

The Value Mining and Utilization of Archive Management in the Era of Big Data

Haocheng Yuan

College of Physical Science and Engineering Technology, Guangxi University, Nanning 530004, Guangxi, China

Abstract: *With the continuous development of big data technology, archival management is facing unprecedented opportunities and challenges. Big data technology has brought new directions of data integration, storage optimization, and intelligent management to archival management, significantly enhancing the management efficiency and utilization value of archival resources. This paper explores the strategies for value mining and utilization in archival management in the era of big data, covering aspects such as data integration and storage innovation, data analysis and value discovery, and in-depth development of archival resources. The paper proposes the construction of an intelligent management system, data visualization, and training to enhance the capabilities of archival managers, thereby promoting the transformation of archival management towards intelligence and refinement.*

Keywords: Value Mining; Utilization Strategies; Archival Management; Era of Big Data.

1. INTRODUCTION

In the context of the information age, archive management is increasingly facing many challenges and tests, the most important of which is how to effectively manage large amounts of data and maximize the utilization of archives. The emergence of big data technology has opened up new solutions for archive management, and data integration and analysis can fully tap into the value of archive data. The purpose of this article is to explore the value and strategies of archive management in the era of big data, with the aim of transforming archive management towards a better development direction.

2. OPPORTUNITIES AND CHALLENGES FACED BY ARCHIVE MANAGEMENT IN THE ERA OF BIG DATA

In the past, the management of archives was mainly based on paper archives and physical space, with a relatively single working method and low efficiency, which could not adapt to the increasing amount of information. The retrieval and management of paper archives are time-consuming and labor-intensive, and can easily lead to damage and loss of archives. Traditional systems for managing archives often encounter problems such as low retrieval efficiency, limited storage space, and difficulty in sharing archives. With the advent of the information age, especially the rise of big data technology, there is an opportunity for great changes in archive management. The digital storage, fast query, and intelligent analysis of archives effectively promote the work of archive management, making the efficiency and accuracy of archive utilization higher.

Although big data applications have begun to be applied in some archive management systems, their popularity is limited and they are still in the early stages of application. There are several challenges: firstly, many units still have a large number of paper archives, and progress in implementing digitization is slow; Even if electronic storage has been implemented, the difficulty in integrating these archival data can result in serious data silos, where data cannot be shared and integrated, and the value of data utilization cannot be maximized; Secondly, existing big data technologies are mostly used for storage and basic classification management, without intelligent and deep level applications, which cannot meet the increasingly complex needs of managing data. Data confidentiality and privacy have also become thorny issues in archive management. It can be seen that the application of big data must also be upgraded from hardware to software, from methods to applications, in order to optimize the utilization of archive management.

3. VALUE MINING OF ARCHIVE MANAGEMENT IN THE ERA OF BIG DATA

3.1 Innovation in Data Integration and Storage

From the perspective of data scale, traditional archives, due to their small amount of data, are no longer able to

meet the needs of current archive management work in terms of storage and management. The introduction of big data technology can help integrate and store archives, accelerating the digitalization of archive management work. In this process, it is necessary to focus on the following two points: (1) data integration technology and digital processing of archives. Data integration technology mainly refers to the ability to organize data from archives with different source types and formats through data integration functions, and unify different formats into a single data model. This helps to break down barriers between archive data and effectively integrate archive types. The application of data integration technology in digital archive work mainly refers to the unified and centralized storage and management of various archive forms such as paper archives, image data, audio and video materials, and electronic archives on the same platform. With the continuous development and deepening of digital archive management, data integration technology will significantly improve the efficiency of archive management work through its own effective means. Digitization of data not only frees archival materials from the constraints of physical space, ensuring the efficiency of data transmission, processing, and sharing, but also facilitates effective retrieval of data by staff, achieving good application results. (2) Optimization of storage efficiency for cloud computing and archives. Based on cloud computing technology, archive data is stored in the "cloud" end under this condition, and the data stored through cloud computing technology has the characteristics of elasticity and scale. Staff can expand or reduce the storage capacity at any time and freely according to needs, especially for large-scale archive data in data volume management, which has obvious advantages. It can make the storage and reading efficiency of data higher, and necessary backup processing can be carried out to ensure the security and stability of archive data. The intensive storage of cloud computing also provides convenience for timely sharing and collaboration of archive information, promoting the overall efficiency of archive management.

3.2 Data Analysis and Value Discovery

Archive management is not simply the storage, archiving, and management of large amounts of information. It also requires the application of data analysis and mining techniques to discover the potential significance of archive data and support decision-makers' analysis and decision-making. The key includes: (1) the application of data mining technology. Data mining can extract information, patterns, or patterns from a large amount of archival data, and help managers discover potential trends and tacit knowledge. For example, deep level mining and analysis of archival data can be carried out through data classification, clustering, and association rule mining. In the management of historical archives, data mining can provide forward-looking guidance for the collection, management, and utilization of future archives by analyzing factors such as document content, time, and region, and identifying trends in the development of a certain field or event. In addition, it can help discover abnormal patterns or error information in the document, which can improve the accuracy and correctness of archive data. In the context of digital management of archives, data mining can organically integrate and analyze different types of data, such as unstructured and structured data, in archive data. For example, in the management of literature archives, it is possible to automatically extract keywords, classify and cluster data through content analysis and pattern recognition of documents, and automatically associate different types of data to generate a more comprehensive and high-quality collection of archive information. (2) The application of data analysis methods. The application of data analysis methods, such as statistical analysis, regression analysis, pattern recognition, etc., in archive management can help promote the full utilization of archive resources. For example, through statistical analysis methods, archive data information can be analyzed through statistical means to uncover patterns or trends in the data. For example, statistical analysis of historical archives can reveal the number of archives, time distribution, regional distribution, and other information related to the topics or events recorded in the archives, supporting subsequent planning and management of the archives. Regression analysis and pattern recognition techniques can be used to support the prediction and decision-making of archival data. Based on regression analysis models, it is possible to predict the occupancy of archival resources in a certain field or type, thereby optimizing the allocation and management of archival resources. Based on pattern recognition technology, it helps archive management personnel identify potential patterns in archive data, such as automatically classifying and labeling archive files, or automatically generating relevant content recommendations based on file content, supporting file retrieval.

3.3 Reuse and Deep Development of Archive Resources

The key to improving the effectiveness of archive management is to carry out the reuse and deep processing of archive resources. Through cross domain collaboration and intelligent processing of archival information, archival resources can meet the different needs of cross domain decision-making and research, and enhance the potential value of archival information. Specifically, it includes: (1) cross domain collaboration and sharing of archive information. By collaborating with different fields such as government, research institutes, and enterprises, archive

information can be shared and utilized collaboratively across different industries. For example, archival information such as historical archives and cultural heritage archives can be provided to related industries such as scientific research and public services through shared platforms, forming multi-directional support for the value of archival information. Cross disciplinary collaboration based on archival information can enhance the effectiveness of archival information utilization, promote knowledge exchange and innovation between different industries, and help promote comprehensive socio-economic development. (2) Intelligent processing and utilization of archival information. Based on artificial intelligence, big data analysis, and machine learning technologies, automated classification, annotation, and retrieval processing of archival information are carried out to reduce manual intervention in the intelligent application of archival information and improve the effectiveness of archival information utilization. In the archives management system, natural language analysis technology can automatically identify and analyze key information related to the content of archives, conduct intelligent induction and analysis, and provide refined services for decision-making. At the same time, through deep learning, the implicit correlations in archival information can be analyzed, providing deep insights for scientific research, policy-making, and more.

4. UTILIZATION STRATEGIES FOR ARCHIVAL MANAGEMENT IN THE BIG DATA ERA

4.1 Construction of an Intelligent Archival Management System

The construction of an intelligent archival management system aims to enhance the efficiency of archival storage and retrieval, as well as the sharing and utilization of archival data. It also aims to increase the depth of data information mining and provide more precise services for decision-making. The application of big data technology can improve the data processing process of the archival system, increase the utilization rate of archival resources, and transform the traditional archival management model. The system construction should include the following aspects: (1) Design of big data technology and intelligent archival management system. The design of the archival management system should be able to support the storage and processing of a large amount of archival information, and enable flexible access and real-time updates of archival data resources. Based on big data technology, the intelligent archival management system realizes real-time monitoring, management, and operation of archival information big data. The combination of distributed storage and cloud computing architecture enables efficient storage and flexible access of data archives in different regions and platforms, with strong fault tolerance and high reliability. By using data warehouses and big data analysis techniques, the system can perform real-time classification and analysis of archival data, ensuring the real-time flow and efficient utilization of archival information [3]. The platform can integrate various data interfaces to achieve cross-departmental and cross-domain archival information sharing and collaboration. (2) Application of AI + machine learning in archival information classification and retrieval. AI can intelligently mark and automatically classify archival information based on the semantic of the archival text and context, no longer being limited by traditional manual classification. Using NLP (Natural Language Processing) technology, the system can perform semantic information analysis of document-form information, identify keywords and information fields, and achieve classification, significantly improving the efficiency and accuracy of archival management. Machine learning algorithms can automatically adjust the retrieval algorithm based on users' browsing history and preferences, providing personalized recommendations and rapid retrieval services for archives. Through intelligent classification and retrieval methods, the accuracy of archival retrieval can be improved, providing users with more relevant information that better meets their requirements, further increasing the value of archival management.

4.2 Strengthening the Visualization and Accessibility of Archival Information

In the big data era, archival management should not neglect the dissemination and utilization of archival information. The display of information enables users to have a deeper understanding and utilization of archives, providing decision-makers with more intuitive and efficient data support. Specific measures include: (1) Using visualization tools. To present massive archival data in graphical, chart, and other visual forms to users, making archival data easier to observe and utilize. Common visualization tools such as dashboards, heat maps, tree diagrams, and time series diagrams can visualize and graphically represent complex massive data and information, enabling users to quickly perceive the key contents and trends of the data. Data visualization is conducive to the display of archival content, and through images, the correlation and potential problems between archival data can be visually expressed [4]. Visual analysis allows users to intuitively see the distribution, change process, and historical development trend of archival data, and can analyze abnormal information and potential risks through monitoring real-time changes of data. (2) Integration of data visualization and decision support systems. In

archival management, through decision support systems to integrate various data analysis and visualization tools to achieve real-time data collection and presentation, helping decision-makers comprehensively understand the usage of archival resources and the management status of archival resources. For example, the system can generate real-time reports based on factors such as archival usage frequency, storage status, and update requirements, assisting managers in optimizing the allocation of archival resource usage and scheduling. Data visualization and decision support systems can also help predict future archival needs, make scientific plans and strategy adjustments. The system analyzes historical data and trends, which can provide forward-looking predictions for archive management, optimize resource allocation and management processes. This can enhance the intelligence level of archive management and help managers make precise decisions, avoiding resource waste or management errors [5].

4.3 Enhancing the Data Analysis Skills of Archive Managers

The development of big data technology has raised new requirements for the basic data literacy of archive managers, including the ability to analyze and apply data. Therefore, through regular data analysis and big data technology training, the understanding of archive managers regarding data collection, storage, processing, and analysis methods should be improved, enabling them to master the application of various data analysis tools. Different training plans can be proposed for different positions, using a hierarchical training approach. This is beneficial for promoting the improvement of the corresponding data analysis capabilities of managers at different levels and providing support for the implementation of archive management work. Additionally, managers should also adapt to the new technologies and change their thinking patterns in archive management. Specifically, internal exchanges and training learning should be organized to enhance the interest and recognition of archive managers towards new technologies. Starting from practical cases, the managers' understanding of the application of big data in archive management should be strengthened, such as data analysis in big data applications, data sharing, and intelligent search. Efforts should be made to improve one's technical ability and enhance the confidence and ability to execute complex tasks in archive management under the big data environment.

5. CONCLUSION

In summary, big data provides historical development opportunities for archive management work. With the help of data classification, scientific management, and intuitive presentation, archive management is bound to provide better services. Intensify the cultivation of data analysis skills for archive management personnel, continuously improve their adaptability to new technologies, and enable talents to promote the sustainable development of archive management. New technologies will promote the continuous development of intelligence and refinement in archive management.

REFERENCES

- [1] Chen Xinlei. Some Thoughts on Archival Management at the Grassroots Level in the Era of Big Data [J]. Shaanxi Archives, 2023, (06): 56-57.
- [2] Sun Huanying. Practice and Reflection on the Construction of Digital Archives in the Era of Big Data [J]. Lan Tai Inside and Outside, 2023, (36): 37-38 + 41.
- [3] Jiang Jing, Huang Changhao. Analysis of the Comprehensive Characteristics of Archival Data Mining in the Era of Big Data [J]. Heilongjiang Archives, 2022, (02): 107-109.
- [4] Xu Wenwen, Wang Yaqiong. Exploration of Digital Transformation of Archival Management in the Information Age [J]. Heilongjiang Archives, 2023, (06): 129-131.
- [5] Xiao Yanli. Research on the Data Transformation of Archival Management from the Perspective of the Internet [J]. Science and Technology Wind, 2022, (36): 70-72.