Innovative of Archives Management in China Based on the Application of Big Data postprint

Authors: Caiming Zhang, Wei Wang, Xiaoyi Wen, Taiping Pan, Xinglong Ning, Menghan Yang, Jiantao Wang, Peng Wang, Caiming Zhang

Date: 2022-11-17T00:00:00+00:00

Abstract

This article provides a comprehensive review of the relevant literature and conducts a thorough systematic analysis of big data applications in archives management innovation. It identifies the need to create a file management system based on big data applications across multiple dimensions: file data acquisition, data transmission, data storage, archives utilization, staff training, and security construction. Finally, this article analyzes the practical implementation and effects of archives management innovation based on big data application in China, and draws corresponding value conclusions.

Full Text

Preamble

Innovative Archives Management in China Based on the Application of Big Data

Received date: 20 March 2021
Revised date: 31 July 2021
Accepted date: 18 September 2021

Caiming Zhang[1], Wei Wang[2], Xiaoyi Wen[3], Taiping Pan[4], Xinglong Ning[5], Menghan Yang[6], Jiantao Wang[7], Peng Wang[8]

[134]China University of Labor Relations, Beijing, China
[25678]Gansu Jiyoupin Network Technology Co. LTD

Corresponding author, e-mail: zhangcaiming@cule.edu.cn

Abstract: This article provides a comprehensive review of the relevant literature and conducts a thorough systematic analysis of big data applications in

archives management innovation. It identifies the need to create a file management system based on big data applications across multiple dimensions: file data acquisition, data transmission, data storage, archives utilization, staff training, and security construction. Finally, this article analyzes the practical implementation and effects of archives management innovation based on big data application in China, and draws corresponding value conclusions.

Key words: Big Data Application; Archives Management Innovation; Chinese Archives Management

Funding:
1. Educational Reform Project of China University of Labor Relations: Construction of Experimental Curriculum System for the Construction of National First-class Undergraduate Major—Taking Human Resource Management Major as an Example (No. JG006)
2. Postgraduate Education Reform Project of China University of Labor Relations: Research on the Implementation Path of Intelligent Education (No. YJG006)
3. Archival Science and Technology Project in Gansu Province: Development and research of archive collection system (No. 2018-13)
4. Gansu Science and Technology Project: Development and Application of Special Archives Information Management (No. 18CX5JG047)

1 Introduction

With the continuous advancement of information technology, the digitization of document management using computer and network technologies has become a critical component of reform efforts across enterprises, government agencies, and social organizations. Early implementations of paper-to-digital document conversion were often crude and fragmented. The advent of Chinese character OCR technology enabled document scanning with character recognition for computer storage, yielding significant achievements in exploration and practice. However, notable shortcomings persisted: despite substantial improvements in computer processing efficiency for paper documents, the overall input process lacked a complete information system, making comprehensive monitoring and coordination difficult to guarantee. Compounding this problem, the diversity of archive types led to management systems with disparate functional designs and implementations, while data specification standards remained ununified across archives management systems.

Paper document sorting is extremely time-consuming and labor-intensive, demanding high levels of knowledge and management experience. Large-scale decluttering projects often require dozens of professionals and dozens of days or more to complete, with errors easily occurring throughout the process. Consequently, many enterprises and departments fall into a pattern of "archiving without collating" when handling paper documents. The application of information

technology in archives management is therefore both necessary and important. Using information technology during file collection can enhance the quality and level of archives management while ensuring proper software operation, error prevention, and scientific validation of archival data.

2 Literature Review

Liu Yong et al. (2017) analyzed archival work within the big data context through the organic integration of platform, management, technology, resources, services, and talent, positioning archives departments as management centers, data supervision centers, and information service centers. According to Hou Jia (2013), smart archives management construction faces information challenges that provide unprecedented opportunities for archival departments, demonstrating the irreplaceable importance of archival work through four transformations: from "dead archives" to "living information," interconnected archival information, in-depth development of archival information, and service consciousness for the overall situation. To ensure the authenticity and security of foundational "front-end" data, archives departments must integrate relevant original information resources scattered across various systems within government organizations to build a unified, authoritative information resource system and service platform that enables information sharing.

Archival data in the big data era exhibits characteristics of risk clustering, comprehensive crossover, dynamic ubiquity, and hidden relevance. Archival data are stored not only on paper carriers but also in binary-based archival information systems. Program problems and technical vulnerabilities in these systems make archival data more susceptible to distortion and tampering (Jin Bo and Yang Peng, 2020). The training system for archival data security managers remains imperfect and unsustainable, and a security talent team with broad interdisciplinary knowledge and high-level skills has yet to be established (Qin Qiaoyun, Zhou Feng, Yang Zhiyong, 2017). The application of big data in archives management innovation is both urgent and significant across several key dimensions:

2.1 Archive Data Collection

Archive collection represents the first step for archival data entering the system, and ensuring data security during this stage lays the foundation for comprehensive security governance. First, archival data formats must be standardized. Due to diverse sources and structures, archival data exists in structured, semi-structured, and unstructured forms. Integrating and preprocessing these data sources is essential for subsequent provision of high-quality archives and ensures data compliance, consistency, and legitimacy (Feng Dengguo, Zhang Min, Li Hao, 2014; Shao Qifeng, Jin Chee-qing, Zhang Zhao, Qian Weining, Zhou Ao-ying, 2018). Second, ensuring the authenticity of archival data entering

the system is critical, particularly against the risk of maliciously created “fake files.” Archival data traceability technology, blockchain technology, and timestamp mechanisms can effectively protect the authenticity of archival data from tampering (Yu Yarong, Zhang Zhaoyu, 2020).

2.2 Archival Data Transmission

Due to their high value and large volume, archival data face security threats at various levels during network collection and utilization transmission processes. First, archival data are vulnerable to hacker attacks during transmission. Archival data encryption and anonymity technologies should be employed to protect transmission security, using symmetric and asymmetric encryption to protect key archival data. Simultaneously, relevant network protocols should be signed during transmission to coordinate relationships between parties and clarify security responsibilities. Second, many data files in the system depend on specific processing environments, and archival data migration can result in missing or unreadable file data. To address file data heterogeneity, a national strategic-level archival data storage platform should be constructed to solve platform heterogeneity problems (Qin Qiaoyun, Zhou Feng, Yang Zhiyong, 2017).

2.3 Archival Data Storage

As archival data are stored in archival information systems, system functionality directly affects archival data security. It is necessary to improve the safe operation and management systems of archival information systems, enhancing prevention and control capabilities, authority settings, and management levels. The archival information system manages the entire lifecycle from collection to destruction of archival data. First, software and hardware equipment must be continuously strengthened and updated, with security prevention and control capabilities constantly improved to reduce system vulnerabilities. Second, archival information system risk and crisis response capabilities should be enhanced through anti-intrusion detection technology and firewall technology. Firewall technology can effectively block illegal access, enable data encryption, and prevent illegal Internet intrusion. Improving the fine-grained access control of archives information systems and continuously enhancing vulnerability repair capabilities help identify malicious attacks and monitor data processes to determine malicious invasion behaviors (Liu Xuan, Dong Xin Luna, Ooi Beng Chin, 2013).

2.4 Archival Data Utilization

Archival utilization is the fundamental purpose of archival work and should be realized to the greatest extent possible. Yang Zhiyong and Zhou Feng (2016) discuss four archival service modes under the smart city background, examining how different combinations and interactions of subject, object, content, and service elements can achieve diversified, personalized, accurate, efficient, and convenient archival information dissemination. Lyu Yanbing (2016) proposes

archival data resource construction based on open data concepts, identifying three key areas for future digital archival resources: open data, credit data, and crowdsourced data. Archival data utilization represents both the starting point and foundation of archival data management. In the big data era, the core issue is not who owns the data, but who uses the archival data, for what purposes, and what value can be created. The most fundamental significance of archival data management lies in realizing deep-layer efficiency, protecting data, and ensuring utilization. Therefore, establishing security sharing mechanisms and archival data alliance sharing platforms during utilization is essential. For example, the Australian National Data Service Center has successfully integrated data information resources from over 100 Australian research institutions, governments, and universities, improving data resource integration and development sharing processes (Chen Wenjie, Cai Lizhi, 2016). Access control mechanisms for archival information systems should be established with appropriate strength and granularity according to compliance requirements, including identity authorization and data tracing to reasonably limit visitor access scope. During archival data use, identity identification and access control rights should determine which archival data can be used and which personnel can access which data (Liu Yuenan, 2020). Based on different subjects, archival data types, and network systems, the security protection system must combine archival data permission control with user permission control to protect key archival information.

2.5 Archives Professionals Training

Archivists serve as keepers and guardians of archival information across various units and periods, and the realization of archival value depends on their expertise. Therefore, with the development and innovation of archival work under the big data background, archivists have become key determinants of archival work development. Along with the rapid development and application of Internet of Things, big data, cloud computing, and new media technologies, strong knowledge, wisdom, and information literacy have become not only necessary qualities for archivists but also technical support for archival work innovation. Gao Qixiang (2017) examines the necessity of information literacy promotion through national policies, industry norms, institutional evaluation mechanisms, and personal professional identity, encompassing personal accomplishment (service quality and media literacy) and information capability (data management and data mining skills). Yang Zhiyong and Meirong Fei (2017) argue that archivists should be both “professional” and “knowledgeable” to meet digital age and smart city requirements. Wang Shang (2017) contends that archival talent construction should include personnel selection, evaluation, utilization, education, and retention to grasp the latest field developments and continuously expand knowledge scope. Tang Yizhi (2014) argues from the perspective of archives management repositioning that archivists should learn, use, and master new technologies, especially new information technologies. Xue Jinling (2012) discusses archival talent construction from perspectives of professional

quality, information awareness, professional knowledge, information technology, and foreign language proficiency.

2.6 Archive Security Construction

Archival security refers to implementing effective protection measures to safeguard archival entities and contents from natural disasters or human damage. Jin Zhizhi (2016) elaborates on archival information security categories (including transmission security and media security), proposing promotion strategies across four aspects: strengthening relevant security business training, building an "Internet + Internet of Things" security monitoring system, implementing cloud grading and security indexing, and conducting timely security backups. Zhai Fei (2016) analyzes practical security risks in archives information systems, exploring five aspects for archives security: establishing and improving safety management systems, scientifically predicting security risks, ensuring long-term effective information storage, improving information security awareness, and detailing information security standard systems. Corresponding prevention measures should be taken in practice, with network environment archival information security strategies proposed across seven aspects including network security evaluation implementation.

---

3 Archives Management System

The system comprises two subsystems: the front-end file collection system and the archive collection system.

3.1 Front-End File Collection System

The front-end file acquisition system is a newly developed digital file processing program designed to meet file attribute requirements, comprehensively covering various file types. Combined with self-developed OCR region automatic recognition technology, it can quickly and accurately identify various descriptive items required for catalog data and intelligently generate electronic files, replacing the traditional "manual catalog description—scanning—connection" digital processing workflow.

3.2 Archive Collection System

The archive collection system breaks traditional filing methods by restructuring the 15-step file collection process into a three-step model of "document scanning, sorting, and online review." Combined with improved TextRank methods and digital watermarking technology, it intelligently generates page numbers, package numbers, file numbers, chapter files, archive file directories, reference appendix tables, etc., completely replacing manual input to achieve automated and networked archive collection.

4 Application of Archives Management System

The archives digitization process avoids rework and resource waste, saving storage costs and space while providing convenience and speed, preventing repeated printing and consequent paper and personnel waste. In archives management, the application of the archives management system significantly improves collection and collation efficiency. Previously, extensive manual work in sorting, collection, classification, recording, searching, and management consumed substantial human resources, reduced record management accuracy, and severely depleted personnel energy. However, archival information management system applications easily overcome these limitations. More sophisticated archival management methods and systems have been implemented, enabling relevant personnel to conduct detailed processing, differentiation, and screening of archival information, thereby achieving sustainable archives management. The system establishes a classification management module that can adjust instructions at any time according to circumstances.

5 Conclusions

The innovative practice of archives management systems in China based on big data application can generate substantial value and benefits.

5.1 Saving Costs and Reducing Rework

Previously, the extensive model relied on increasing office staff and costs as the sole solution, causing resource waste, extensive rework, and significantly increased management costs. The archives digitization process avoids rework and resource waste. The digital archives management system transforms traditional paper-based archives into machine-readable files, saving storage costs and space while providing convenience and speed, preventing repeated printing and consequent paper and personnel waste.

5.2 Achieving Sustainable Archival Work

The application of archives management systems significantly improves collection and collation efficiency. Archival information management technology systems easily overcome manual management limitations, enabling more sophisticated archival management methods. Relevant personnel can conduct detailed processing, differentiation, and screening of archival information. Through information technology application, more rapid workflows can be identified, reducing data loss probability and enabling integrated data information management, thereby achieving sustainable archives management.

5.3 Establishing a Standardized Archives Management System

The classification management mode based on tagging applications has transformed traditional file management encompassing documents, infrastructure data, personnel files, and unit development history. With archival information management technology systems, classification management modules and systems can be easily established. Using computer technology to set instructions, electronic archives can be categorized according to unit production, finance, personnel, and administrative management, establishing a comprehensive and standardized file management system. The application of archives management systems significantly improves archive collection and collation efficiency.

References

Jin Bo, Yang Peng. (2020). Analysis of archival data security governance strategy in the era of big data. *Information Science*, 38(9):30-35.

Wang Shiwei. (2017). Smart society is a new realm of smart library development. *Library Journal*, 36(12):9-13.

Qin Qiaoyun, Zhou Feng, Yang Zhiyong. (2017). Research on information security of digital archives in big data environment. *Beijing Archives*, (6):18-21.

Shao Qifeng, Jin Cheqing, Zhang Zhao, Qian Weining, Zhou Aoying. (2018). Blockchain technology: architecture and progress. *Journal of Computing*, 41(5):969-988.

Feng Dengguo, Zhang Min, Li Hao. (2014). Big data security and privacy protection. *Chinese Journal of Computers*, 37(1):246-258.

Yu Yarong, Zhang Zhaoyu. (2020). Design of electronic archive evidence collection and verification scheme based on trusted timestamp service. *Archives Management*, (1):66-68.

Liu Xuan, Dong Xin Luna, Ooi Beng Chin, Srivastava Divesh. (2013). Online data fusion. *Proceedings of the 39th International Conference on Very Large Data Bases (VLDB '2013)*. Trento, Italy, 97-108.

Chen Wenjie, Cai Lizhi. (2016). Big data security and its evaluation. *Computer Applications and Software*, 33(4):34-38+71.

Wu Jiaqi, Chen Xiaoling. (2015). Research on Co-construction and sharing Mode of Regional Archival Information Resources under the background of smart City. *Archives Management*, 1:33-35.

Lv Yanbing. (2016). Digital Archive Resources in Smart City Framework. *Journal of Zhejiang Archives*, 1:24-26.

Yang Zhiyong, Zhou Feng. (2016). Research on Archival Information Service Mode for Smart City. *Bulletin of Archives Science*, 4:44-49.

Jin Zhizhi. (2016). Information Security Promotion Strategy of Smart Archives. *Management Observation*, 6:46-48.

Zhai Fei. (2015). Research on Digital Archive Information Security Guarantee System. *Office Business*, 2016, 22:59-60.

Wang Yihan. (2015). Research on the Security System of Archive Information Management in Network Environment. *Technology Development of Enterprise*, 34(17):123-124.

Note: Figure translations are in progress. See original paper for figures.