

Research on Collaborative Data Governance Action Research Under the Integration of Colleges and Universities*

Jiankun Ma

School of Economics and Management
Beijing Jiaotong University
Beijing, China

Abstract—According to the theory of collaborative governance, this paper explores how to apply data governance to solve the problem of information resource sharing among organizations. Based on the action research of the Institute of Economics and Management of Beijing Jiaotong University, this paper proposes a method of collaborative data governance under the integration of institutions. It establishes a more effective sharing model through the collaboration between organizations, business and technology, significantly improving the efficiency of sharing information resources, data quality and clarifying the relationship between the rights and obligations of data providers and demanders.

Keywords—information resource sharing; collaborative data governance; action research; collaborative governance

I. INTRODUCTION

In the early stage of informationization, various departments within the organization established a large number of information systems based on their own business needs. After years of development, business data accumulated in the system has gradually increased. However, due to the lack of experience at the beginning of construction, all departments have their own data, and the standards are different [1]. This makes chimney systems exist in large numbers and serious heterogeneity problems. It is difficult to share information resources among systems, especially in big data era, the explosive growth of data volume makes the sharing of information resources more difficult, and it is difficult for the decision maker to obtain valuable data from the data. This paper takes the School of Economics and Management of Beijing Jiaotong University as an example to discuss how to implement more efficient information resource sharing through the data governance method from the bottom up and make the data assets play a value. Finally, realize the goals of data traceability, quality assurance, and clear relationship between rights and responsibilities. This article adopts a method of action research, theory-guided practice, and practice-rich theory, starting from the college level to conduct data governance work, and exploring the

two-level more efficient information resource sharing model in future colleges and universities.

The structure of this paper is as follows: The second part is the literature review, the third part is the research method, the fourth part is the action research program under collaborative data governance, and the fifth part is the summary and outlook.

II. LITERATURE REVIEW

The information resource sharing model has been developing along with the development of information resource sharing practices. The centralized model and distributed model are two types of information resource sharing models recognized by the academic community [2]. In the era of big data, there are more entities to be shared and the volume of data is increasing. The centralized model cannot manage all data in a unified manner. Based on this, the distributed model has received extensive attention from scholars. Daniella Meeker et al. pointed out that the distributed research network is a widely accepted architecture and that data is maintained locally by agencies and coordinated through a common infrastructure that shares practices and software [3].

Collaborative governance theory is an emerging interdisciplinary of the synergetic theory of natural sciences and the governance theory, so it is the supplement of the governance theory. Luning Liu and others studied the construction of government governance platforms in developing regions and proposed a cooperative e-government platform workflow modeling framework, including three scalable dimensions of requirements, methods, and models, which were contributed to inter-agency collaborative governance [4].

III. RESEARCH METHODS

Action research was first proposed by Kurt Lewin in 1944. It is a scientific method with a plan, steps, and reflection for the purpose of solving problems. This article adopts action research for the following reasons: First of all, data governance is often used as a method in practice, and there is relatively little research on relevant theories. Action

*Found Project: China Railway Corporation Science and Technology Research and Development Program
Project No.: 2017X006-B

research emphasizes the combination of theory and practice, and through interventions in the practice process to enrich theoretical knowledge, so that the theory and practice promote each other and improve each other. Secondly, researchers can discover new theoretical knowledge or research frameworks from the action research, and then apply it to other areas. The collaborative data governance methods discussed in this study can provide reference for organizations to implement distributed information sharing model. The process of action research in this article includes the process of identifying the spiral of the problem, action planning, action implementation, evaluation and reflection. Through the cycle of the action research cycle, discover problems, propose improvements, and continue to optimize until they reach the target state.

IV. ACTION RESEARCH PROCESS

A. Problems

Investigation of the current situation of informatization found the following problems: 1. The problem of serious

data quality: First of all, the data at the source had a variety of data quality levels, such as nonstandard data, inconsistent data, and data redundancy. Secondly, there is ambiguity in data, and technical personnel do not understand the meaning of business data and cannot effectively manage data. 2. Ambiguity of data rights: The college's existing dual-center architecture system has brought a lot of data to the college. However, there is no corresponding record of the origin and whereabouts of the data. Because the quality of the data itself is uneven, once the wrong data is available, it is difficult to find the person responsible for the data.

B. Planning

The researchers decide to establish a distributed information resource sharing model within the entire school to solve the above problems, ensure the quality of the data from the source, and adopt a peer-to-peer sharing method to achieve the goal of "one source and one source". The overall architecture diagram is shown in "Fig. 1".

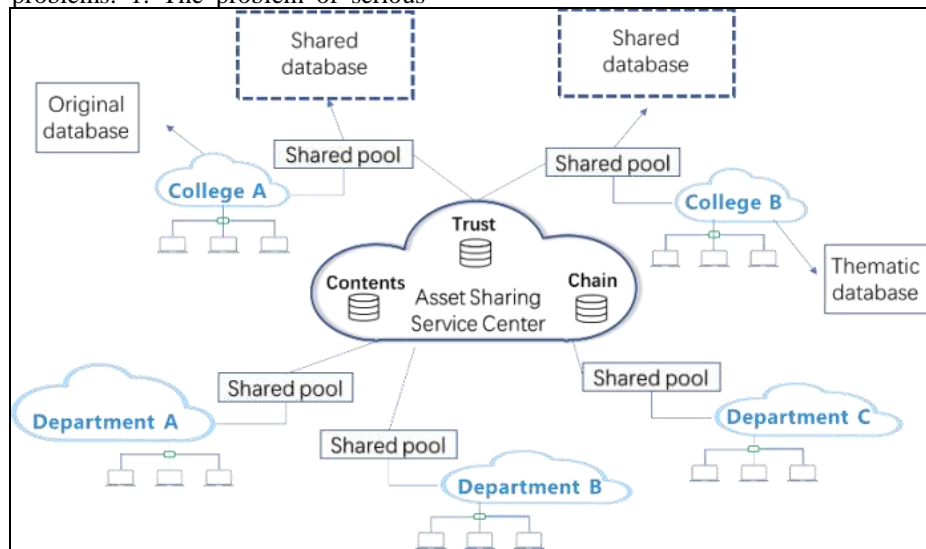


Fig 1. Collaborative data governance architecture.

Each school and Department of the school shares an "asset sharing service center". The center is maintained by the school information center and is used to store the details of the information resources of each unit. By querying the directory, each unit can find the storage of required data. Location, and then establish a shared relationship with the unit that owns the data, extract the required data from the unit's shared pool, and then store the extracted data in its own shared pool for subsequent use. In order to achieve the above distributed information resource sharing model, after discussion with the researcher, each unit stores the business

data after cleaning and stores it in the shared pool, which is divided into a supply shared library and a demand shared library. The supply shared library indicates that the unit can be shared externally. Data, demand shared library refers to the data that the unit uses from other shared pools. Each unit is responsible for the quality of the data in the shared pool. Moreover, in order to facilitate the application of data, the demand side really uses the value of the data assets and analyzes the data through a special library. The business operation flow for College A providing data for Department A is shown in "Fig. 2".

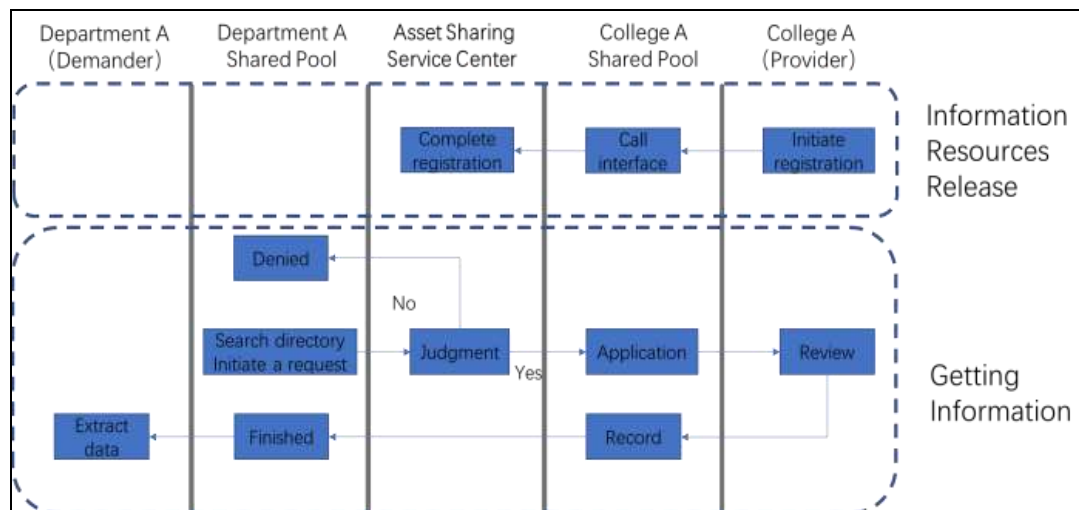


Fig 2. Information resource sharing process.

It includes "information resource release" and "information acquisition." When college A publishes its information, at first it initiates a registration application with the asset sharing service center, and then it invokes the registration interface in the Resource Sharing Pool of college A, and finally completes the registration of information in the asset sharing service center. When the Department B wants to obtain information, at first it initiates a request for search catalog to the asset sharing service center, and then the service center makes a judgment. If the search is unconditional, the request process ended. If it is conditional, a searching request is made to the Resource Sharing Pool of the provider (College A). Then the provider (College A) checks and records the request and transmits the data to the Resource Sharing Pool of the requester (Department B). Finally, the Department B extracts the data into its own database and the information acquisition process is completed.

C. Implementation

Researchers want to establish a school-wide distributed information resource sharing model through collaborative data governance methods, that is, collaboration between schools and colleges, and collaboration between business and technology. However, the school has not yet carried out related work, so the School of Economics and Management as a pilot unit simulated the entire sharing process through collaborative data governance methods, and finally promoted to the entire school. Through the method of collaborative data governance, the whole implementation process realizes

collaboration between schools and colleges and collaboration between business and technology from both organizational and technical aspects.

In terms of organization, the division of power and responsibility of providers and demanders of information resource sharing has been divided, as shown in "Table I".

In terms of technology, practitioners built a three-tier database at the School of Economics and Management: original database, shared database, and thematic database.

The original database stores all the data retrieved through the school and college interface, and the data imported manually in Excel and so on. The functions of the original database include: data classification, data source analysis, data cleaning, logging records, and dirty data processing. The shared database, as its name suggests, is a database that is created to facilitate the sharing of information resources. Therefore, the data stored in the shared database is clean data after the original database is cleaned. The rights and responsibilities are clear and the data quality is guaranteed. Thematic databases are business-oriented and database-specific. The ultimate goal of managing data assets is to make full use of the value of data assets. If you can't make good use of data and cannot integrate information resources well, then all the work done in the past is to increase costs. Therefore, the established thematic database is to use business requirements to display the standards-compliant data in the database in the form of various image reports and to support managers in making correct decisions.

TABLE I. ORGANIZATIONAL RIGHTS AND RESPONSIBILITIES

		<i>Original database</i>	<i>Shared database</i>	<i>Thematic database</i>	<i>Metadata Management System</i>
Provider	Business people	Import Data; Define cleaning rules...	Provides quality data required by demanders		Maintain business metadata; directory maintenance
	Technical staff	Data cleaning	Provide data interface		Maintenance Technology Metadata; Directory Maintenance
Demander	Business people	Define cleaning rules	Extracting data shared by providers	Report analysis	Maintain business metadata
	Technical staff	Execute data cleaning according to cleaning rules		Auxiliary business staff to do the report	Maintaining technical metadata

D. Evaluation and Reflection

Through the method of collaborative data governance, the problems under the existing information resource sharing model are solved: in terms of data quality, the quality is guaranteed from the source, and the semantic problem is solved; In terms of powers and responsibilities, the rights and responsibilities of all parties are divided. The source of each data in the metadata management system also has a corresponding record. Once a problem arises in the data, it can quickly find the person responsible for the data.

However, this round of action research is only a pilot project at the School of Economics and Management. It has not been truly applied within the scope of the entire university. In the future, a new round of data governance research is needed to extend this method to the entire university.

V. CONCLUSION

This study takes the School of Economics and Management of Beijing Jiaotong University as an example. In order to solve the problem of information resource sharing, the collaborative data governance method under the integration of institutions is proposed to improve the ease of use and efficiency of the shared model. After the action research, it finally reached the goal of the expected conception, the efficiency of information resources sharing was improved, the quality of data was greatly improved, the relationship between the rights and responsibility of data providers and demanders was clearly defined, and the ecological relationship of collaborative data governance was gradually established.

However, this study also has many deficiencies and requires further research in the future. On the one hand, this paper only simulates the distributed information resource sharing model of the whole school through collaborative data governance. It has not been truly promoted. What kind of problems will be encountered in the future will require further research. On the other hand, for the theory of collaborative governance in data governance, there are few studies in the industry, and there are not many documents in this area. This article is only a simple exploration in practice, and further research is needed in the future.

REFERENCES

- [1] Huang Ning. "A Preliminary Study on the Integration Model of Sharing Information Resources." *Information Technology* 31.8(2007):150-153.
- [2] Ma Feicheng, Pei Lei. "Progress in China's Practice and Theoretical Research on Information Resource Sharing. *Journal of Intelligence* 24.3(2005):277-285.
- [3] [3] Meeker, Daniella, et al. "A system to build distributed multivariate models and manage disparate data sharing policies: implementation in the scalable national network for effectiveness research." *Journal of the American Medical Informatics Association* 22.6(2015):1187-1195.
- [4] [4] Liu, Luning, J. Ju, and Y. Feng. "An extensible framework for collaborative e-governance platform workflow modeling using data