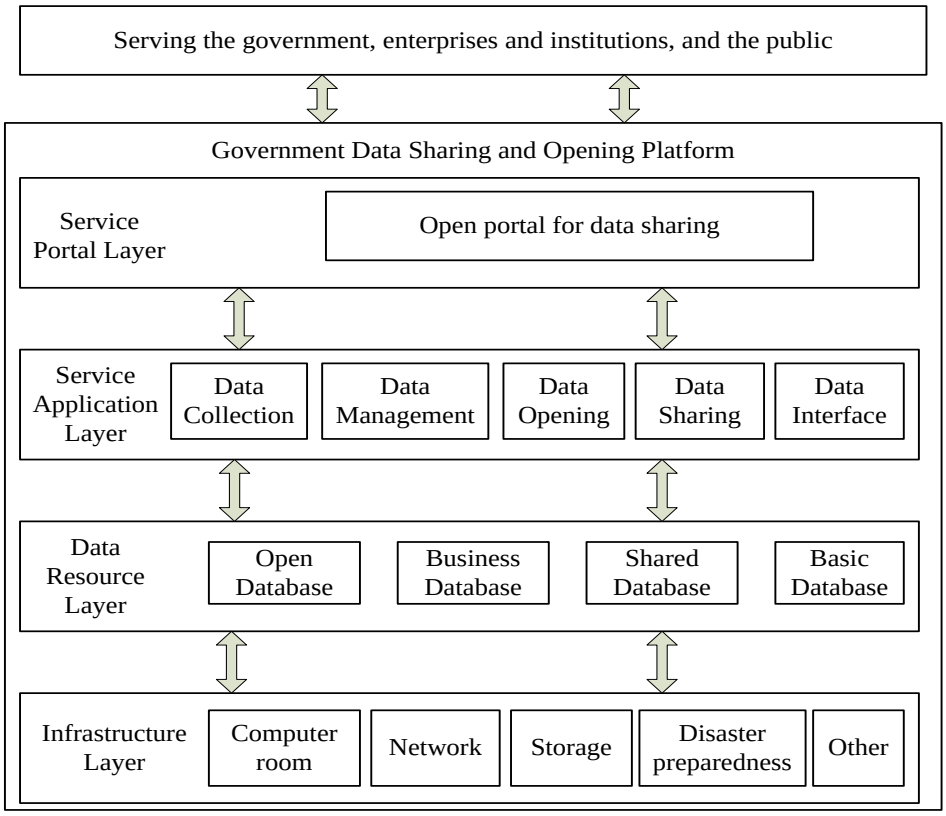

Figure 2. Functional architecture of the platform

The functional architecture of the GA SOA platform is shown in Figure 2, and the functions of each layer are introduced as follows.

(1) Infrastructure layer

The construction of the infrastructure layer is to provide the platform with a reliable physical environment for the server room, network resources, storage resources, disaster recovery and backup hardware to ensure that these key equipment resources can operate safely, stably and reliably, which is the basic support for the whole platform.

(2) Data resource layer

In the data resource layer includes two levels of database construction, one is the basic database and system platform database, which is the basic data support of the whole platform, and the other is the business-related open database, shared database and business database, which are directly used for the upper system application services and data resources of the data open sharing platform portal.

(3) Service application layer

Data collection service. The data collection service (system) of the GA sharing and open platform is a data docking interface connecting the data and information resources of various government business departments and the platform, and is used to collect data resources of the business systems of various government departments. The data collection system is deployed on the front-end machine, relying on message communication middleware and data collection driver component design to realize the collection of business data information of each government department. The data collection system can flexibly carry out data collection driver configuration to collect data information according to the type of business data of each government department, and can carry out data transmission configuration according to the data characteristics, such as batch update, incremental update, update period, update frequency and other configurations to realize data synchronization between the upload database of the collection system and the basic database of the platform.

Data management service. Data management service is a comprehensive, systematic, long-lasting and visualized unified management of all data information in the basic database of the platform. It displays data base information, data level and update operation records, including creation time, modification time, update content, user at the time of update and other information, and adds, deletes, and checks operations on database, table, table structure and data according to different types of data. Data interface service. Provide API interfaces for data interaction, forming a complete set of platform data interface calling services to meet the needs of the community for data interface services.

(4) Service portal layer

This layer provides provincial, municipal and county-level GA SOA portals.

3.3 Metadata Design

The original design of the government affairs data SOA platform is to realize safe and controlled data exchange between departments and let the masses realize one-stop affairs processing, so the key factor in metadata is ID card number, so it is proposed to use ID card as the index to store and find metadata on the chain.

Each government department has an independent storage system with different storage formats and large data volume. Too much data will be transferred to the blockchain in a short period of time, which will lead to the degradation of the blockchain performance, and it is not worthwhile to share all data through the blockchain. At the same time, because all nodes keep a complete blockchain data, on the one hand, it is insecure for data with high privacy requirements, on the other hand, for the government system with more data will undoubtedly increase the storage pressure as well as hardware cost, and it is not possible to share all information directly through the blockchain for DS, so it is proposed to use smaller metadata to represent larger and complete shared data in different data formats on the chain. The platform program is stored in the database as new data.

$$sim(m_i, n_i) = 1 - \frac{2 \times |m_i - n_i|}{|m_i + n_i|} \tag{1}$$

m_i and n_i are numerical attributes, respectively, and $sim(m_i, n_i)$ is the similarity between attributes.

$$sim(p_1, q_1) = \frac{\sum_{i=1}^{n} w_i sim(p_i, q_i)}{\sum_{i=1}^{n} w_i} \tag{2}$$

p_1, q_1 are two data in the database whose similarity is used to represent $sim(p_1, q_1)$, and w_i is the weight of the ith data attribute.

4. Platform Testing and Application

4.1 Platform Testing

To test the response time of the system, the application system performs operations such as adding data source tables, deleting data source tables, adding files, deleting files, resource hooking, resource sharing, etc. Each operation generates or deletes only one record, and then tests the system response time of one record operation, and the test results are shown in Table 1.

Table 1. Running time of each operation

Request Type	Run Time
Add library table	16ms
Delete Library Tables	20ms
Adding a file	345ms
Delete a file	62ms
Resource Linking	183ms
Resource Sharing	36ms

As can be seen from Table 1, for single-table add, delete, and check operations, the response is relatively faster, and the speed drops for multi-table modifications. Operations involving file upload and download are the slowest due to the size of the file and the network speed.

The data synchronization efficiency of the system is queried. The test object is set as six groups of data tables of data front library, and different data tables are used to insert data with different data volume to detect the synchronization time delay of data synchronization library. The synchronization delay refers to the time required from the time the data is inserted into the front library to the time the data is updated in the data center library. The data synchronization delay is calculated based on the insertion time and the synchronization time. The test results are shown in Figure 3.

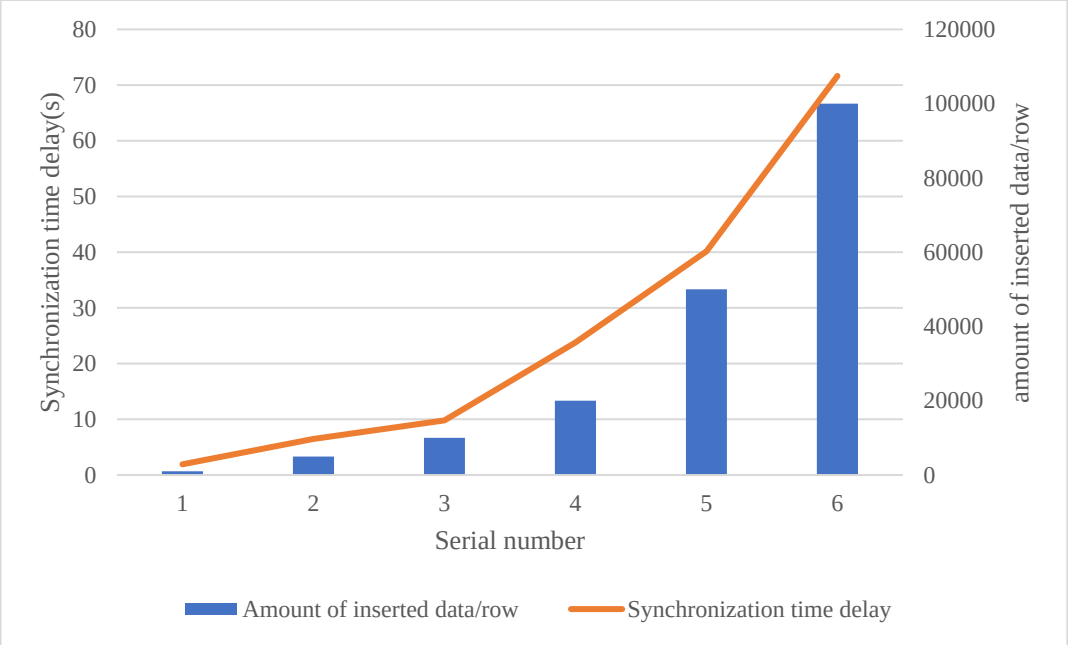

Figure 3. Data synchronization delay test

From the test results, it can be seen that when the data volume is less than 10000, the synchronization delay of the test is less than 10s, and the main bottleneck of data synchronization delay is network time consumption; when the data volume is more than 10000, the synchronization delay also starts to increase as the gap between single group data becomes larger, and the synchronization delay reaches 71.64s when inserting 100000 rows of data volume, which does not exceed the threshold of 100s. The test result meets the requirement.

4.2 Application Effect of the Open GA Sharing Platform

Putting the GA sharing and open platform into major provincial and municipal applications provides the society with a channel for data query, download and application. The data usage of a city applying the platform was collected, and the results responded generally well. From the city's application of the platform to date, the number of visits to the open GA platform has reached more than 120,000, and the number of downloads has exceeded 65,000. According to the platform data statistics, the top 5 data sets downloaded from the platform are: medical institution practice license

information (36,742 times), local taxation tax registration information (4,109 times), city population size and distribution (annual) (1,435 times), basic information of commercial entities (587 times), and information of senior executives (434 times); the top 5 data sets in terms of interface usage They are: querying water and rain information according to the name of reservoirs (21 times), querying the number of change registration information (15 times), querying the list of kindergartens in the city (9 times), querying the basic information of travel agencies according to the name of travel agencies (8 times), and querying the use of foreign capital in the city according to the name of indicators (5 times).

5. Conclusion

In the current environment of rapid development of the Internet and the rapid progress of big data technology, the smooth promotion of the construction of the GA sharing and open platform requires strong support from the government and continuous innovation of data technology. Open GA can collect more public opinions, provide better public services for citizens and improve the government's public management ability, thus creating more social and economic values. On the one hand, this paper realizes the centralized storage and dynamic management of GA through the construction of an IP for SOA GA; on the other hand, the sharing of various departments based on application protocols is conducive to reducing risks and can well meet the demand of sharing GA among functional departments.

References

[1] Sheshadri Chatterjee, Sreenivasulu N. S.: Personal DS and Legal Issues of Human Rights in the Era of Artificial Intelligence: Moderating Effect of Government Regulation. Int. J. Electron. Gov. Res.15(3): 21-36 (2019).

[2] JasonJ. Saleem, Nancy R. Wilck, John J Murphy, Jennifer Herout: Veteran and Staff Experience from a Pilot Program of Health Care System-Distributed Wearable Devices and DS. Appl. Clin. Inform. 13(3): 532-540 (2022).

[3] Youngseek Kim: Reputation, trust, and norms as mechanisms forming academic reciprocity in DS: an empirical test of theory of collective action. Aslib J. Inf. Manag. 74(6): 1174-1195 (2022).

[4] Manuel Rueda, Roberto Ariosa, Mauricio Moldes, Jordi Rambla: Beacon v2 Reference Implementation: a toolkit to enable federated sharing of genomic and phenotypic data. Bioinform. 38(19): 4656-4657 (2022).

[5] Pavel Vazquez , Kayoko Shoj, Steffen Novik, Stefan Krauss, Simon Rayner : Globally Accessible Distributed DS (GADDS): a decentralized FAIR platform to facilitate DS in the life sciences. Bioinform. 38(15): 3812-3817 (2022).

[6] Rabeeha Fazal, Munam Ali Shah, Hasan Ali Khattak, Hafiz Tayyab Rauf, Fadi M. AI-Turjman :Achieving data privacy for decision support systems in times of massive DS. Clust. Comput. 25(5): 3037-3049 (2022).

[7] Murat Tahir Caldag, Ebru Gokalp: The maturity of open GA maturity: a multivocal literature review. Aslib J. Inf. Manag.74(6): 1007-1030 (2022).

[8] Martin Lnenicka, Anastasija Nikiforova, Stuti Saxena, Purnima Singh : Investigation into the adoption of open GA among students: the behavioural intention-based comparative analysis of three countries. Aslib J. Inf. Manag.74(3): 549-567 (2022).

[9] Benedict Bender, Moreen Heine: Government as a Platform? The Power of Platforms to S upport Personalization of Public Services. J. Data Intell. 3(1): 169-187 (2022).

[10] Brendan T. L awson: Realizing the benefits of open GA: Journalists' coverage of the NHS winter crisis, 2016-17. Inf. Soc. 38(1): 25-35 (2022).

[11] Lindy-Anne Siebritz , Ahmad Desai, Serena Coetzee, Antony K. Cooper : Capacitating local governments for the digital earth vision: lessons learnt from the role of municipalities in the South African spatial data infrastructure. Int. J. Digit. Earth 14(12): 1897-1917 (2021).

[12] Majid H. Alsulami : Government Services Bus (GSB): Opportunity to Improve the Quality of Data Entry. Int. J. Semantic Web Inf. Syst. 17(3): 35-50 (2021).