

Visual Measurement Analysis of Domestic Data Governance Research based on CiteSpace

Qi Zhu

School of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

Abstract: Data, as a valuable resource, occupies a more and more important position, and data governance has gradually become a research hotspot. This paper uses the bibliometric software citespace to study 1130 documents selected from the CNKI journal database in 2010-2020, and makes a comprehensive analysis of the data governance from several aspects from the number of the number of articles, the distribution of authors, publishing institutions, research hotspots and frontier trends. By studying the current situation of data governance research in China, this paper is helpful to understand the development trend, explore hot issues, make up for the defects and deficiencies in the research, and provide reference for promoting the theoretical research and practical application research of data governance research in China.

Keywords: Data governance, Citespace knowledge graph, Bibliometry.

1. Introduction

With the advent of the era of big data, the data forming all kinds of information occupies an increasingly important role in organizations and institutions." The view that data is a valuable resource" has increasingly become an academic consensus [1]. Data governance has also gradually become the focus of organizational and academic research. Data governance began to emerge with enterprise-wide regulations [2]. In the early studies, data governance was the governance between people and technology [3]. With the continuous enrichment of research, data governance is considered an organizational design task, and is organizational governance based on data [4]. The existing research themes on data governance research vary. For example, Liu Guifeng and Qian Jinlin (2020) pay attention to university data governance and put forward the construction of data governance model in universities in China [5]. Xia Yikun (2020) pays attention to the dimensional analysis and path optimization of the government data governance, and proposes that it must promote the healthy development of the data factor market and deepen the reform of the government data governance mode [6] [7]. An Xiaomi and Guo Mingjun et al. (2019) organically integrated the macro, medium and micro system components of data governance, and put forward the big data governance system framework of multi-dimensional factor connection and win-win results and its effective path of realization [8]. Liu Yiming (2020) builds a VCA model of publishing enterprise data governance based on the value chain analysis model and the practical needs of data governance of publishing enterprises [9]. Han Xu (2018) [10] And Gu Bin (2017) [11] Focus on the quality and management of the data assets themselves, etc.

To sum up, China's research on data governance needs to be further developed, clarify the problems and grasp the research status and development trend, and carry out more professional research on the basis of the existing research results. This paper uses the citespace software developed by Professor Chen Chaomei to visually analyze the relevant domestic literature, dig out the frontier of the discipline from a large number of literature and find the hot spots in the research field, laying a foundation for deeper research in the

future.

2. Data Sources and Research Methods

2.1. Data Source

This study used the CNKI (CNKI) journal database as the data source. In the advanced search selection of keywords include "data governance", the time is set from 2010 to 2020, the time span is nine years to search, the search time is July 7,2020, a total of 1130 documents were obtained. The selected literature was exported in Refworks format, and the obtained data included the title, author, keywords, abstract and corresponding citation information, etc. The data search deadline is July 7,2020.

2.2. Research Methods

This paper uses the citespace information visualization software, a tool adopted by scientometry. Citespace is a kind of citation analysis, the so-called citation analysis is the use of various mathematical and statistical methods and comparison, induction, abstract, generalization, the scientific journals, papers, authors and other analysis of the object cited and cited phenomenon, in order to reveal the quantitative characteristics and internal laws of a bibliometric analysis method [12]. Using citespace software to obtain an intuitive and clear knowledge map has overcome the defects that previous literature reviews may have subjective factors influence.

By using citespace software, the literature data from 2010-2020 were analyzed and processed, the collected data was processed, the time slice was set to 1, the node was set as the research institution, authors and keywords, and a series of knowledge maps were obtained.

3. Analysis of the Visual Results

3.1. Statistics of the Chronological Distribution of the Published Volume

The functional function of China Knowledge Network (CNKI) journal database was used to analyze the chronological distribution of published literature and obtain the general trend chart (Figure 1).The relevant literature on

data governance generally showed a continuous growth trend from 2010 to 2020. In addition, the distribution of relevant documents can be roughly divided into two stages: the first stage, the germination and exploration period of 2010-2015, the number of relevant documents is small and the growth trend is not obvious. The second stage, from 2015 to now, from 47 in 2015 to 326 in 2019, predicts that about 384 related articles will appear in 2020. The reason for this is that in the 21st century, people began to pay attention to the role of data. In 2012, the word big data is more and more mentioned. In 2012, the Chinese government put forward the Big Data Research and Development Plan in the United States and approved the "12th Five-Year National Government Information Construction Project Plan". China's era of

openness, sharing and intelligent big data really began in a large area. In 2015, the service control group of ITSS Branch of CECTA, on behalf of China, proposed the "International Standard of Data Governance" to ISO / IEC JTC1 / SC40, and was approved and carried out the development of national standards of data governance norms[13]. The continuous growth trend of relevant literature also shows that the number of data governance research publications is related to the importance of data governance in China. At the same time, influenced by the continuous deepening and refinement of the research theory, the research methods, more flexible research means are more diversified in studying the problems related to data governance, so the number of publications is also increasing.

总体趋势分析

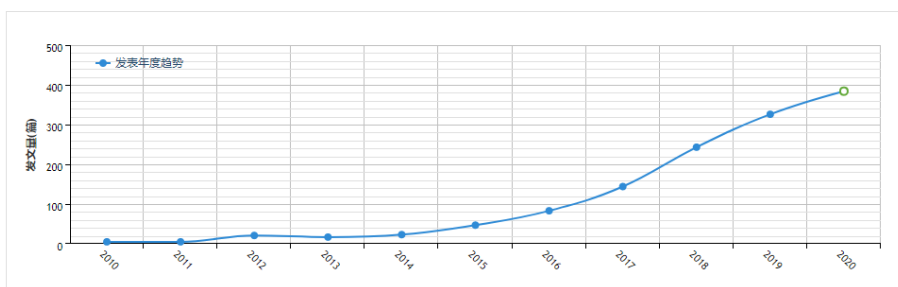


Figure 1. General Trend Analysis

3.2. Statistics of Document-Issuing Institutions

Citespace software was used to set the node type as the issuing institution, the time slice length is 1, the data with nine years' time span from 2010 to 2020 was extracted, and the institutions with two or more publications were selected to obtain the visual map of the research institutions studying the problems related to data governance (Figure 2). According to the map, N=218, E=237, and the density is 0.006, indicating that most of the 1130 documents retrieved are from 218 research institutions, but there is less cooperation between the research institutions. The resulting map is shown in Figure 2. The results obtained from citespace can select the five institutions with the most publishing institutions to Table 1.

As can be found from Table 1, the five main institutions to publish data governance related literature are the School of

Information Resources Management of Renmin University of China, Information Technology Department of China Everbright Bank, China Institute of Electronic Technology Standardization, Information Resources Research Center of Wuhan University and China Academy of Communications. Specifically, most of the five major publishing institutions are scientific research institutions, and only two are university information resources research institutions. According to the knowledge graph, the connection between research institutions is loose. However, the lack of communication of research results is easy to lead to lagging research development. At present, the exchanges and cooperation of research institutions can be strengthened to promote the continuous development and progress of data governance research.



Figure 2. Visual map of core institutions of Data Governance Research

Table 1. Main issuing agencies

Ranking	Frequency	Document organization	Proportion (%)
1	18	School of Information and Resource Management, Renmin University of China	1.60
2	15	China Everbright Bank Information Science and Technology Department	1.33
3	11	China Institute of Electronic Technology Standardization	0.97
4	11	The Information Resources Research Center of Wuhan University	0.97
5	9	China Academy of Communications Research Institute	0.80

3.3. Distribution Statistics of Published Authors

Citespace software was used to set the node type as the publishing organization, the time slice length is 1, the data with the time span of 2010 to 2020 for nine years were extracted, and the authors with three or more publications were selected to obtain the visual map of the authors studying the problems related to data governance. According to the map, N=348, E=306 and density are 0.0051, indicating that most of the 1130 documents retrieved were published from 348 authors but relatively independent cooperation between authors. The resulting map is shown in Figure 3.

On the whole, anomi has published data governance related articles in recent years, which published the framework of government big data governance system and its effective path for implementation (2019.2) and the research concept for the construction of government big data governance rule system (2018.5). Secondly, Liu Guifeng published research on the construction of scientific research data governance model in Chinese universities (2020.6) and research progress of data

governance at home and abroad: connotation, elements, model and framework (2017.12). By analyzing the distribution data of the research authors related to data governance, we can find that most of the existing studies on data governance focus on the construction of government data governance research and data governance research model, which reflects the lack of rich data governance research to some extent. In addition, it is known that although the research on data governance is gradually increasing, the authors studying issues related to data governance are in a minority compared with other studies. This shows that the domestic research on data governance is not mature and the attention is not high. As is also known from the atlas, the authors are more independent, and have less cooperation between the authors and the authors. High-level paper publication requires in-depth cooperation. Therefore, based on the current research situation, researchers should start from reality, seek cooperative joint organizations or cooperative institutions to strengthen relevant research and expand the research team.



Figure 3. Cooperative visualization atlas of the Core authors of Data Governance Research

3.4. Keyword Analysis

Citespace software was used to set the node type as keywords, the time slice length is 1, the data with the time span of 2010 to 2020 was extracted, the keywords appeared five times or more, and then the keyword word frequency distribution table and the keyword collinear map were obtained. shown in the keyword collinear map are N=511, E=1861, a density of 0.0143 and a modular Q value of 0.6335 (> 0.3) demonstrate that the clustering results are significant.

The contour value MS=0.5217 (> 0.5) proves that the clustering is significant and reasonable. The collinear graph of the keywords is obtained (Figure 4). The top ten keywords were obtained in citespace software in the keyword word frequency distribution, and the keyword word frequency distribution table was obtained (Table 2).

Through the frequency of keywords, you can withdraw the research hotspot. The higher the frequency of keywords, the more likely it is to be the research hotspot. The frequency of the citespace is reflected in the strength of the information

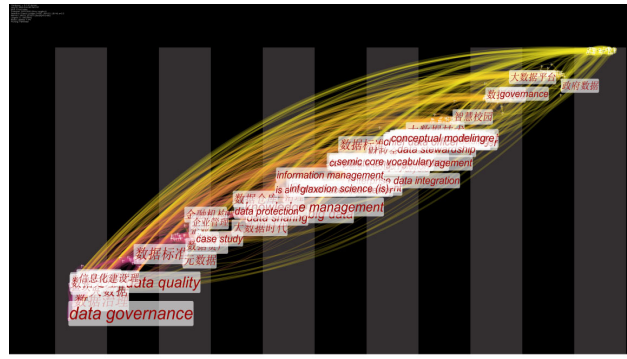


Figure 5. Time diagram of migration of data governance keywords

4. Conclusion

This paper is the first CiteSpace tool on the 2010-2020 domestic data governance research literature visual network, respectively from the general trend of the literature, the main organization, the main author, and keyword default view, keyword migration timing diagram and keyword present analysis, understand the development trend of the research, and determine the field with prospective and potential research direction. These findings can lay the foundation for the theoretical improvement and practical application in the field of data governance, and provide a useful reference for further deepening and expanding research in the future.

Combined with the visual map, it can be found that the research related to data governance has been in a growing trend, which has a relatively obvious positive relationship with the attention to data governance in China. However, there are still some problems: (1) insufficient core authors, the research team and core author group have not been formed. (2) There are many categories, large differences and relatively independent research institutions, and the low cooperation rate of institutions. (3) Research hotspots mainly include data governance, data quality, big data, government data, etc., with few research hotspots and insufficient empirical research. (4) From the analysis of keyword migration timing and the sudden occurrence of keywords, we can find that the research on data governance has less innovation and insufficient research depth in recent years. With the continuous development and deepening of big data, data governance will also receive more and more attention. In the future, it can fully mobilize the enthusiasm of researchers, strengthen the information cooperation and exchange between researchers and research institutions, broaden the research fields, strengthen the depth of research, and diversify the research methods.

References

- [1] Lai K C, Chang V(2007). The need for data governance: a case study[C]. Proceedings of the 18th Australasian Conference on Information System, p.999-1008.
- [2] Be gg C, Caira T(2012). Exploring the SME quandary: Data governance in practise in the small to medium-sized enterprise sector[J]. The Electronic Journal Information Systems Evaluation, vol.15, no.1, p.3-13.
- [3] Coleman S(2008). Foundations of digital government[M]//Chen H.Digital Government.Integrated Series In Information Systems. Boston, MA: Springer, vol.17, p.3-19.
- [4] Zhang Kang(2018). Data governance: the Direction of understanding and construction [J].E-government, no.1, p.2-13.
- [5] Liu Guifeng, Qian Jinlin, Zhang Jiyong(2020). The Construction of Scientific Research Data Governance Model in Chinese Universities [J].Intelligence Science, , vol.38, no.6 p.17-24 + 31.
- [6] Xia Yikun(2020). Dimension analysis and Path optimization of government data Governance [J].E-government, no.07, p.43-54.
- [7] Xia Yikun(2020). Market allocation of data elements and deepening the reform of government data governance mode [J].Books and Intelligence, no.03, p.14-16.
- [8] An Xiaomi, Guo Mingjun, Hong Xuehai, Wei Wei(2019). The framework of Government big Data governance system and its effective path to implementation [J].Big Data, vol.5, no.3, p.3-12.
- [9] Liu Yiming, Xie Zehang(2020). Research on Published Enterprise Data Governance Based on the VCA Model [J].Publishing Wide Angle, no.10, p.11-14.
- [10] Han Xu, Cao Zengyi, Wang Zhaoyang(2018). Enterprise Data Asset Governance and Management [J].Electronic World, no.23, p.95+97.
- [11] Gu Bin, Peng Tao(2017). Che Wei.Research on the Full Life Cycle Management System of Enterprise Operation Data Assets [J].Enterprise reform and management, no.05, p.4-5+14.
- [12] Qiu Junping(2007). Informational metrology [M].Wuhan: Wuhan University Press.
- [13] Zhang Shaohua, Yang Lin, Gao Hongmei (2017). Song Jundian.Interpretation of national Standard of Data Governance Standards [J]. Information Technology and Standardization, no.12, p.25-29.