Prospect of Archive Dataization in Big Data Age: Significance and Dilemma

GUO Wei-ya[a]

Abstract

The rising tide of datafication in big data era has aroused the attention of the archival community on the trend of archives datafication. Archives datafication is regarded as a new stage of the development of archives digitization. It refers to the process that archives departments transform digital archives resources into archives data resources that can be read, analyzed and processed with the orientation of users' needs and business needs. Archives datafication is of great practical significance to the data transformation and in−depth development of archival work in big data era. It is helpful to break through the bottleneck of the in−depth development of archives information resource, innovate the demand−oriented archives service model, and promote the realization of the intelligent management of archives business process. However, due to the existence of many uncertain factors, archives datafication is also faced with dilemmas, which are mainly reflected in the policy and motivation level and the talent and technology level.

Keyword: Big data; Archives datafication; Archives digitization; Archives information resource development; Smart archives; Archives service

Introduction

The emergence of computer technology, especially Internet technology, promotes the rise of digitization and brings about the revolution of digital memory storage. The big data strategy has a profound influence on the concept, technology, method and mode of archival work, and puts forward that it is necessary to adopt big data, intelligent management and other technologies to improve the informationization of archives business and the deep development and service level of archival information resources[1] .Through the research on the development present situation of our country's financial market and records management application, and finds some problems exist in archives management of financial market research shows that we should give full play to the role of archives in the era of big data, should be based on the data archives supervision idea as the direction, through big data technology in existing information platform establish scientific and perfect the big data file management mechanism, based on the change of regulation, the characteristics and value of original certificate file, file only, as the data storage and the final effective traceability

a. *Qilu Hospital of Shandong University, China.*
mickeypan114@126.com

certificate, make by means of supervision after inspection, the prosecution changed prior effective monitoring, rapid positioning data. Afterwards, accurate and effective tracking, prosecution, responsibility. To protect the financial security, safeguard the rights and interests of investors and the modernization of financial supervision.

1. Literature Review

Some scholars have tried to emphasize the value of "data" of archives, and think that in the data age, archives should be regarded as a collection of data that needs to be excavated[2]To make archives valuable through dataization[3]Some scholars take the excavation or deep development of archival information resources as the starting point, and emphasize the further data development of archival digital achievements[4][5][6]It is believed that only by transforming into "data state" can we carry out the association, mining and analysis of archival data and realize the data value of archives[7]There are also scholars based on the need of innovative archives service mode to promote the transformation of archives digitization to data[8]These research results reveal the coming of

the trend of archival dataization, but do not deeply analyze the connotation and development prospect of archival dataization. Based on this, this paper will explain the connotation of archival data, and look forward to the significance and dilemma of promoting archival data in the era of big data, in order to enlighten the archival work and archival research in the period of data transformation.

This paper first analyzes the current situation of market development and big data technology application supervision at home and abroad, summarizes the focus of current research through the results of research and investigation, and then puts forward some thoughts on how the domestic functional departments apply archives to realize supervision in the era of big data. Finally, how to face and solve the current difficulties to make some suggestions and prospects.

2. Trends of Data and the Intension of Archives Data
2.1 Trends in data

Digitalization realizes the transformation from analog state to digital state, and then promotes the social digital to be sure, digitization brings the analog world into the digital environment, which enables human society to store more information and process it faster. But in the digital age, digital information is still treated as analog, often used for the same "single purpose ", and the purpose of collecting this information is tied to its value.[10]After big data attracted wide attention, data became a popular term and trend.Victor Mayer-Schoenberg (Viktor MayerSchonberger) and Kenneth Cookie (Kenneth Cukier) introduced "digitizing " as a technical term with specific connotations into the public eye in 2013. At present, academic circles usually interpret data from three angles:

First, digitization represents a technological trend that translates many aspects of our lives into computerized data. Michael Schoenberg and Cookie believe that dataization transforms social behavior into online quantitative data for real-time tracking and predictive analysis[10]Some scholars also believe that dataization is to structure and granularize uniform and continuous digital bits to form standardized, open, nonlinear and universal data objects, and to implement relevant applications and activities based on different forms and categories of data objects[11

This kind of organization trend, it transforms the organization into the data-driven organization, forms the new value creation form the above

interpretation of data from the technical, organizational and social perspectives reveals the development trend of data transformation. This generalized trend of dataization mainly depicts the new data formation and processing methods in cyberspace, such as environment-aware data, scientific observation data, Internet of things big data, user-generated data, network and social media data, and the collection, aggregation, quantification and analysis of these data. The trend of data transformation will undoubtedly have an impact on archives, libraries and other traditional collection institutions, as well as enterprise document information departments. In recent years, the archival circles of our country have realized the influence of big data and data trend on archival work.

2.2. The connotation of archival data

"Digitization" is a high-frequency term in the digital age. It refers to "converting analog information into binary code in 0 and 1 that can be processed by a computer"[18]The corresponding basic units are bits (b i t), while the typical objects that are datated are bytes (byte) and words (word).Visible, data is endogenous in digitization, but also different from digitization. In recent years, under the impact of the trend of data, archival data has gradually become a new word of concern to the archival community. Many scholars have made a narrow understanding of the concept of archival data processing from the point of view of digital archives data processing. They believe that archival data processing is based on digitalization, using optical recognition (OCR), speech recognition (ASR) and other technologies to identify, classify, record and index digital content in images, audio and video, to form structured data, and to establish a directory database, full-text database and media resource database, so that the contents of archival resources can be linked, accessed, accessed, called, updated, analyzed and processed by computers, so that they can be retrieved with search words.[3][6][8]Some scholars believe that the dataization of archives is not only the category of data processing of digital archives, but also includes the process of structured data analysis, application-level encapsulation, media organization and chain management[19]

Compared the digitization of archives with the digitization of archives, it is pointed out that the main purpose of the early digitization of archives is to obtain the results of the digitization of archives in image format. The control and utilization of archival content are mainly realized by archival

catalogue database, forming a typical "digital image + archival catalogue database" file digitization mode. This model meets the needs of traditional carrier file computer reading, and does not directly solve the problem of file content control and development and utilization. Archival dataization refers to the process of transforming archival information into archival information resources that computers can read and understand. The aim is to realize the automatic analysis, understanding and disposal of archival information by computer, and to change the way of using archives from "page reading" to "content control" and "information development ". The way to realize it is to read the archival text by computer, and then convert it into the content that the computer can understand with the help of intelligent tools, so as to realize the deep development of the archival utilization value.[4]

3. Significance of Advancing the Process of Archival Dataization

In the data age, data to achieve decision-making and improve business, through data to reveal historical and cultural "phenomena" will have a certain impact on archival work. First, big data analysis can realize the deep development of multi-source heterogeneous low-density value data resources. Compared with social media big data and other external big data, archival resources with high density value as internal assets of enterprises are placed in the ranks of useless dark data for a long time because of the lack of data. In order to realize the value of archival data resources, it is necessary to realize archival data processing.Secondly, with the wisdom archives from the idea to the practice, the wisdom appraisal, the wisdom indexing, the wisdom retrieval, the wisdom compilation and research and so on file business wisdom realization also put forward the request to the file data. In addition, the demand for a large number of historical and cultural data sets in the digital humanities is also a driving factor in the digitization of archival data.

In recent years, the archival community has noticed the opportunity of archival dataization, and some scholars have proposed that "providing basic data for big data analysis is the value of archival work in the information age"[24]Only after the file data is made, can the data foundation of the file big data be laid, and it is possible to discuss the data mining of the file and the deep integration of the data[8]Some scholars also emphasize that through the "archival data leading service model innovation ", the traditional business of archival department

will be transferred to the direction of data analysis and data mining of archival resources, and the analysis and processing of a large number of data will become the main business of archival department[25]The author thinks that for the archives department (including the archives department of enterprises and institutions and the national comprehensive archives), it is necessary to define its own function in the data age, not only to provide high-quality archival data resources, data services and knowledge services, but also to explore the technical methods of improving business ability, service level, improving file utilization rate and improving work efficiency in the data age. Therefore, in the data age, archives departments need to re-understand the value of archival data resources in a development-oriented way, and realize the transformation of digital archives resources from machine reading to machine understanding through deep-level archival data development.

3.1. It Helps to Break Through the Bottleneck of Deep Development of Archival Information Resources

The bottleneck of the case department. From the perspective of development mode, the information resource development and utilization in library and information circles generally includes "classification, sequencing, digitization, networking, analysis and research, knowledge mining, service provision and decision making"[26]In contrast, the development and utilization of archival information resources mainly emphasize the two links of archival retrieval and archival compilation and research, in which the former is the shallow processing of archival information and the latter is the deep processing of archival information.[27]However, at present, the development depth of archival information resources with retrieval and compilation as the core is obviously insufficient. The archives department has a large number of valuable "primary materials ", and the archival compilation and research work has long become the routine work of the archives department, but it is always difficult to break through the bottleneck of" there is no research "in the archives department.

The level of development is closely related to the depth of digital processing of archives. Referring to the way of dividing the level of digital development of ancient books in the field of digitalization of ancient books[29]Taking paper

archives as an example, this paper divides the depth of digital development into four levels (see Table 1). Among them, the first layer is to solve the problem of scanning input and digital storage of paper files; the second layer is to basically solve the problem of network release and retrieval of digital archives resources; the third layer is shallow annotation and data processing of archival content; The fourth layer is file intelligent retrieval and intelligent compilation. Although the digitized archives have changed the form of archives existence and broadened the access of archives information, the essence of the archives is that the archives themselves and the contents of the archives are stored, and the further development of the archives information resources is not realized. At present, the digital development level of archives in most institutions in China is still in the "surface layer ".

Table 1. **Levels of Digital Development of Paper Files**

Level		Main Content of Digital Development of Paper Archives
Surface layer	One	Document files scanned; catalogued; indexed; OCR identified
	II	Metadata standards for archival description and description; web-based full-text retrieval of archival directories and full-text publishing
	III	Annotation and text mining of archives content; file data processing and data service
Deep	IV	Intelligent file retrieval; archival data association and knowledge discovery; intelligent identification and release; intelligent compilation and research

It is required to strengthen the development of state key archives, arrange a number of major topics for archives development every year, and carry out the construction of the national key archives catalogue resources infrastructure system. However, at present, the main task of the development of the information resources of the state key archives is to collect, record and build the catalogue database, the special database and so on. The emphasis is still on the collection, processing, organization and information management of the information resources, but the development of new information and new knowledge by using the information resources of the state key archives has not been fully carried out.[30]Based on this, some scholars put forward the concept of "integration of national key archives information resources ", emphasizing the comprehensive processing and cooperative utilization of national key archives information resources with the help of information fusion technology, so as to generate a new information space and knowledge structure.[31]

3.2. Contribute to the innovation of demand-driven archival service models

Constantly meeting the social needs is the eternal theme of archival service. For a long time, the archives department has developed a variety of service methods, such as file access, file compilation and research, file exhibition, cross-hall certification, information push, reference consultation and so on, but these service methods are difficult to meet the needs of archival users in the new era for network, convenience, efficiency, knowledge and personalized services. In the era of big data, some scholars believe that it is necessary to use big data technology to analyze structured, semi-structured and unstructured data such as user identity, borrowing content, storage behavior, search methods, words and deeds records, mining and predicting users' implicit demands, improving service methods, making it possible to upgrade service level, realize humanization and knowledge of archival service, and change from "supply-oriented" service mode to "demand-oriented" service mode centered on users.[33]It is true that the demand-oriented archival service model places the archival scholars on the archives of the big data era High hopes for service innovation, but the innovation of archival service cannot be separated from archival capital.The development of archival resources refers to the process of processing various carriers and forms of archives and archives collection in order to meet different archival needs in order to form various archival products or services. Its fundamental purpose is to deeply explore the valuable archival information contained in archival resources, to find and acquire more systematic or specific value knowledge and wisdom, and to effectively provide archival users with specific needs in various fields of society. In order to achieve archival resources and archival users demand docking, resource association and service matching.[34] Traditional archival services depend to a great extent on the depth of indexing and the compilation of retrieval tools such as document catalogue, file catalogue, thematic catalogue, thematic guide and full-family guide. With the development of digital archives, indexing of digital archives resources based on metadata has become

a mature technology, which is of great significance to the description, retrieval, management and long-term preservation of digital archives resources. At present, it can not only carry on the basic retrieval to the file volume level and the file level directory description item, but also the keyword retrieval, the topic retrieval, the metadata retrieval, the full text retrieval and so on. However, it is still impossible to define, analyze and identify the complex features of information units and the complex semantic relations between information units.

3.3. Helps to Promote the Intellectualization of Archives Management Business Processes

In recent years, the construction of intelligent archives has become the frontier of archival attention, representing the good vision of archivists for the construction of archives in the future. Yang Laiqing of Qingdao Archives pointed out that "wisdom" means that "the information system of intelligent archives can perceive, collect, memorize and analyze information independently according to the preset model and process, implement identification mode, trigger plan, control operation, evaluate business, predict trend, risk early warning and control information, excavate knowledge,

discover relation and so on, and complete the established archival work task ." He believes that there are two ways to build intelligent archives: one is to bring digital archives business and building information management business into intelligent archives, and to develop a broad sense intelligent archives system covering all the contents of archives information; the other is to retain or appropriately upgrade digital archives system, building intelligent management system, and develop a narrow sense intelligent archives system to meet the requirements of intelligent management.[37]However, from the current practice and research progress on the construction of intelligent archives, most of the archival circles have adopted the construction path of narrow wisdom archives. This means that intelligent archives only mean the upgrading and updating of hardware and software, and most so-called intelligent archives only realize the safe and intelligent control of archival entities, archival warehouse environment and buildings, which can be called the external wisdom of pursuing hardware and software systems. Instead of realizing the internal wisdom of archives and archives business processes.[38]

		C	$R_{mt} - R_{ft}$	SMB_t	HML_t	ISI_t	$Adj - R^2$
PB1	PE1	0.944857 （0.0614）	1.038831 （0.0000）	0.550068 （0.0000）	0.104951 （0.5367）	1.000869 （0.0402）	0.788896
	PE2	0.260408 （0.6375）	1.093442 （0.0000）	0.845857 （0.0000）	0.294997 （0.1179）	0.809175 （0.1310）	0.775600
	PE3	0.3223364 （0.5633）	1.149474 （0.0000）	1.150986 （0.0000）	-0.026086 （0.8903）	1.472856 （0.0072）	0.815220
PB2	PE1	0.898126 （0.0520）	1.098635 （0.0000）	0.580244 （0.0000）	-0.494265 （0.0019）	0.992725 （0.0264）	0.845182
	PE2	0.681544 （0.2020）	1.051619 （0.0000）	0.933141 （0.0000）	-0.200789 （0.2674）	1.104671 （0.0334）	0.796751
	PE3	0.428627 （0.4139）	1.129083 （0.0000）	1.217710 （0.0000）	-0.259296 （0.1468）	0.740043 （0.1450）	0.830156
PB3	PE1	0.847318 （0.0916）	1.115213 （0.0000）	0.617757 （0.0000）	-0.716554 （0.0001）	0.816362 （0.0918）	0.830629
	PE2	0.555777 （0.3259）	1.124220 （0.0000）	0.954568 （0.0000）	-0.610744 （0.0019）	0.801142 （0.1432）	0.804381
	PE3	0.611188 （0.2682）	1.111144 （0.0000）	1.151232 （0.0000）	-0.611870 （0.0014）	1.071953 （0.0457）	0.822263

The author believes that the core of the construction of intelligent archives lies in the intelligent operation of archives management business processes. The fundamental purpose of our pursuit of intelligent archives is to assist decision-making, control and operation, save

manpower, improve business efficiency and business level. Therefore, we should explore how to use the technology in the field of big data and artificial intelligence to realize wisdom collection, wisdom filing, wisdom indexing, wisdom identification, wisdom retrieval, wisdom

compilation and research, wisdom utilization and wisdom service through the establishment of relevant business rules and data models and process models. For example: collecting archival user needs, mining archival knowledge of user concern, embedding business intelligent question and answer application in intelligent archives system based on archival knowledge base, providing intelligent service function; automatically analyzing archival data resources and carrying out intelligent compilation and research based on big data analysis results of social hot spots and archival users using hot spots; realizing intelligent management and control of archives open wisdom identification and information security risk by

identifying and learning policies and regulations and sensitive words in the fields of confidentiality, information security, archives and so on. In short, in the era of big data and artificial intelligence, the convenience of meeting the high-level and personalized needs of archivists and archival users is amazing. To promote the above-mentioned archives management business to achieve wisdom, need to have the support of archival data. Only by realizing archival data can we realize the analysis of archival data level and lay the foundation for the application of big data and artificial intelligence technology. It is the tone and epitome of the construction of digital archives to the construction of intelligent archives.

B/M	C	$R_{mt} - R_{ft}$	SMB_t	HML_t	ISI_t	$Adj - R^2$
L1	0.544993 (0.2946)	1.118374 (0.0000)	0.905056 (0.0000)	-0.829812 (0.0000)	0.702859 (0.1620)	0.832737
2	0.662212 (0.2249)	1.130549 (0.0000)	0.859881 (0.0000)	-0.653547 (0.0006)	1.038830 (0.0498)	0.817304
3	0.765047 (0.0000)	1.095284 (0.0000)	0.959561 (0.0000)	-0.478086 (0.0106)	0.789485 (0.1328)	0.806833
4	0.725786 (0.1530)	1.109889 (0.0000)	0.899534 (0.0000)	-0.396462 (0.0226)	1.070040 (0.0301)	0.828975
5	0.855185 (0.1055)	1.068716 (0.0000)	0.896833 (0.0000)	-0.454906 (0.0120)	1.221466 (0.0174)	0.812486
6	0.467842 (0.3542)	1.102617 (0.0000)	1.038345 (0.0000)	-0.286465 (0.0963)	0.632323 (0.1950)	0.826756
7	0.596540 (0.2389)	1.101809 (0.0000)	0.799025 (0.0000)	-0.134338 (0.4334)	1.170195 (0.0179)	0.818296
8	0.465023 (0.3742)	1.085648 (0.0000)	0.869361 (0.0000)	0.046242 (0.7941)	0.623676 (0.2173)	0.795523
9	0.561070 (0.3125)	1.098873 (0.0000)	0.829286 (0.0000)	0.185229 (0.3257)	0.860037 (0.1098)	0.777752
H10	0.514557	1.100624	0.837462	0.207994	1.669939	0.792005

Under the background of big data, the archives department should set up the concept of big data, big archives and big service, take the initiative to act and actively integrate into the era of big data. According to the important way of resource construction, it is of great significance to strengthen national soft power.[17]Therefore, in the face of difficulties, it is urgent for archival institutions to carry out research and analysis on the needs of archival data, clarify the objectives and prospects of archival data, and determine the priority and selection strategy of archival data implementation. This paper studies the integration of archival data into the future archives information construction plan and the national key archives protection and development plan, and promotes the national

archives data development work as a whole. Local archival institutions should also strive to seek the integration of archival work with the big data strategy and the development of intelligent society, give play to their enthusiasm, act as vanguard, and actively strive for the support of local governments and finance. Explore the feasible path and implementation plan of archival data. Reference books, ancient books and other areas of advanced experience, as soon as possible to develop archival data related standards and guidelines.

4. The Dilemma of Promoting Archives Dataization

The big data era promotes the development of archives, but also puts forward higher requirements for archival work. Archival

dataization is a new trend in the future development of archives digitization. It is of great practical significance for archives departments to break through the bottleneck of archives information resources development, innovate the file service mode with user as the demand, and promote the realization of intelligent archives management business process. However, because the theory and practice of archival dataization are still in the initial stage, archival dataization still faces many uncertain factors. For example, the uncertainty of the realistic demand of archival data, the uncertainty of the policy support of archival data, the uncertainty of the standard specification of archival data, the uncertainty of the technical scheme of archival data, etc. Faced with difficulties, the most important difficulties are reflected in the level of policy and motivation, talent and technology.

4.1. Lack of planning, lack of motivation

Policy guidance and planning guidance has always been an important driving force for China's archival work.Since the 21st century, the digitization of archives has been included in the "15th ","11th ","12th Five-Year "and "13th Five-Year" plan of the national archives cause, which has become an indispensable part of the construction of archives information. Reflects the national level of archival digitization work attention. The outline of the 13th Five-Year Plan for the Development of Archives in China, issued in 2016, once again clearly puts forward the goal of "comprehensively promoting the digitization, incremental electronization and utilization of network" of archives management information.Not only that, this plan also mentioned to explore the integration with big data action ," establish open file information resources socialized sharing service platform, formulate file data open plan ". Although the concept of "open archival data" was creatively used at the national level for the first time, it reflected the change of thinking of the national archival department in the context of the big data era [39] However, the concept of archival digitization has not yet been clearly used, nor has an action plan for archival digitization been proposed.

All along, under the influence of the top-down archives management system, the local archives institutions in our country have formed a strong dependence on the overall planning at the national level. Because there is no clear policy trend on archival data at the national level, many local archival institutions also lack the initiative and enthusiasm to explore the road of archival data. Although some local archives institutions (such as Zhejiang Provincial Archives Bureau) have made bold and beneficial explorations and attempts in the emerging trends of archives department integration into big data, Internet + and so on, they have also not clearly formulated the relevant policies and action guidelines for archival digitization. Therefore, due to the uncertainty of the action of archival data in the national policy and planning layer, the motive force of archival department to explore archival data is obviously insufficient. Lack of planning, lack of motivation and the resulting low level of attention of local archives department leaders, lack of corresponding financial support, has become a major dilemma to promote archival data.

4.2. Shortage of Talent and Technology

In the early stage of archives information construction, the archival circles have fully realized the importance of talents in archives information construction. Zhang Zhaoyu pointed out that "information talent is the driving force to promote the development of archives information and the most valuable resource for the construction of archives information"[40] However, the shortage of archival information talents has not changed. Especially the serious lack of technical talents with computer and other professional background has become the bottleneck of archives information construction. With the arrival of the big data era, the archives department needs more technical and complex talents in the direction of database construction, data analysis, data mining, knowledge organization, artificial intelligence and so on. The current situation of archival professionals is not optimistic: first, the structure of archival professionals is unreasonable. The existing archives professionals in our country are older, more women, less highly educated and seriously short of modern technical talents; second, the skills of archives professionals are not high. According to statistics, according to the professional level of archives, the proportion of the existing full-time personnel of archives administration departments at all levels and comprehensive archives in the country is only 0.39 percent [41] It can be seen that the education level of most archival staff in our country is not high and the professional level is low; third, the information literacy and information mining ability are extremely weak. Most archival staff have low information literacy and weak ability

to collect, mine and analyze archival data and archival information resources.[42]

Database development, knowledge organization and knowledge discovery, such as metadata, ontology, semantic web, associated data, GIS 、 text mining, visualization, knowledge map, corpus, social network analysis, etc. However, many of these techniques have not been applied to the archives department. In the digital stage, most archival institutions solve the dilemma of talent and technology shortage by outsourcing. But at present, not only the archives department is not enough for the archival data technology, but also the preparation degree of many third-party companies on the technical solution of archival data is insufficient. Therefore, the shortage of archival data development talents and the shortage of technology bring great challenges to the deployment and implementation of archival data development projects.

In the face of the shortage of talents, some scholars from the perspective of education and talent training, think that archivists in the era of big data should have data sensitivity, data custody, data analysis, data processing and understanding of the data background ability. Based on this, they put forward that higher education of archives major should pay attention to training data processing talents, data analysis talents, data management and supervision talents, data storage and security protection talents, cultivate data management talents adapted to the docking of information science and archives management, and improve students' information literacy and information skills.[33Under the concept of "thick foundation and wide caliber ", the education of archives specialty emphasizes the importance of cultivating" compound talents "who are proficient in archives management and familiar with data analysis, data management, data mining and other technologies, but the complex talents of archives specialty cannot completely replace the role of technical talents trained in computer science, data science and other specialties. Therefore, the author thinks that for archival institutions and enterprise archives departments, the most important thing is to reform the recruitment mechanism of technical talents, introduce technical talents boldly, and alleviate the contradiction of unbalanced talent structure in archives department in big data era. In the face of the challenge brought by the lack of technology, archival scientific research institutions and third-party archival science and technology companies should actively explore the technical methods of

archival data development and provide feasible technical solutions for archival data development.

5. Research Meaning

Through the research and analysis of the current domestic market situation and existing problems, in the long run, how to realize the effective monitoring, rapid positioning and accurate tracking of the relevant information of market transactions based on big data technology and archival data is of great significance to the protection of national financial security and the legitimate rights and interests of investors. As a kind of archives, financial archives are the most authoritative records recording the occurrence of the financial market at all times, and also provide a reliable basis for the continuous development of the financial market. To do a good job in financial archives and use financial archives to serve the financial market is not only urgently required by the current objective environment, but also an important means to maintain financial stability in China in the future. Based on big data technology, we can improve the efficiency of financial archives supervision, and use the original and unique characteristics of archives and the value of archives voucher as evidence (the original record of transaction operation process), so as to realize effective monitoring in advance, rapid positioning and comparison of data in the event, and accurate tracing and responsibility determination afterwards. Making good use of this sharp tool of financial supervision will play a key role in the determination of illegal acts, and it will be of great help to realize the modern supervision of China's financial supervision department, resist and prevent the malicious shorting and long of international financial institutions, and monitor the programmed operation.

6. Limitations of Research

With the development of financial market and the enhancement of management consciousness, file management has gradually shown its unique guidance. It is of great significance to carry out file management in the current system. It is the most direct and effective means to implement financial supervision. Even before, the domestic understanding and application of archives management is still relatively weak. However, the further research on using big data technology financial archives to serve the financial market will be expanded in the future.

However, due to the limited knowledge and lack of cognitive ability of the author, there are some disadvantages in the research of this paper, such as

the research has not been demonstrated and feasible suggestions have not been put forward, and the research work needs to be deepened. Therefore, the author hopes that in the future study can make up for their own shortcomings, strengthen the weak points, clear research direction, in order to better carry out the research.

7. Conclusion

Big data, which includes all information, such as text, pictures, audio, video, and so on, is considered a set of technologies with the potential to change paradigms. It has been accepted by many industries, including medical, transportation, media and government, and has been successfully applied in the fields of Internet, retail, etc., but big data technology and its applications. The change of the environment. Knowledge discovery of multimedia archives, content mining and visualization of traditional genealogy, discovery of archival value oriented to knowledge association and realization technology in the National Social Science Foundation Project 2019 Project are the emergence of "technical tendency" topics, which reflect the change of archival research direction in the data age. Based on this, the author thinks that the research direction of archival information resources development, archival compilation and research will be further turned to the construction of archival data resources oriented to intelligent services, knowledge services, digital humanities, and archival data will become a new direction of archival information and an important means of institutional data resources construction

In this paper, the author analyzes the significance and difficulties of the big data era to promote the process of archival dataization. It is hoped that it can arouse the attention of archival data, further demonstrate the social demand and development prospect of archival data, and explore the technical solution, realization path and promotion strategy of archival data.

References

[1] Akella, Janaki. (2016), The value promotion of archives management in the era of big data, Journal of Finance, 23(4), pp.589-609.

[2] Beaver, William H. (2015), Challenges and opportunities of archives management in data Era, (4),ournal of Documentation pp.71-79.

[3] Carmichael. (2011), Construction and sharing mechanism of enterprise archives information resources under the background of big data, ournal of Documentation, No.2, pp.6-8.

[4] Fitzpatrick, P.J. (1932), On the logical management of Archives, ournal of Documentation, (10), pp.598—605.

[5] Giovanni Caggiano.(2016), Thinking about the development trend of archival science under the influence of big data, (5), pp.1-2.

[6] Jeehan Almamy.(2016), New ideas of archives information collection in the era of big data , Journal of Corporate Finance, (36), pp.278-285.

[7] Majeed Simaany.(2020), Digital archives and its service system for big data, European Journal of Operational Research, 280(L), pp.49-53.

[8] Mosh Hagig.(2009), Discussion on Archives Management under the background of big data Era, Journal of International Management, (L5), pp.286—295.

[9] Ohlson AJ. (1980), Impact and requirements of big data era on archives modernization, Accounting Research, (18), pp.107-L33.

[10] Odour MD. (1990), Research on knowledge mining technology in big data analysis of online digital archives, San Diego.

[11] Soltanizadeh S. (2014), Archives in the era of big data: an analysis based on SWOT,Social and Behavioral Sciences, 164, pp.332-337.

[12] Virginie Coudert. (2018), Information construction of enterprise archives in the era of big data, International Economics, (61), pp.98—116.

[13] Willett AH. (2009), Characteristics and development strategy of archives in the era of big data, Columbia University Press, (4), pp.142-L44.

[14] Zabio Oliva. (2016), Research on knowledge mining technology in big data analysis of online digital archives, International Journal of Production Economics, 173, 66-79.