Comparative Analysis of Data Governance in Domestic and International Universities Based on Cite Space

Boya Yang[1#], Junhao Liang[2#], Zhiping Li[3,*]

[1]School of Information Science and Technology (SIST) Yunnan Normal University, Yunnan, Kunming, 650000, China
[2]Fogang Eastern Junior Middle School, Guangdong, Qingyuan, 511600, China
[3]Network and information center Yunnan Normal University, Yunnan, Kunming, 650000, China
[#]Yang Boya and Liang Junhao are co-first authors of the article.

*1966@ynnu.edu.cn

Abstract. With the rapid development of information technology in recent years, colleges and universities have accumulated a large amount of data resources, and data governance has gradually developed into a hot research field in colleges and universities. Enhancing data quality with the help of data governance is an inevitable choice for universities to realize high-quality development. Based on the data governance literature of colleges and universities in WOS and CNKI databases, this paper compares and analyzes the status quo, hot topics, and evolution paths of data governance research in colleges and universities at home and abroad by using Cite Space software. It is found that: foreign university data governance mainly focuses on creating organizational structure and decision-making mechanism, and domestic theoretical system research starts late, focusing on infrastructure construction to application deepening. Therefore, in the future development of data governance in domestic universities, it is necessary to deepen the understanding of the importance of data governance and build a perfect data governance system.

Keywords: Data Governance, Universities, Cite Space.

1 Introduction

Definition given by the Data Governance Institute International (DGI): Data governance is a system of decision-making authority and segregation of duties through a series of information-related processes executed according to a consensual model that describes who can take what action, on the basis of what information, at what time and in what circumstances, and in what manner[1].

In April 2018, the Ministry of Education (MOE) released the Education Informatization 2.0 Action Plan "striving to build a new model of education services, a new model of talent cultivation, and a new model of education governance." The Data Security Law of the People's Republic of China requires "international exchanges and

Y. Feng et al. (eds.), *Proceedings of the 4th International Conference on Internet, Education and Information Technology (IEIT 2024)*, Atlantis Highlights in Social Sciences, Education and Humanities 26,
https://doi.org/10.2991/978-94-6463-574-4_66

cooperation in the fields of data security governance, data development and utilization." Big data is a new stage in the development of information technology, and we should be good at acquiring data, analyzing data, applying data, and enhancing the ability to use data to promote all work, so that big data can play a greater role in all work[2]. Strengthen data control through data governance to ensure data quality and fully explore the value of data[3].In the context of "Internet +", big data, the construction of data governance in colleges and universities is a trend of network security in the new era, and a means of deep integration of informatization and modernization of education. The use of informatization means to provide teachers and students with better data support[4].

2 Data Sources and Research Methodology

2.1 Data Sources

This paper utilizes CNKI and WOS databases to carry out literature search in the field of data governance in colleges and universities, and to review the research results of data governance in colleges and universities in the past two decades since its emergence. Domestic data are selected from the CNKI database, the search mode is "subject", and the search keywords are ("data governance", "universities", "big data governance"); the type of literature is "data governance", and the type of literature is "big data governance", "big data governance"); the type of literature is academic journals; foreign data is selected from WOS core collection database, and the search strategy is theme = ("university" AND "data governance"). governance") will retrieve the "full record and cited references" to "plain text" format after exporting the same to "download_ + file name The name of the file is also "download_+filename". Under the condition of subject area restriction, the time span of the search is January 1, 2004-October 1, 2023, and the valid literature information is 232 and 522 respectively, filtering out the duplicated literature and irrelevant information, and finally obtaining 183 and 433 (616 in total) literature data. The data were imported into Note Express literature management software and named as "download_+filename" after exporting.

2.2 Research Methodology

This article adopts the bibliometric method, using Cite Space visual analysis software as the analysis tool, importing the exported download file for data number format transformation and data analysis. Years Per Slice is set to 1 year; Node Types is selected as keyword. Selection criteria: choose g-index, k=10, and the rest are default values. The keywords are pruned using graph pruning algorithm. The value of keyword clustering module (Q value) > 0.3 indicates that the clustering structure is significant, and the average contour value (S value) > 0.5 suggests that the clustering is reasonable[5]. On the one hand, quantitative bibliometric methods are used to generate information on the number of documents, highly cited authors, and other information of data governance research in universities, and at the same time analyze the co-occurrence of research authors and research institutions, keyword co-occurrence, clustering, and highlighting

maps to reveal the research hotspots; on the other hand, qualitative manual interpretation methods are used to study the disciplinary information and keyword clustering in-depth on the basis of the data mapping, to rectify the bias of the metrological analysis, and to carry out a critical qualitative analysis, with a view to revealing the current state of research in the field of data governance in universities.

3 Current Status of Data Governanein Universities and Colleges in China and Abroad

The analysis of the current status of data governance in universities and colleges at home and abroad is proposed to be carried out from the following aspects.

3.1 Analysis of Annual Number of Publications

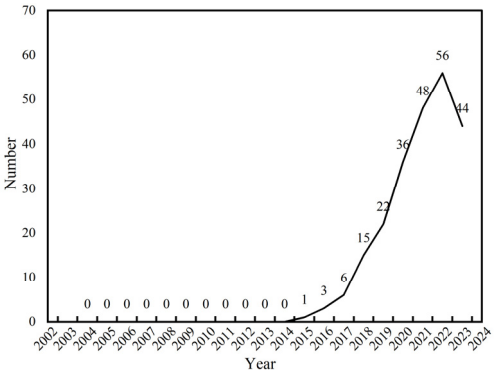

Fig. 1. Annual publication volume of domestic documents

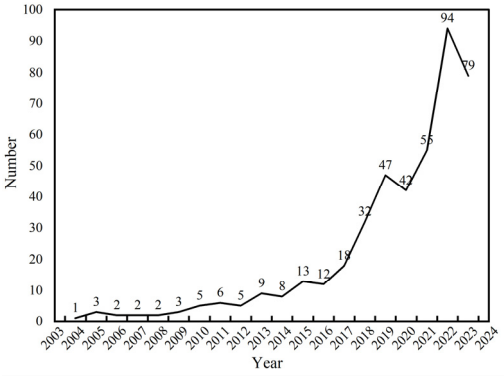

Fig. 2. Annual publication volume of foreign documents

In order to understand the research results of data governance in colleges and universities, this study counts the number of papers published domestically and abroad from 2004 to 2023. Figure 1 shows that domestic research started later than foreign research. The number of domestic publications is in a fluctuating upward state from 2015-2023 and peaks in 2022. From the Figure 2 we can find that the number of foreign publications is in a steady state from 2004-2014, and the number of annual publications in 2015-2019 is multiplied, and after a brief decline in 2020, it continues to rise in 2020-2022 (except for 2014).

3.2 Analysis of Key Authors and Institutions

In the field of data governance research in higher education, the amount of foreign literature is greater than the amount of domestic literature, in the 433 valid literature data from abroad, the authors of the study and the institutions work closely with each other, the amount of English literature issued by the University of London and University College London are more institutions, respectively, 16 and 11 articles. Domestic this research started late, the amount of literature data is small and literature authors and institutions are relatively not close, Chinese literature issued by Liu Guifeng (4), Qian Jinlin (3), Wu Gang (3), Liu Yaqin (3) more to Jiangsu University Institute of Science and Technology Information (3) institutions, overall, domestic and foreign universities in the field of data governance research have not yet formed a high impact of the core authors and research teams. See Table 1 and Table 2.

Table 1. Important foreign authors and institutions

Count	Centrality	Year	Authors | Institutions
16	0.01	2008	University of London
11	0.00	2014	University College London
7	0.00	2020	University of Edinburgh
6	0.01	2007	Monash University
6	0.00	2011	Stockholm University
5	0.02	2007	University of Sydney
5	0.00	2011	University of Queensland
3	0.00	2013	Michigan State University
3	0.00	2022	Linnaeus University
3	0.00	2007	University of Plymouth

Table 2. Important domestic authors and institutions

Count	Centrality	Year	Authors | Institutions
4	0.00	2017	Liu Guifeng
3	0.00	2017	Institute of Science and Technology Information, Jiangsu University
3	0.00	2017	Qian Jinlin
3	0.00	2018	Publicity Department of the Party Committee of Yancheng Polytechnic University
3	0.00	2018	Wu Gang
3	0.00	2021	Network and Computing Center, Huazhong University of Science and Technology
3	0.00	2021	Liu Yaqin
2	0.00	2016	Shanghai Ocean University Modern Information and Education Technology Center
2	0.00	2017	Informatization Office of Tongji University
2	0.00	2017	College of Electronic and Information Engineering, Tongji University

3.3 Keyword Co-occurrence

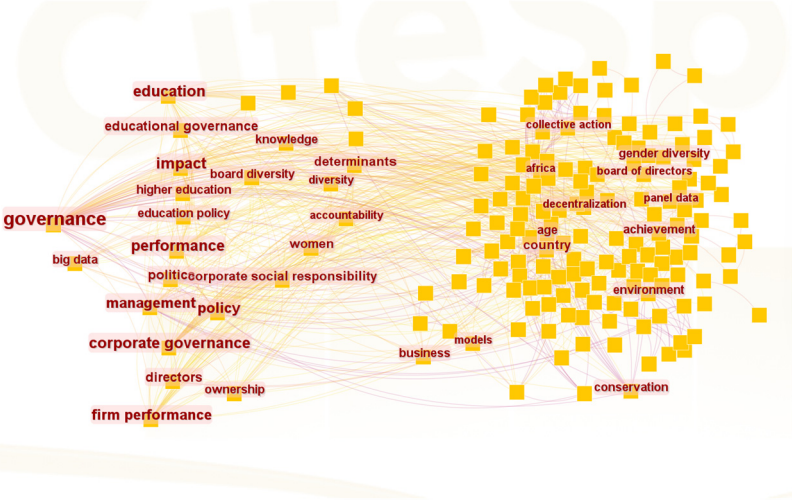

Fig. 3. Co-occurrence of foreign keywords

Using Cite Space software, set the time from January 2004 to October 2023, time slice 1 year, set the network nodes as keywords, import the literature in WOS core ensemble and CNKI database respectively, and conduct centrality analysis, the foreign WOS core ensemble literature ultimately generates 310 nodes, 430 connecting lines, and the network density of 0.0079 of the Keyword co-occurrence map, the literature in the

domestic CNKI database finally generated 328 nodes, 463 connecting lines, network density of 0.0103 keyword co-occurrence map. The results can be seen that in the foreign literature, the top 10 keywords of hot centrality related to data governance in universities are: governance, education, educational governance, impact, higher education, performance, policy, management, corporate governance. In the national literature, the hot keywords related to data governance in universities are data governance, big data, data standards, data management, smart campus, education, data quality, university, university governance, data center. See Figure 3 and Figure 4.

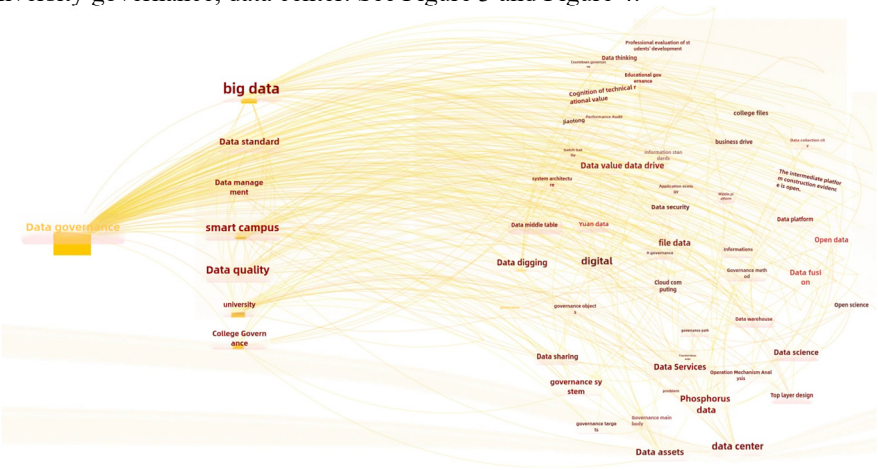

Fig. 4. Domestic keyword co-occurrence

4 Comparative Analysis of Hot Topics in Data Governance Research in Higher Education Institutions

4.1 Foreign Research Hot Spots

High-frequency words are words that appear more frequently in the literature and can be used as one of the references for the key points and focus of the study[6]. It often reflects the research theme of the literature. In this paper, we use the word frequency analysis function of Cite Space software to get the high-frequency keywords of foreign research, such as Table 3, the data show that the high-frequency keywords in foreign research are School governance, Board diversity, Devolution, Bigdata, Environment, Forests and so on, which It shows that the policy impact and technology application research carried out in combination with university data governance is a more theoretical content of foreign research, and the nodes with mediator centrality greater than 0.1 in the relevant nodes in Cite Space are called key nodes[7]. The chart shows that SCHOOL GOVERNANCE is the most representative key node in foreign research, with a mediational centrality of 1.30, which indicates that foreign research on data governance in higher education mainly centers on university management.

Table 3. High-frequency words in foreign research hotspots

No.	Key word	Times	Central-ity	No.	Key word	Times	Centrality
1	School governance	169	1.30	11	Performance	11	0.00
2	Board diversity	36	0.56	12	Management	11	0.01
3	Devolution	34	0.67	13	Directors	10	0.00
4	Big data	31	0.43	14	Firm performance	10	0.00
5	Environment	29	0.30	15	Knowledge	9	0.00
6	Forests	21	0.28	16	Policy	9	0.00
7	Corporate	16	0.09	17	Ownership	7	0.00
8	Education	16	0.01	18	Country	5	0.00
9	lmpact	13	0.01	19	Business	5	0.00
10	Higher education	13	0.01	20	Age	4	0.00

Keyword clustering can visualize the distribution of topics in a domain[8]. In order to further clarify the thematic distribution of data governance research in foreign universities, the keyword clustering function of the software was utilized, and the LLR algorithm was used to extract the cluster identifiers to obtain the foreign keyword clustering view (Fig. 5), which resulted in a total of seven major clusters[5]. The clustering weighted average profile value S>0.7 means the clustering result is exact, the Q value in this clustering atlas is 0.7103 and the S value is 0.85, which indicates that the internal structure of this clustering is closely connected and the clustering result has a high confidence value. The clustering map shows that foreign research topics are mainly distributed in the fields of #0 board diversity, #1 school governance, #2 devolution, #3 big data, #4 environment, #5 forests, #6 local government, #7 corporate strategy, etc.

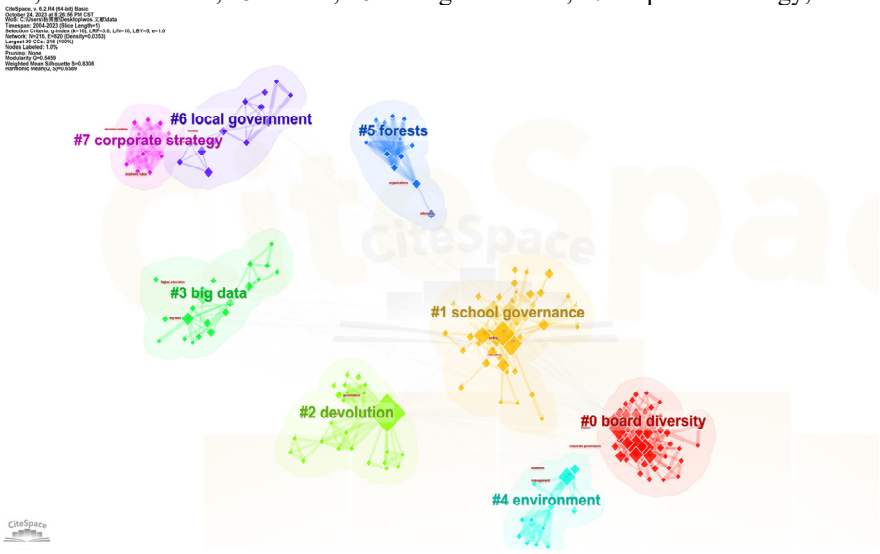

Fig. 5. Keyword clustering map of WOS source documents

4.2 Domestic Research Hot Spots

The data show that the high-frequency keywords in the domestic research are data analysis, big data, data center, data center, research data, data quality, etc., which indicates that the infrastructure construction and technology application research carried out in conjunction with the data governance of colleges and universities is the content of more domestic research, and nodes with intermediary centrality of relevant nodes in Cite Space greater than 0.1 are called key nodes[7].

Table 4. High-frequency words, a hot topic in domestic research

No.	Key word	Times	Centrality	No.	Key word	Times	Centrality
1	Data governance	147	1.29	11	data center	5	0.01
2	Big data	40	0.29	12	data sharing	5	0.00
3	Wisdom campus	36	0.11	13	Data assets	5	0.00
4	University	23	0.21	14	Data security	4	0.00
5	Data quality	21	0.04	15	University governance	4	0.05
6	data standard	11	0.04	16	data warehouse	4	0.00
7	Data service	7	0.01	17	informatization	4	0.00
8	Scientific research data	6	0.05	18	Open science	3	0.03
9	Data platform	6	0.00	19	Information standard	2	0.00
10	Data management	5	0.04	20	master data	2	0.00

Table 4 shows that data governance and big data are the most representative key nodes in domestic research, with mediated centrality of 1.29 and 0.29, which indicates that data governance research in domestic colleges and universities mainly focuses on infrastructure and practical application.

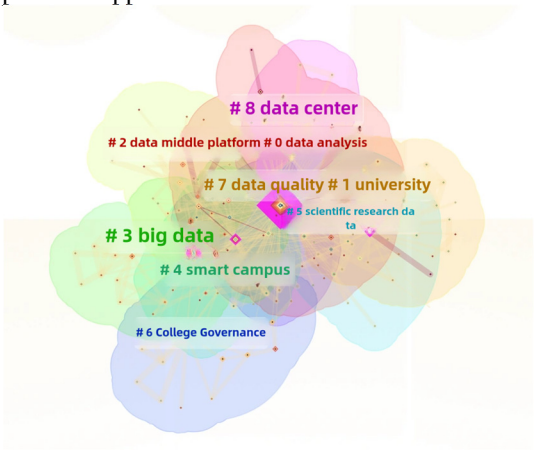

Fig. 6. Keyword clustering map of CNKI source documents

The clustering operation on the high-frequency keywords obtains the domestic high-frequency keyword clustering map (e.g., Figure 6), which forms a total of nine major

clusters, with a Q-value of 0.554, indicating that the structure of this cluster is obvious, and an S-value of 0.864, suggesting that the results of this clustering are credible[5]. This clustering map shows that the domestic research topics are mainly distributed in the areas of #0 Data Analytics, #1 Higher Education, #2 Data Middleware, #3 Big Data, #4 Intelligent Campus, #5 Research Data, #6 Higher Education Governance, #7 Data Quality, and #8 Data Center.

5 An Evolutionary Path for Data Governance Research in Higher Education

5.1 Evolutionary Paths Abroad

Cite Space software's TimeLine View can show the development path and inner connection between the clustered keywords from the time dimension, and the result can reflect the development of the research topic, research frontiers, etc.[9]. Using this software to draw a timeline graph of foreign keyword clustering (e.g., Figure 7), where the size of the nodes represents the heat of the keywords, and the connecting lines between the keywords represent the evolutionary path of the research content[10]. Its burst monitoring algorithm is utilized to identify the turning points of data governance in foreign colleges and universities and the research content in different stages (e.g., Figure 7). Combined with the graph of the volume of foreign research publications and the timeline graph, the research on data governance in foreign colleges and universities is categorized into the following 3 stages by manually combing the relevant literature.

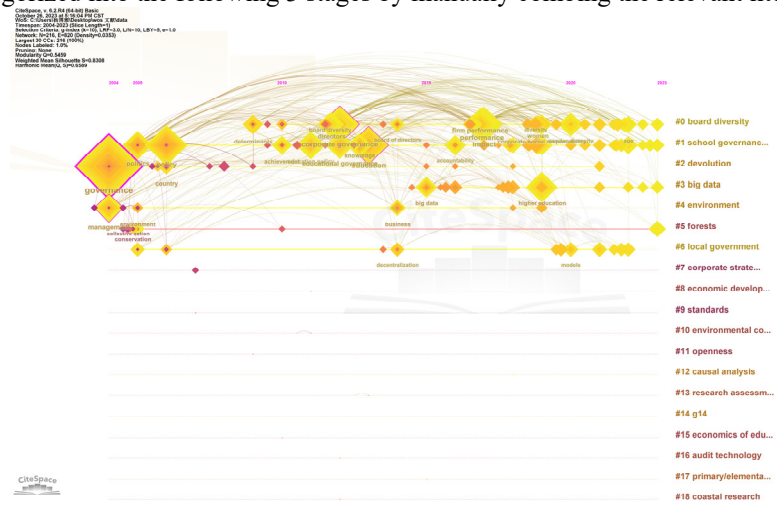

Fig. 7. Evolution Path of Data Governance in Foreign Universities

Incipient stage: policy theory formulation (2004-2009)

The proposed stage of policy theory is mainly accompanied by the continuous development of data science, information management and organization theory and

gradually formed. The research themes in this stage mainly focus on governance, management, policy and so on. In the process of proposing the theory of data governance management system, foreign colleges and universities draw on the successful experience of enterprise data governance. So that colleges and universities can more quickly establish a data governance system that meets their own characteristics.

Development phase: policy operationalization (2010-2014)

With the continuous progress of big data, cloud computing, artificial intelligence and other technologies, the research topics in this phase are focused on the big data, determinant, direct and so on. MIT scholars launched the Total Data Quality Management (TDQM) program in 1988, which aims to help companies improve data quality and optimize teaching and management through data analytics by establishing a solid theoretical foundation and required tools in the field of data quality management. It focuses on cross-departmental collaboration and communication, promotes information sharing and cooperation among different departments, breaks down information silos, and ensures the accuracy and consistency of data.

Stabilization phase: applied practice in higher education (2015-2023)

At this stage, data governance in colleges and universities has formed a relatively mature and stable system. The concepts and practices of data governance have penetrated into all levels of colleges and universities and become an important support for the development of colleges and universities. The research themes at this stage are focused on the study of HIGH EDUCATION, FIRM PERFORMANCE, MODELS and so on. It focuses on the continuous optimization and innovation of data governance, constantly improves the system and process of data governance, strengthens the monitoring and management of data quality, and actively explores new data application modes.

5.2 Domestic Evolutionary Path

Using the same method to draw the timeline map of keyword clustering in China (as shown in Figure 8), combined with the publication volume of foreign research hotspots, the domestic research on data governance in universities from 2004 to 2023 is divided into the following three stages according to the year:

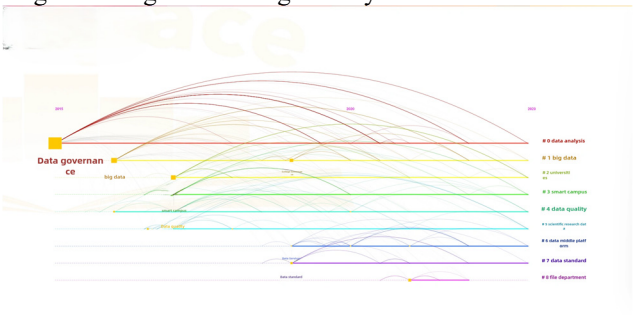

Fig. 8. Evolution path of data governance in domestic universities

Initial phase: technology development and foundation building (2004-2016)

China's research in this field is now mainly focused on practical applications, this stage has experienced a long development process, the development of information technology in the early stage is to better cope with the challenges of data, and to provide adequate technical support for data governance in universities and colleges. Du Xiaoyong et al. in "Data Organization-Key Technology for Big Data Governance[11]. The article analyzes the key technologies of data organization and emphasizes their impact on the effectiveness of data governance. Other scholars have also conducted in-depth studies on the technical level of data governance, covering various aspects of data governance, including data quality control, data security and privacy protection, data mining and analysis.

Construction phase: standard-setting and quality monitoring (2017-2020)

There are few scholars in the domestic research, Dai Hong, Zhang Qun, Yin Zhuo and others from the China National Institute of Electronic Technology Standardization (CNIETS), who in their article Research on Big Data Governance Standard System, the construction and development of big data governance standard system is discussed[12]. Prof. Qingguo Meng and Teng Zhang et al. proposed a data governance system centered on three levels: institutional, technological, and market. And they also have unique insights and contributions in standardization and quality supervision.

Stabilization phase: smart applications and business synergies (2021-2023)

This stage establishes a unified data management platform to integrate and standardize the data generated by various business systems within the university. Constructing an intelligent decision support system to help universities make smarter decisions in strategic planning, resource allocation, risk management, etc.

6 Conclusions and Implications of the Study

6.1 Conclusions of the Study

In terms of organizational structure and decision-making mechanism, the data governance systems of foreign universities tend to be more mature and systematic. For example, the organizational structure of data governance in some foreign universities is divided into three levels: decision-making, coordination and implementation, forming a complete closed-loop system. The data management officer has a central position in the data governance system, with responsibilities spanning multiple levels and coordinating the relationship between different departments and areas.

In terms of technology application and tool selection, the country can be summarized as a process from infrastructure construction to application deepening. At the initial stage, colleges and universities mainly focus on the construction of data infrastructure, such as the establishment of data centers and the improvement of data networks, to provide basic protection for data governance. With the continuous progress of technology, colleges and universities began to focus on the collection, integration and sharing of data to promote the effective use of data resources.

6.2 Research Implications

Although data governance research in domestic colleges and universities has achieved promising results in recent years, foreign research still has important reference and reference value for domestic related research. In the future, relevant domestic research should be based on practice combined with cutting-edge intelligent technology to conduct prospective exploration, expand the scope of research, deepen the research on data governance segments, and promote the transformation of theoretical research into the value of practical development. Summarizing the results of the comparative analysis of domestic and foreign studies, the following insights are obtained: domestic university data governance in the future development needs to continuously deepen the understanding of the importance of data governance, build a perfect data governance system, focus on technological innovation and application, and strengthen the data security and privacy protection, and through the implementation of these measures, the level of data governance in colleges and universities can be further improved to provide strong support for the sustainable development of the university.

References

1. Anonymous. DGI data governance framework[EB/OL]([no date])[2024–04–07]. https: // dgiframework.com/.
2. CHENG HONGYI. Xi jinping emphasized at the second collective study of the political bureau of the communist party of china central committee (cpc central committee) that the time and situation are ahead of schedule, elaborate planning and layout, and strive to take the initiative to implement the national[J]. Practice (Theory of Thought Edition), 2017(12): 7.
3. CAI CUIHONG, WANG YUANZHI. Global data governance: challenges and responses[J]. Research on international Issues, 2020(6): 38–56.
4. YANG QIANG. Analysis and practice of the current situation of internal data governance in colleges and universities[J]. Modern Information Technology, 2023, 7(12): 120–123, 127. DOI:10.19850/j.cnki.2096-4706.2023.12.030.
5. CHEN YUE, CHEN CHAOMEI, LIU ZEYUAN, et al. The methodology function of citespace mapping knowledge domains[J]. Studies in Science of Science, 2015, 33(2): 242–253. DOI:10.16192/j.cnki.1003-2053.2015.02.009.
6. HUANG HAIYAN, MI YUANYUAN, GUO HUILING. Bibliometric analysis of blood glucose management in elderly patients with severe disease in china[J]. Nurs J Chin PLA, 2020, 37(1): 56–59.
7. YANG LI, WEI XIAO, GAO ZIHAN. Knowledge graph visualized analysis of osce in the application of medical education in china[J]. CHINA MEDICAL EDUCATION TECHNOLOGY, 2020, 34(5): 586–591. DOI:10.13566/j.cnki.cmet.cn61-1317/ g4. 20 2 005012.
8. FENG YIJIE, SHAO BO. Development plan and implementation strategy for library smart spaces[J]. Library Forum, 2023, 43(1): 71–77.
9. PAN LI, LIU JIANRU, HOU JIANHUA. Hotspot domains and frontiers of international situational learning[J]. Comparative Education Review, 2014, 36(6): 42–46. DOI: 10. 20013/ j.cnki.ice.2014.06.009.

10. HOU JIANHUA, HU ZHIGANG. Review on the application of cite space at home and abroad[J]. Journal of Modern Information, 2013, 33(4): 99–103.
11. DU XIAOYONG, CHEN YUEGUO, FAN JU, et al. Data wrangling: a key technique of data governance[J]. BIG DATA RESEARCH, 2019, 5(3): 13–22.
12. DAI Hong, ZHANG Qun, YIN Zhuo. Study on big data governance standard system [J]. BIG DATA RESEARCH, 2019, 5(3): 47–54.