http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (17): 4277-4280, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4277.4280

Study on Case of Application and Practice of Archives Digitization Work

Jinquan Zuo
Archives of Jinan University, Guangzhou 510632, China

Abstract: This study is based on the practice of Jinan University establishing digital archives workflow model, aiming to analyze archives digitization work flow practice. Conscientiously, from the specific work of the archives digitization, covering all aspects of digital archive in the university Archives, formulate archives digitization work flow scientifically and rationally and wishing to provide theoretical support and practical reference in order to carried out University Archives digitization work.

Key words: Archive, digitalization, workflow

INTRODUCTION

University Archives digitization can effectively protect the original archives, a digital archive can be transmitted to the remote place via computer LAN or WAN easily which expands the use of archives in different areas and shares archives information in the university (Zhang, 2009). However, the University Archives have a long time span, share a wide variety of categories and are frequently used. Advanced hardware equipment and digital technology is preconditions and means of the University archives digitalization and standardized workflow makes archive digitalization progress effectively.

PRESENT RESEARCH AND SITUATION OF COLLEGE ARCHIVES DIGITIZATION WORK

"The National archives Twelfth Five Year Plan" printed and distributed by the National Archives says that, one of the Main task of future archival is that we should speed up making copies of archives, protect the original files, generally carry out Archival Digitization and build the archives information Security management system. This document promoted the process of archives digitization work.

From the research status of University Archives digitization, all universities have put forward ideas and suggestions of digitization work (Wang, 2009; Ding, 2009). These studies mainly focused on the theoretical research aspects of digitization work, including importance and related countermeasures of archive digitization (Li, 2009), the problems exists in the process of archive digitization, you should follow the basic principles and development strategies of archive digitization (Guan, 2010), etc.; also part of the research scholars from a technical perspective to probe into establishment of archives digitization management platform system (Xue and Huang, 2004) or directly study digitize OCR text recognition technology (Xu, 2011) and bar code technology what applied in the archives digitization. Overall (Jin, 2008) these studies played a positive role in promoting the Information Construction of china' university archives and development of archives digitization but they don't elaborate about how to proceed the archives digitization systematically and comprehensively from a practical perspective.

In order to clarify the current situation of the university's archives digitization, I interviewed about 40 colleges and universities in Guangdong province via the telephone, these universities have carried out archives digitization. But, in fact, because of shortage of funds, those universities' have not yet make a long-term plan of archives digitization, so that the archives digitization process is different among those universities. The outsourcing model of archives digitization is a bold attempt in the construction of archival digitization (Wang, 2011) and has been widely used in the archives of universities in Guangdong province and they have summarized a lot of supervision and management strategies which will contribute to improve the achievement quality of archives digital. Also, the staffs of archives could not carry out their work effectively and could not match the skills that is needed in archival digitization, so it's difficult for them to control the security in the process of archival digitization and it's will dispute between the archival digitization outsourcer and the archives due to the unclear responsibilities and obligations which is conducive to the work of digital archives. Furthermore, Outsourcing businesses is a short form of work in archival digitization it can't carry out the daily work of the archives digitization timely and effectively.

Therefore, this study will base on Jinan University Archives of archival digitization work practices; Aims to analyze archives digitization work flow practice.

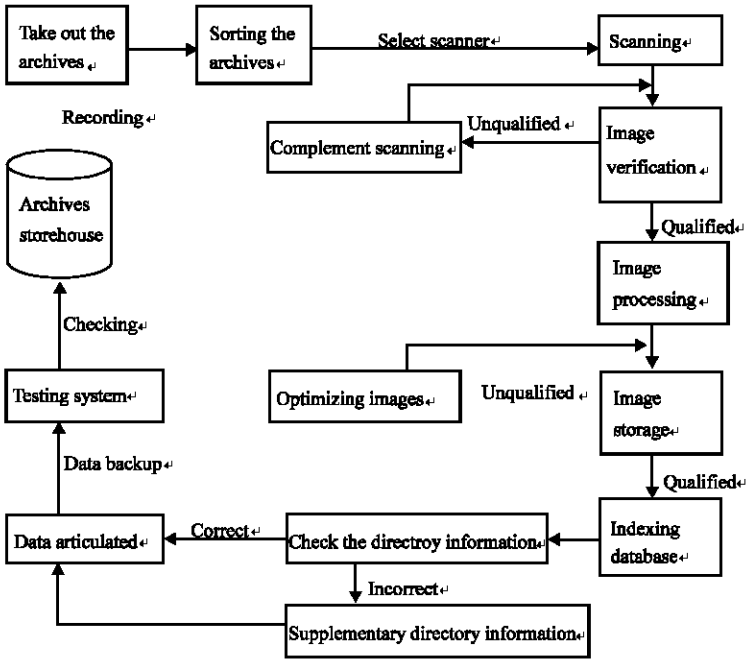

Fig. 1: Archives digitization process figure

Conscientiously, from the specific work of the archives digitization, cover all aspects of digital archive in the university Archives, Formulate archives digitization work flow scientific and rational. Hope provides theoretical support and practical reference in order to carried out University Archives digitization work.

ARCHIVES DIGITIZATION WORK

Archives digitization of the basic aspects include: Sorting archives, Building a database, Archives scanning, Image processing, Image storage, Data quality, Data articulated, Data inspection, Data backup, Results-based management and so on.

Base on Jinan University Archives of archival digitization work practices, this study will design the archive digitization work flow As follows chart1-1, ensure that archive digitization work: Clearer, Definition and Visualization.

Take out the archives: As the name implies it is take out the original archives of daily digital processing that needed to complete from archives storeroom and fill out the registration form out of storage. For amount of this grant of archives divided batches, Placed in the pending archives scan storage and to ensure the security of study archives.

Sorting the archives: This step is different with the sorting when the files into Archives. Specifically it is take the files out of the box and remove the binding, one thing should pay attention is to protect archives from damage, If there is serious damage, cannot be directly scanned files, Fold original archives should first conduct a technical repair and should be flattened or ironed, etc., Based on the original archives choose the right scanner, That is high-speed scanning or flatbed scanning.

Scanning: According to the archives format size (A4, A3, A0, etc.) and study quality, select the appropriate scanner, archival material which is better condition Study can be used in A4 high-speed scanners and maintenance of Status and admission be used in A3 high-speed scanners, infrastructure projects drawings can be used in A0 engineering scanner; for the purpose of protecting archives, The older years study files in poor condition that can only use the flatbed scanner or packaged a transparent film surround on the file to be scanned. During scanning it must ensure the file arranged in order and not missing pages, Guaranteed scanned image file: Clear and complete, in the right direction, not skewed, black edges. Choose a scanning resolution, Should be in accordance with suggested of the "study archives digitization technical specifications" Black = 100dpi; Color = 200dpi, Scanners currently on the market can meet the standards. But if you want to used bar code labels

paging or scanning the archival, You must use color scan = 300dpi, Otherwise, the scanner will not recognize the RFID bar code labels.

Image verification: Checking each page of scanned image, check whether the same of pages between Physical and electronic document, if files pages upside down, we must exchange the pages in the right order; If it has missing part of the scanned files, we must added Scan; and if an image is not clear, we must rescan and replace the original scanned image.

Image processing: Processing the defect of scanned image file, including: Correction, decontamination, image stitching, cutting edge processing and OCR text recognition. It should be noted that, OCR recognition accuracy depends entirely on the image quality and the use of text font, If the image quality is high, recognition accuracy is high.

Image storage: Image Storage mainly to the scanned image files stored in the computer hard disk, for future access to the system to use. The focal point of the step is selected image storage format and named images.

Selected image storage format, the contents of the main consideration including: Degree of fidelity of the original file, size of store the file, Format versatility and standard degree (Zhang, 2003). Due to long-term preservation and high reliability, the electronic documents in PDF format can be recommended by experts as the best format for image storage (Liu, 2002). But we should choose TIFF or JPG format in the hard disk and optical disk backup. Use storage in a variety of different media formats help the preservation of electronic documents and resistance to outside information environment changes.

Whole images naming and storing, can be used computer scanning systems and bar code identification system automatically processed.

Indexing database: You can find the Index database metadata information, including various types of descriptive entry content. If the archive management system has entered the entry information, Common data format export the index of directory. No input files entry information management system, according to the data format system into production index directory.

Check the directory information: If you do not have a directory of archive, Archives administration system should be follow the rules of archives cataloging, DA/T18, for cataloging entries, field length and content of the archive directory. Currently, more scientific approach is Used bar code management system is integrated with the scanning system (Sun and Cheng, 2011).

Data articulated and test: Index and image database of digitized Form conversion is in progress, use the network Load data into the server-side on time. Through the preparation of program or with the software, can be achieved automatically search for catalog data associated with the digital image, Adding the corresponding electronic address. After image information imported we must accessing and testing in sampling of type. Ensure that the images data can be correctly circulate and access in archives management system.

Data acceptance and backup: We must check index data of completion of the digital conversion in sampling way. Including directory databases, image files and mount the overall quality of the data. All the procedure need for test report with the corresponding.

Electronic image data of vetted and approved must backup at once. To ensure data security backup carrier must be diversifications, we can take Implement multiple sets of backup combination of online and offline and attention to ex situ conservation.

Binding and restore: The last, take the archival material which has been complete burning form the verification of Image storage place, In accordance with Sequence checking and sorting, consistent must be done with achieve which before scanning and re binding as the same as the type before the Binding.

SUMMARY

Archives digitalization is fundamental to archives information. Through the Jinan University's application and practice of archive digitalization workflow, we can see archive digitalization is not mysterious and tedious. By the means of scientific management and modern computer technology, parts of archive digitalization process could be done automatically (step 5, 6, 7, 8, 9). Jinan University has a work-study program which helps needy student, so some steps which need manual process could be finished by employing some students from work-study program which saves costs. We have investigated over 20 archive digitalization outsource company and their prices are between 0.5 Yuan and 1 Yuan per black-white page and between 3 Yuan and 4 Yuan per color page. But, in our practice, we only need to pay some commodity cost and work-study program wages which gains economic benefits and saves funds in the department.

But we must be reminded that all the digitalization process must be supervised and involved by the archive specialist and especially the inventory and quality assurance must be done by the archive specialist which guarantees the electronic archive information correctness and validity. Of course, the digitalization is a long-term job, the workflow and process devised in this study may have some fallbacks as the time goes and information technology develops, so this process my needs to be improved in the future.

ACKNOWLEDGMENT

This work is supported by the Fundamental Research Funds for the Central Universities and 2013 Annual National Archives Technology Projects, No. :2013-X-10 and The Fifteenth Installment of Education Reform Project No.: JG2013041.

REFERENCES

Ding, Y.Z., 2009. Discussion on university library archives digital. Archives, 4: 41-42.

Guan, M.L., 2010. University archives digital construction problems and countermeasures. J. Harbin Inst. Vocational Technol.

Jin, S.F., 2008. Use barcode technology to modernize archives management. Lantai World.

Li, J.Y., 2009. Discussion on the digitization of the construction of university archives. Sci-Tech. Inform. Dev. Econ., 20: 97-98.

Liu, J.Z., 2002. File format PDF document. Arch. Sci. Res.

Sun, Y.F. and Y.D. Cheng, 2011. A management system of digital archive based on RFID. Value Eng., 23: 139-140.

Wang, R.Z., 2009. Thinking of strengthening university digital archives. Yunnan Arch., 8: 42-43.

Wang, X.P., 2011. Talking archives practice of digitization and development strategy. Arch. Sci. Bull., 6: 54-57.

Xu, C.C., 2011. Archives digitization process OCR technology. Arch. Manage., 1: 38-40.

Xue, S.X. and Z. Huang, 2004. Archival information system integrated management platform. Arch. Constr., 4: 21-23.

Zhang, S.L., 2009. Archives digitized intellectual property strategy-to digital Dunhuang project as an example. Arch. Sci. Bull., 3: 50-53.

Zhang, Z.Y., 2003. Digital files select a file format. Arch. Sci. Bull., 6: 41-44.