Research on the Data Governance Model in College Library

Meng Wang^{1,a}, Xiaoshuang Wang^{2,b,*}

¹Library, Zhejiang Ocean University, Zhoushan, China ²School of Information Engineering, Zhejiang Ocean University, Zhoushan, China ^a65238382@qq.com, ^b11986972@qq.com *Corresponding author

Abstract: The construction of library data governance model aims to measure an assessment model of ability, promote the implementation process and improve the process, which have a guiding way to enhance library data. This paper studied the composition of existing data governance factors, and found out the indicators that affect data governance. What's more, it put forward a new model of library data governance, and used matrix equation in mathematics to weigh. It was concluded that those factors in the process of data governance were more important, so as to compare which elements do better and which aspects need to be improved in the process of data governance. For the follow-up improvement, we can use this to clarify the governance aspects, so as to make the data governance more perfect and mature. This paper constructs a new library model through Matrix, which provides a new measure for library data governance, plays an effective role in library data governance, quickly makes the weak points of factors, and puts forward effective strategies.

Keywords: Model; Library; Governance; Information Technology; Algorithm

1. Introduction

Data is an important resource of a country, a place and an organization, and the library is the gathering place of books and data resources, which has a large number of important data. With the continuous development of information technology, it is inevitable that the center of gravity of college libraries would shift to digital resources. Under the exponential growth of data and the development of digital technology, college libraries need to continuously acquire and mine the potential social value of data, so as to give full play to the role of data. As the data and service center of teaching, cultural communication and scientific research, library is a typical big data industry with "data concentration" and "data as its core asset and business"^[1] The large-scale data assets owned by college libraries are facing both good opportunities and great challenges.

2. Current College Library Data

2.1 Current situation of data governance in college libraries

As the source of data, library has a large number of valuable data. If it plays the role of data effectively, it would provide important data reference and data value for schools and even local areas. As a brand-new concept and field, there are few theoretical achievements, empirical research and practice in this field in library and information institutions, libraries at home and abroad, especially in college libraries. Cases are even rarer^[2]. Mainly in the following:

2.1.1 The amount of data is huge and the types are complex

College libraries have purchased a large number of Chinese databases and foreign language databases, which have innate data advantages, but the databases are basically independent, data independent, database independent, and a large number of data are repeated, which occupies the storage resources of college libraries, and gives the original huge data even more huge. Coupled with the development of various data types, the complexity of library data is further increased ^[3]

2.1.2 Lack of awareness of data governance

Data governance is a systematic governance. Librarians should be conscious and improve their awareness of data governance. Managers are not aware of data governance, and the training of the whole library staff is seriously insufficient. Everyone is groping, without relevant experience and environment^[4].

2.1.3 The value of data utilization has not been brought into play

The above two points lead to the inability to reflect and play the value of data, and the extraction and mining of many important data can not be reflected.

2.2 Significance of data governance in college library

There is still no standard and unified definition of data governance, which is a complete set of workflow including policy, technology, management and control for data assets, and the key lies in controlling data quality, so as to serve the management and decision-making of all links(Yan 2020)^[5]. In conclusion, data management is the most common and important activity among various "data" activities, with the most universal meaning. It is a very important measure for data governance on finding a model to judge the effectiveness.

3. Current Mature Data Governance Models at Home and Abroad

3.1 Foreign data governance maturity model

Data Management Maturity Model (DMM) was developed by CMMI Research Institute, a subsidiary of Carnegie Mellon College based on the basic principles of Capability Maturity Model Integration (CMMI) and released in August 2014^[6].

The Data Management Capability Assessment Model (DCAM) is led by the Enterprise Data Management Council (EDM Council), which organizes financial industry enterprises to participate in compilation and verification, and is compiled based on the experience summary of many practical cases. It was officially released in 2014 and the second version was released in May 2019^[7].

Data Management Capability Maturity Model (DCMM) was initiated by the China National Information Technology Standardization Network in 2014, led by China Electronics Standardization Institute, and developed by a working group composed of Yushufang, Tsinghua College, China Construction Bank and China Everbright Bank. In April 2018, DCMM became the national standard "Data Management Capability Maturity Assessment Model GB/T 36073-2018" ^[8].

3.2 Domestic data governance maturity model

In 2018, China issued the national standard of data management capability maturity assessment model, "Data Management Capability Maturity Assessment Model (DCMM)" GB/T36073-2018, which is the first national standard officially released in the field of data management in China. Eight data management capability domains are proposed from four dimensions: organization, system, process and technology, including strategy, data governance, data architecture, data application, data security and quality, data standards and data life cycle. Each capability domain includes several capability items of data management domain, There are 29 in total, namely data strategic planning, implementation, assessment, data governance communication, organization and system, metadata management, data model, distribution, integration and sharing, data analysis and development, data service, data security strategy, management and audit, data quality requirements, inspection, analysis and promotion, business terminology, reference data and master data, etc..

3.3 Comparative analysis of data governance maturity models

The maturity models of data governance at home and abroad are shown in Table 1:

For the division of maturity models, they all have something in common, which basically starts from data governance, and then all mention policy system, data quality, data life cycle and other elements. There are differences in some elements of data management function, DCAM of EDM lacks reference data and management of main data, etc. The equationtion of national standard DCMM refers

more to the development status and characteristics of domestic data management, and adds assessment elements such as data standard, application and security

Model	Assessment elements	Quantity of	Rank
name		elements	
CMMI DMM	Data management is divided into five categories + one supporting process, including: data management strategy ,data governance, data quality, data operation, data platform and architecture, and supporting process	25	5
EDM DCAM	Data management is divided into seven capability domains: data management strategy and business case, data management project and funding model, business and data architecture, data and technology architecture, data quality management, data governance and data control environmen	31 Capacity subdomains and 106 capability terms	6

Table 1: Maturity models of data governance at home and

4. Thoughts on the Selection of Maturity Elements of Data Governance in Colleges

The selection of maturity elements is very critical, which directly affects the final data governance results, and is very important to choose elements. Generally, the basic large-scale selection is made first, and then the screening is carried out on this basis. Choose the elements suitable for your own data governance through some software. NVivo is a powerful software specially used for qualitative analysis, which can effectively analyze words, videos, sounds and pictures, and has rich types. It can intelligently get results from complex manual operations such as classification and sorting in data.

Maturity factor is to collect the data of college library data governance based on the existing research results, encode it with NVivo software, and then analyze it. Then the model needs to sort out the factors, and then these elements through some professionals to these elements for a judgment; Adjust and modify the elements to get accurate influence results

First of all, literature collection and collation. Through the well-known databases at home and abroad, the related literature of college data collation is inquired, and the query literature is coded by NVivo qualitative analysis software, and the related elements of college library data governance are obtained.

Secondly, get the rough maturity elements, compare and integrate the related factors of literature query and some well-known models at home and abroad, roughly determine the unfamiliar elements needed by the college data collation model, and explain some characteristics of each indicator.

Finally, the final maturity elements are determined according to the above.

5. Construction of Maturity Model of Data Governance in College Library

5.1 Analysis of maturity factors of data governance

Maturity elements collect the current data governance literature. The factor that maturity needs is to sort out and screen the existing documents in the database. The author selects CNKI and Web of Science databases, takes "library data governance" as the search term in CNKI database, and selects the last 5 years. If 5 years are not set, the query result is too big, and 26 documents are retrieved from the journals in the last 5 years. In the Web of Science database, 104 literatures in recent five years were retrieved. Node coding of 130 documents was carried out by NVivo software.

The above documents are coded in an open way, and then combined with the characteristics of data governance, the documents of data governance in college libraries are summarized. Through NVivo, "user management", "technology update", "information release", "data management", "data security", "quality", "scientific research" and "librarian" appear frequently. After classifying these words with high frequency, and then integrating sensitive words with the elements of domestic maturity models and my own work in the library, it is divided into 6 cores and 20 element indicators as shown in Table 2 below:

International Journal of New Developments in Education ISSN 2663-8169 Vol. 6, Issue 1: 1-6, DOI: 10.25236/IJNDE.2024.060101

core	Factor indicator					
User data	User management, user maintenance, librarian					
Business data	Books, sharing, quality					
data base	Database update, database security, database quality, database maintenance					
technology	Technical support, technical update and technical maintenance					
data management	Data management, data security, data update					
service	Information release, information update, information push and scientific					
	research					

Table 2: Six Cores and Twenty Element Indicators

5.2 Determination of maturity elements of data governance

In order to further determine the of these elements and indicators, we can interview some professionals through questionnaire survey, and conduct a questionnaire on these indicators. They score these indicators to show the recognition of experts on these indicators. We selected some doctors, associate professors, professors and people who have a certain understanding of the library, some of whom are teachers with rich library experience who have worked in the library for many years, to score the questionnaire. Each index adopts a score system of 1, 2, 3 and 4. 4 indicates great agreement. It is very important to recognize this index; 3. They agree that this index has a great impact on the model and constitutes an important factor of the model; 2. Agree that this index has a certain impact on the model; 1 indicates that they don't agree. They think this index has little impact on the index and will not affect the results of the model.

This time, 7 teachers were selected, including 5 doctors, who have worked in the University for many years, 3 teachers who have worked in the library for many years, 2 masters and 1 undergraduate.

Analysis method: as an excellent software product used for statistical analysis and operation, prediction analysis and decision-making, SPSS is used in various fields in data analysis. We use SPSS software to gap the answers, and use SPSS software to analyze the survey results of 7 interviewers.

techer essential factor	1	2	3	4	5	6	7	Mathematical expectation	variance
user management	4	4	4	3	4	2	4	3.57	0.53
User maintenance	4	3	4	3	4	4	4	3.71	0.20
staff member	4	4	3	3	4	3	3	3.43	0.24
books	3	4	4	4	4	3	2	3.43	0.53
share	3	3	2	2	3	3	3	2.71	0.20
quality	4	3	4	4	4	3	4	3.71	0.2
Database update	2	4	3	4	4	4	3	3.43	0.53
database security	4	3	4	4	3	4	4	3.71	0.2
Database quality	4	4	4	4	4	4	3	3.86	0.12
Database maintenance	3	4	3	4	4	3	4	3.57	0.24
Technical support	4	4	4	4	4	3	3	3.71	0.20
Technology update	3	3	4	4	4	4	4	3.71	0.20
Technical maintenance	4	4	3	3	4	4	4	3.71	0.20
data management	3	4	4	4	3	4	4	3.71	0.20
Data update	3	3	4	4	4	4	4	3.71	0.20
data security	4	4	4	4	3	3	4	3.71	0.20
Information Delivery	2	4	4	4	4	3	4	3.57	0.53
Information update	3	4	4	3	4	3	4	3.57	0.24
Information push	4	4	4	4	4	3	3	3.71	0.20
scientific research	4	3	4	4	4	4	3	3.71	0.20

Table 3: Scoring analysis

Mathematical expectation is the average score of each index. The average value is represented by m,

n represents the number of each index, and X represents the value of each item, $m = \frac{x_1 + x_2 + \dots + x_n}{n}$ Variance indicates the deviation degree of each index from the average value.

The smaller the data is, the more stable the data value is, and the larger the deviation from the average value is. S represents the variance, n represents the number of indicators, X represents each value, and

scores is: $\frac{1+2+3+4}{4} = 2.5$ It can be seen from the average score of each index that except that the

average value of element sharing is 2.5, the average value of other indicators is more than 3.43, and the performance is relatively excellent. Looking at the calculated value of variance, the smaller the variance, it shows that the score of teachers tends to be basically the same, either high or low. For example, the quality of element database, 6 teachers get 4 points and one gets 3 points, indicating that the score of teachers tends to be high. For example, in element sharing, five teachers give 3 points and two give 2 points, which also shows this element. Teachers give low scores for trend words, which is also similar.

Most of the factors with large variance result have stable scores, but the rise and fall of a teacher's score is relatively large, and the variance is too large, which is not controversial. For example, the variance value of the first factor is 0.53, five teachers score 4 points, one 3 points and one 2 points. If the variance result is too large, one teacher scores 2 points, and other scores are very high, indicating that this element is certain.

Based on the above two indicators, the average value of element sharing is low and the variance value is small, indicating that teachers can ignore the importance of this indicator to the model. Other index values are very important to the model. Finally, it is determined to eliminate the element of sharing, leaving 6 cores and 19 elements.

6. Construction of Data Governance Model of College Library

According to the determined 6 cores and 19 elements, a new model of data governance in college libraries is constructed. For 6 cores and 19elements, we use matrix to build a new model.

For the indicators of 6 cores and 19 elements, a_{ij} , $i = 1, 2, \dots, 6$; $j = 1, 2, \dots, 19$, a large matrix A with 6×19 is generated, and each refined indicator has a weight $\beta_i (j = 1, 2, \dots, 19)$ according to the collected data, and these weights form a 19×1 matrix B, and the score of each component can be obtained by multiplying AB the matrix. In addition, the weight α_i (*i* = 1, 2, ..., 6) of each component of data governance in data governance is considered to form a 6×1 matrix C of element weight matrix, and the equation is used Achieve the specific score of library data governance.

$$G = C^{T} A B = (\alpha_{1}, \alpha_{2}, \dots, \alpha_{6}) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,19} \\ a_{21} & a_{22} & \dots & a_{2,19} \\ \vdots & \vdots & & \vdots \\ a_{61} & a_{62} & \dots & a_{6,19} \end{pmatrix} \begin{pmatrix} \beta_{1} \\ \beta_{2} \\ \vdots \\ \beta_{19} \end{pmatrix}$$

7. Conclusion

This paper studies the composition of existing data governance factors, and finds out the indicators that affect data governance. What's more, this paper puts forward a new model of library data governance, and uses matrix equation in mathematics to weigh it. It is concluded that those factors in the process of data governance are more important, so as to compare which elements do better and which aspects need to be improved in the process of data governance, thus making data governance

more perfect and mature.

Acknowledgements

Academic research project of Zhejiang Library Society in 2021: Research on the Data Governance Model in College Library (Ztx2021A-7)

References

[1] Aspidistra. Comparative study and enlightenment of data management ability maturity model. Book and Information work. 2020, 64(13):51-57.

[2] Data Management Capability Maturity Assessment Model GB/T 36073-2018; General Administration of Quality

[3] Supervision, Inspection and Quarantine of the People's Republic of China; 2018-03-15 essentials. [2015-8-31]

[4] Fu Bo. Exploration of Data Governance Path of Domestic College Libraries. Journal of Shandong Radio and TV College. 2019, (03). 77-82

[5] Niu Lixue. Research on Maturity Model of Government Data Governance [D]. Hebei: Hebei College, 2020.

[6] Qin Zhongyun. Research on Data Governance and Maturity Model of University Library Century Library. New World Library. 2019, (11): 62-67

[7] Wu Jinchi, Yu Weijie. Construction of Library Data Governance Maturity Assessment System [J]. Information Science, 2021, 39 (01): 65-68.

[8] Yan Xin. Research on the Construction of Data Governance Framework of Public Library. Library. 2020, (05): 58-63.