

Research on Electronic Archives Management and Service under the Background of Big Data

Chen Lei

Coal Economy Academy
Shandong Institute of Business and Technology
Yan Tai, China
chenl77@163.com

Chen Shuying

Key Laboratory of Intelligent Information Processing in
Universities of Shandong
Shandong Institute of Business and Technology
Yan Tai, China
304779566@qq.com

Abstract—In the era of big data, the types of electronic archives are increasing, and the scale is expanding, which brings difficulties to the management and service of electronic archives. In order to solve the problems existing in the management of electronic archives, such as the loss of archives, the out of control and low level of development and utilization. After careful study, this paper puts forward the relevant measures of electronic archives management. It is necessary to establish the idea of front-end control, strengthen the management of electronic archives from the source. In the system to develop archives management regulations, standardize the use of archives to avoid irreparable damage. In the method of electronic archives preservation, use the system of "double sets" to save the whole archive information. The most important thing is to strengthen the management of staff, improve the awareness of archives protection. At the same time to do a good job in the management and preservation of electronic archives, the article finally designed the innovative way of electronic archives services.

Keywords—*Big data; Electronic archives; front control; Electronic archives management; Electronic archives service*

I. INTRODUCTION

In today's society, the era of big data has been roaring. Massive data not only profoundly affect people's work and life, but also bring new opportunities and challenges to the archive work. Archival science also needs to innovate in theory, management and service. Only in this way can we make the archival management service sustainable development, and promote the continuous progress of archival work.

A. Big data concept

First, Big data itself is a more abstract concept, literally, it represents the size of the data. But only the number of large apparently unable to see the concept of big data and the previous "massive data", "very large data", and what is the difference between them. At present, Academia has yet to have a generally accepted definition of big data, Different definitions are based on the characteristics of big data, through the elaboration and induction of these characteristics, and we try to give the definition. Among these definitions, the more representative is the 3V definition [1]. That big data needs to meet 3 characteristics: volume variety and velocity. Wikipedia big data are simple and clear [2]. Big data refers to the data set

that use of commonly used software tools to capture, manage and process data over time to tolerate. In the history of mankind, there has never been a generation of such a huge amount of data. Data generation has been completely free from time and place constraints, As Google's chief economist Hal Varian said [3], Data is widely available and what is lacking is the ability to extract knowledge from it. The fundamental purpose of data collection is to extract useful knowledge from the data according to the requirements, and apply it to specific areas.

There are several features of the era of big data: Multi-source isomerism, widely distributed, Dynamic growth, after the first data model. It is different from the traditional data management features, making big data era of data management is facing new challenges. Similarly, Electronic records of the era of big data also exist as above characteristics. From the type of electronic archives, not only includes a large number of text files, but also graphics, images, audio, video, multimedia and other documents. The value density of electronic archives is low, and the massive information which is accompanied by social activities, only a few of them are valuable, and the data processing speed is very fast.

B. The research situation of domestic archival science data

Taking CNKI as the data source, the key words are searched by the "big data", According to the published time statistics, time span from 1985 to 2017. Prior to 2010, most of the literature is on the "big data", "big data set", "big data transfer", "big data volume" and other related documents and computer technology. They are not the real meaning of big data literature. Starting in 2011, Wang Shan "Big data architecture: challenges, status and Outlook", Bai Yunchuan "Wait for the big data Era", are the first time in the domestic scholars CNKI big data for the study of literature. Therefore, the author believes that the domestic big data research started in 2011. In terms of the number of documents, there were 11 articles before 2013, from the beginning of 2013, according to a large number of literature data showing an annual rise trend, big data research shows in the past five years has become a hot research topic. In the name of "big data" and "file" as the key words for retrieval, access to the literature 1009, we can see that the

combination of archival research and big data began in 2012, in 2014 showed a trend.

II. THE CURRENT STATUS OF ELECTRONIC RECORDS MANAGEMENT IN CHINA

A. *The electronic archives management is out of control*

From the microscopic point of view, most Company do not develop electronic archives management rules and regulations, No one is responsible for the collection, archiving and preservation of electronic archives, The electronic archives in the state of emerge of itself and perish of itself. A large number of electronic documents are stored in the file of the drafters of the computer hard disk file, many of them in the revision process covers the original manuscript; A large number of emails are handled by the sender and recipient without screening. A large number of database files have lost their previous records in automatic updates; a large number of digital photos, video files, audio files, multimedia files scattered everywhere. Staff can change, delete, copy of the electronic file system unit information. Most of the electronic files by the formation of their own custody are free disposal. Up to the central, down to the place, the country has no unified management of electronic files and electronic records management has not been effective supervision, standardized control, So that we cannot know the number of electronic documents, content and distribution. It is impossible to carry on the macroscopic allocation to the national electronic files resources, scientific research and development, which has caused the enormous waste of social resources.

B. *The loss of electronic records is very serious*

At present, many organizations do not take any measures to retain the database, e-mail, multimedia files, web files and other types of electronic files; because the management method is simple, the vast majority of electronic files retained by the unit do not have a real file value. Some of the important electronic documents saved cannot be read due to improper management. January 6, 2000, the United States, Washington Post published articles published: In June 1999, the National Archives of the United States about 43000 minutes of e-mail missing, these emails include 125 users of internal documents. According to the analysis, it may be because someone mistakenly deleted an e-mail, causing the accident [4].

C. *Low level of development and utilization of electronic archives*

In recent years, China has accelerated the pace of building a service-oriented government, as the basic means of transmitting the government information, the electronic files is an important information resource. It is estimated that about 80% of China's information resources are generated and preserved in government departments. The core information resources of the government departments are mostly in the form of electronic files. Therefore, the development of the use of electronic files resources, so that the public to enjoy low-cost, high-quality information service is an important way to strengthen the construction of government services in the new era. At present, in addition to a number of government websites

and the regional electronic document center centralized storage and limited to provide part of the current electronic documents for government agencies and public services, the development and use of the program is extremely rare. Not only a large number of electronic files mastered by the government cannot be used for the society, but also the government cannot fully share the electronic documents, so as to realize the value of the information resources. According to the Information Office of Beijing in 2004, the data shared by all the authorities in Beijing are about 45% of the total data, and the vertical sharing data in each industry system is only about 36%, and the cross sector data is only 7% [5].

III. ELECTRONIC ARCHIVES MANAGEMENT COUNTERMEASURES

The main measures to improve the management and utilization of electronic archives are as follows :

A. *To strengthen the management of electronic archives and establish the idea of front-end control*

Due to the easy to loss, easy to change and other characteristics of electronic archives, it's easy to make mistakes, and the damage is irreversible. In view of the characteristics of the electronic archives "fleeting", the time of error prevention and control should be extended to the beginning of the life cycle of the document, From the early formation of the document to do a good job of maintenance, can effectively prevent the damage and destruction of electronic archives. The principle of "front-end control" emphasizes that the measures should be taken to the beginning of the formation of electronic archives, which will be taken from the source to ensure the authenticity and integrity of electronic archives [6].

B. *Establish the system of archives utilization and integration all kinds of archives resources*

We need to develop the system of archives utilization. When you need to access and use archives, you must first get the approval of the head of the archives. We also need to develop archives management system, Archives collection department can use network technology to carry out the archives collection management activities. This can make the archives collection more quickly, and it can also ensure the rapid updating of electronic archives, to better meet the needs of users, and to achieve the use of electronic archives and paper archives complementary.

C. *Regulate the use of archives to avoid irreparable damage*

We set the electronic archives content based on user situation classification categories are controllable management. Not fully open electronic archives, should not be used to provide a copy of the way, the process of browsing such documents must be carried out under effective monitoring. When take the network use, the content of information classified should be encrypted and replaced regularly or irregularly on the use of the "key", to avoid because of disoperation caused irreparable loss.

D. Use the "dual system" model to ensure the integrity of important information

The so-called dual system refers to the electronic archives and paper files submitted to save the archive [7]. "Double management" mode of archivists presents new challenges, requires the archivist to keep pace with the times, grasps the modern information technology, skilled use of computer and modern communication technology, otherwise you cannot participate in the document collection and preservation. In addition, which files should be established "double sets" and which should not set up a "dual system" are the main problem of double sets system, Only in accordance with the characteristics of the contents of various types of archives, there is a choice to establish a dual system, will not cause waste of human, financial and material, but also to achieve a complete preservation of important information.

E. To strengthen the management of staff, improve the awareness of archives protection

Because of the diversity of ways of using and relying on the diversification of technology, the use of electronic archives is very complex. Therefore, we must strengthen the management of the archives staff. From the ideological understanding of the importance of the work of archives management, and ensure the safety of the archives management. It can not only make the electronic archives information be used to the greatest extent, but also make the archive information reliable, safely and effective protection. In short, we must strengthen the electronic document production and management personnel responsibility, training them to develop loyal, responsible, good working attitude, so as to ensure the origin and authenticity of electronic archives.

IV. ELECTRONIC ARCHIVES SERVICE RESEARCH

Big data will affect all aspects of social life. As Gary Kin, an American sociologist, points out, this is a revolution, huge data resources began to change the field, No matter the academic, business or government, all areas will begin this process. Real time data acquisition, accurate analysis, the depth of excavation to become a reality, which brings a lot of growth points for electronic file information services. Big data era electronic file service model has the following points:

A. Personalized service of archives

The era of big data archives also show different characteristics with the past, the main demand is the personalized needs of the archives. With the advent of the era of big data, the number of archives data resources is growing, the growing variety, However, the use of archives for such a huge file data resources have a variety of ways to use and personalized needs. The traditional way of relying on the retrieval of archival data and the thinking mode of the passive service cannot meet the individual needs of the archives users in the big data Era. The arrival of the era of big data, so that the archives can be detailed measurement of the user's reading needs, reading behavior, reading mood and reading satisfaction. Therefore, how to effectively analyze the user's reading behavior data and social data, and accurately find out the user's

reading habits, preferences and needs, that will be the strategy and content of constructing the personalized service of archives; At the same time, it is also the key to the timely adjustment and accurate optimization of the service mode and content based on the actual needs of the users. Therefore, the archives should be based on the individual needs of users and the results of scientific analysis of large data, the archives should introduce the idea of personalized recommendation, personalized recommendation of archives resources. Archives should improve the personalized service system of archives data resources, and explore the use of archives in the era of big data. Archives should provide users with high efficiency, satisfaction, low carbon personalized data reading security services.

B. Intelligent service of Archives

Under the background of big data file services, the technology will be put forward higher requirements, the level of intelligence services will also reach a new level. First of all, from the perspective of the main body of the archives, archives need to use intelligent technology to carry out the advanced, complex data collection and processing, This can save a lot of manpower and material resources to a certain extent, but also to solve the work cannot be achieved manually, Such as the massive file information data intelligent grab, keyword extraction, etc. A great deal of manpower can be saved to study the service strategy. Second, from the user's point of view, the intelligent service and the means of the archives have been improved, and the information such as picture, video, text and so on have been produced, and the individual information such as social information and life data have been produced, A large number of unstructured and semi-structured data provided by the archives service can provide reference for the intelligent decision making of Archives. Third, from the perspective of knowledge flow in intelligent service, the improvement of the intelligence degree of archives service is beneficial to the circulation and dissemination of knowledge from a single subject to multiple subjects; this is conducive to the transformation of tacit knowledge to explicit knowledge, but also conducive to the discovery of knowledge, mining and organization. Therefore, the era of big data archives intelligent services will become increasingly important.

C. The knowledge service of Archives

Under the new situation of the booming development of the knowledge economy, we analyze the knowledge superiority of archives in the era of big data. We need to build a knowledge service engine based on big data from the point of view of data resource, data value and data analysis. We want to fully meet the needs of the era of big data archives of the archives of high-quality, diversified, and personalized needs. The service of archives is the embodiment of its core value and the significance of its existence. The progress of society and the development of technology have injected new vitality and power for the change of the mode of Archives Service.

V. CONCLUSION

To sum up, in the era of big data, cloud computing, social networking and other services are emerging. The type and scale of electronic archives has been increasing and accumulating,

which has brought challenges to the development of archives, and also brought about reform and opportunity. The management and service of electronic archives have become an important issue. At present, there are many problems in the management of electronic archives, such as the loss of control, serious loss, and low level of development and utilization. This paper puts forward five measures for the management of electronic archives. First, we must establish the idea of front-end control, strengthen the management of electronic archives from the source, Second we want to develop file management system, Third we should standardize the use of archives to avoid irreparable damage, Fourth we need to adopt the "dual system" mode to ensure the integrity of important information, Fifth we need to strengthen the management of staff to improve the awareness of the protection of Archives. In the end, the paper designs the service mode of electronic archives in the era of big data. Therefore electronic Archives to make a difference in the age of big data, we must have a clear understanding of the current form, Strengthen the analysis, management and application of the archives. We should take the initiative to innovate the service mode of archives, and constantly enhance the core competitiveness of the archives. Big data era data is complex and a wide range of types, this leads to a lot of

uncertainty about the results of the file service. Therefore, in order to provide a better quality of service in the context of big data, we need to continue to explore.

REFERENCES

- [1] Grobelnik M. "Big data computing Creating revolutionary breakthroughs in commerce", science and society [R/OL]. Oct2012.
- [2] The Economist Data, "everywhere - A special report on managing information" [EB/OL]. Oct 2012.
- [3] The Washington Post, "National archives of USA loses 43000 Emails", Jan2000.
- [4] Beijing information resources management center, "Beijing city information resources development and utilization of the", 11th Five-Year Planning Research Report, 2005 edition.in chinese.
- [5] Nian Yuying, "the characteristics and management of electronic documents", the world stage ,2008 ,pp.17-18.in chinese.
- [6] Sun Zhichang, " strengthen the management of electronic records, do a good job in the use of electronic files", Tokyo literature, Sep2011,PP.69-6-70.in chinese.
- [7] Nian Yuying, "the characteristics and management of electronic documents", the world stage ,2008 ,pp.20-21.in chinese.