

Microfilming and digitization of newspapers in China

Chunming Li, Wei zhang

National Library of China

1. Introduction

Newspapers can preserve vast amount of first-hand historical information for research, and are reliable and valuable sources for studies on politics, economy and culture of a country and a specific period of society. Mr. Fang Hanqi, Chairman of the China News History Society and Professor of Journalism Faculty at Renmin University of China, once said that Newspapers are the draft versions of the history and newspapers are very important reference materials for historical research. It is critical for the newspapers collectors and researchers to pay attention to collecting, storing, and utilizing the newspapers of all times and to make good use of information in the newspapers for historic research.

China's history journalism could be traced back to Di Bao, a kind of news bulletin in the imperial court of the Tang Dynasty (618-907). With the economic development, there appeared the "Kai Yuan Za Bao" (713-734) which was a newspaper hand-written on silk. The progress of the history accompanied with the cultural development. Since the Qing Dynasty (1644-1911), a lot of influential Chinese-owned/managed newspapers had emerged, the well-known of these included "Shen Bao", "Xin Wen Bao", "Zi Lin Xi Hu Bao" etc. "Shen Bao" was published for 77 years (1872-1949) and released altogether 25600 issues, made the record in the Chinese newspaper history. Currently, there are around 1900 published newspapers in China. According to "China Newspapers Development Report" which was published in 2005, China's total published daily newspapers volume has been the world's number 1 in the past five consecutive years.

Notwithstanding the valuable information in the newspapers, it is a big challenge to preserve newspapers. On the one hand, formulated to be inexpensive and expendable, newsprint is manufactured with large percentages of unpurified wood pulp which contains impurities that remain in the paper after processing. These impurities, when exposed to light, high humidity and atmospheric pollutants, promote discoloration and acidic reactions in the paper. Acidity causes the paper fibers to weaken and break, and is the major culprit in causing the paper to become brittle. For instance, National Library of China has collected a series of "Shi Bao" which was published in 1922. Over the past decades, the collection could not be used any more. On the other hand, the publication format of the newspapers has not changed in the past centuries. The design of the

newspapers remains the following features: lots of information, a number of pages, and the short publishing cycle, etc. As a result, preservation of print newspapers needs to occupy a huge space for storage. Also it makes it difficult to retrieve the valuable information.

At present, the only way to preserve newspapers over the long term, while allowing access to the content, is to reproduce that content in another medium. Microfilm is the best surrogate available today. Microfilm will last for hundreds of years if it is stored correctly.

The past decade witnessed rapid new developments in digital technology. The Internet opened up new perspectives for accessing information. Information became more and more digitally based, and finally digitization offered new possibilities for creating surrogates of newspapers.

2. The application of microfilming for long term preservation of newspapers in China

The National Library of China was the first organization in China to introduce the microfilming technology; it imported its first set of 35 mm microfilming machine back in 1948 from United States. Nevertheless, the real beginning of microfilming application was in 1960s. The library microfilmed the "Da Gong Bao" (1902-1966), "Xin Wen Bao"(1893-1949) and "Yi Shi Bao" (1900-1920). At that time there was no corresponding standards to regulate the equipment and materials, and there was no corresponding requirements and procedures on the techniques and operations. Therefore, the quality of density, resolution and contrast of those microfilms was not good enough to satisfy the reading and usage requirements. Staff jokingly called microfilms of that period as "black wok bottom". In 1985, the China National Microfilming Centre for Library Resources (located in National Library of China, referred to Microfilming Centre) was established on back of the support from Chinese government . The goal of the Centre is to rescue and protect the national culture heritages. 22 libraries and dozens of documents depository units participated in the rescue program, and began the research of standard and systematic microfilming. As of the end of 2004, the Microfilming Centre had microfilmed 3,704 titles which were published before 1949 (the founding of People Republic of China) and created 13,985 reels of microfilmed newspapers and 9,723,392 shots. In addition, out of the 700 titles of newspapers which were published after 1949 that need to be rescued, the Microfilming Centre has microfilmed 612 titles and created 12,895 reels and 7,607,593 shots.

2.1 Establishing of a national standard system for microfilming

The aim of establish a standard system is to ensure standard, systematic, and high quality microfilming work. This would also help make sure that the microfilm be preserved longer to facilitate the usage of microfilm and exchange of information. National Standardization of Microphotography Technology Committee (renamed Chinese Document Imaging Standardization Technical Committees in 1999) is officially established in 1987. The committee is the organizer of standardizing the microphotography technology in China. The committee is corresponding to the International Document Management Applications Committee (ISO/TC171). It is comprised of secretarial division and 4-branch committees. Its main task is to establish rules, construct working plan and responsible for the research, drafting and revise of standards. Since its establishment, the committee has introduced many international standards such as ISO6200, ISO6196, ISO4087 and others for constructing national standards. Currently, Standardization Administration of China has officially promulgated 49 national standards for microfilming.

2.2 Microfilming method of Chinese newspapers

Through years of practice, the Microfilming Centre developed a set of their own microfilming method of Chinese newspapers. It can be divided into 4 steps: ① Preparation stage: Preparation is important to make records ready for shooting. If the same title of the newspapers is not collated appropriately, the newspapers may be filmed into different part of microfilms. That will cause great difficulties to later uses. The preparation work consists of the following activities: repairing the newspapers, rearranging the layout of the newspapers and considering the shooting procedure which includes the volume of film, shooting and ordering. ② Shooting stage: to ensure the quality of the microfilm, trial shooting is necessary before the actual shooting. The shooting must be proceeded according to the related technical standards. In addition, the trial shooting should include the printing of the microfilmed film and copying the microfilmed film. Then adjust the techniques of shooting, printing and copying based on the technical indicators of the trial shooting. ③ Printing stage: Testing the water temperature, confirming the formula of prescription and the printing speed. ④ The microfilm copies are used for reading.

2.3 The rules of building metadata for the microfilmed newspapers

According to the "International Standard Bibliographic Descriptions" (ISBD) and the national standard "Bibliographic rules of Non-book information GB3792.4-85", the Microfilming Centre has established the rules of building metadata for the microfilmed product. The Microfilming Centre has published the "Bibliographic rules of microfilmed

product by the Microfilming Centre". In which, the "Bibliographic rules of the microfilmed Chinese newspapers" is focused on the creation of metadata for newspapers. The description of metadata consists of the specification of the film, the reduced rate and the preserved information, the remark of missing pages, etc.

2.4 The preservation methods for microfilmed newspapers

In order to preserve microfilms effectively for a long period of time, it is essential to formulate corresponding preservation methods. The major factors includes: ① the preservation of the master film: the master film is used for long-term preservation. The information in the microfilm is transformed into a microfilm intermedium (also known as production film) to produce duplicate copies. The duplicate copy of the production film is called Positive film or Working films for circulation. Whenever the Working films is damaged or appeared to have problems, the production film is used to reproduce the new working films. The master film is not allowed for duplicating. ②The requirement for the preservation environment: The rate of deterioration of the microfilms can be slowed down in appropriate conditions. According to international standards, the preservation temperature is under 20°C and humidity is under 40%. The environment should be anti-dust and anti-poison gas. ③The construction of the storage site: e.g. The Microfilming Centre has constructed a national master and duplicate microfilm depository underground for preserving microfilms. Special compact shelves for storing microfilms are designed. However, the "vinegar syndrome" poses a serious threat to a large portion of microfilms. It is subject to slow deterioration, releasing acetic acid. Accompanying this degradation are many visible signs of physical deterioration – shrinkage, channeling, buckling, the appearance of crystals and bubbles on the surface of the film – and ultimately, loss of the information carried on the film. The Microfilming Centre has started using polyester films instead of cellulose acetate film since 1996. Polyester films are more stable and have a longer preservation time. At the same time most of acetate films were deacidified.

3. The development of digitization of newspapers in China

The emergence of microfilming technology has alleviated two big problems encountered in the preservation of traditional print newspapers: lifecycle and space. Nevertheless, it does not solve the problem of accessing and sharing. With the development of digital technology, the latter problem is fully resolved. Digitization can provide effective tools and methods to the retrieval and use of information through the change in preserving, managing, disseminating and using of resources with the transmission of documents via the Internet. As a result, it expands the retrieved scope of information and improves the efficiency of information access.

3.1 Existing difficulties and solution of digitization of Chinese newspapers

Although digitization of newspapers has a lot of advantages, it is uneasy to carry out. First of all, the changes of newspaper layouts and fonts are very complicated. For instance, the layout of a regional newspaper was changed from the vertical typesetting to the horizontal typesetting. The font was changed from traditional Chinese characters to simplified Chinese characters. In particular, there are complex cases of mixed typesetting of traditional Chinese and simplified Chinese characters, which makes OCR (Optical Character Recognition) difficult. Secondly, the amount of a newspaper is very large. A regional newspaper has about 100,000 pages counted from its first issue until present. Thirdly, the digitization process is difficult especially when digitizing directly from print newspapers. Since the original carriers have deteriorated, turned yellowish and worn, which makes it difficult to read, there must be some repairing works before the scanning. Moreover, the cost for digital processing and long-term preservation is enormous. If a 50-year-old press agency wants to digitize all newspapers it collects, the total cost would be over 300,000 dollars. The cost of long-term preservation also includes the cost of data migration, data maintenance, etc.

At present, some professional data processing companies in China provide a number of technical solutions to the digitization of Chinese newspapers. For example, Beijing WINTONE Info technology Ltd., Datum Data Co. Ltd., and Green apple data center Co. Ltd. have developed their own digital system for newspapers. The systems include Chinese OCR technology, electronic layout recovery technology, digital production process control, retrieval system, and dissemination system.

3.2 The digitization progress of Chinese newspapers

The digitization of newspapers in China experienced two development stages: The first stage was in the 1980s: the production of newspapers became digitized. Every press agency introduced computer laser typesetting and offset printing technology to replace letterpress-printing method. The second stage was in the 1990s when humans entered the Internet era: press agencies established some integrated news business network and tried to digitize the production and transmission of news contents. At the same time, newspaper collecting agencies and professional data processing companies used digitization technology to produce newspaper databases. The first press agency that tried digitization was "Hangzhou Daily". In December 6th of 1993, "Hangzhou Daily · Afternoon Edition" has launched the online service. In October 1995, "China Trade News · Electronic Edition" begun to provide Internet access services. Later on, "Guangzhou Daily", "People's Daily", "Guangming Daily", etc., went online. The development of CD-ROM edition of "China Information World", "Economic Daily-16 years", "42 Years of Reference News Full-text Database", "Wenhui Bao 60 years

Newspaper CD-ROM”, “Reprographic Materials Full-text Database of RenMin University of China”, and the online edition of “WiseNews”, “Chinese Important Newspapers Database” developed by Tsinghua Tongfang are relatively good examples.

3.3 An investigation on the online services of digitization of newspaper

From January to June of 2006, the National Library of China has conducted an investigation on the online services of digitization of Chinese newspapers. The study shows that there are already 800 titles of online newspapers which account for 42% of print newspapers published in China. These newspapers are distributed throughout the country. The contents cover wide range of subjects. The publication frequency of newspapers includes daily, bi-daily, weekly, etc. The type of publication includes newspapers that are publicly issued and limited issued such as internal newsletter. The typesetting and formats of online newspapers are diverse. There are 243 online newspapers providing the original newspaper format which represents 30% of online newspaper. The common formats include PDF format, JPG format, CEB format (developed by Founder Group), SWF format (developed by Macromedia) (See Table 1 for details). Apart from the above electronic newspapers with graphical versions, some press agencies have issued a type of newspaper with simulation format. The simulated version of newspapers does not only retain the original image of the newspaper, but also provides related links to the text format. Readers can easy browse the contents.

File Format	Qty	Example	Website
Text Format	673	Workers' Daily	http://www.grrb.com.cn
PDF Format	214	Huaxi Dushi Bao	http://www.wccdaily.com.cn
JPG Format	19	Dalian Wanbao	http://www.dlwb.com.cn
CEB Format	12	Sichuan Quality News	http://www.quality.sc.cn
SWF Format	1	Jinri Xinxi Bao	http://www.jrxxb.com.cn
Simulation Format	51	National Business Daily	http://www.nbd.com.cn

Table 1 Statistics of data format of electronic newspapers

Currently most online newspapers provide services free of charge for the public, while only a few newspapers charge users for services. Some press agencies collected their own newspaper to create databases, such as “People’s Daily” and “Legal Daily”, etc.

3.4 Future development of the newspapers digitization

At present, all capable press agencies and collecting institutions in China have attempted digitizing newspapers, including the digital preservation and service of the current newspapers, as well as the digital processing of the old newspapers. The authors realize that the digitization of newspapers in China is still in the beginning stage and main concern for providing services. So far few of the agencies thought over the long-term preservation issues in depth. Although there are some well-developed digitization

techniques and international standards such as the long term preservation of digital information - OAIS (Open Archival Information System), the standard for digitized information formats, related descriptive data of digital information, the standard for encoding descriptive, administrative and structural metadata – METS (Metadata Encoding and Transmission Standard), and standards for publishing ebooks – OEBPS (Open eBook Publication Structure), more thorough researches should be done on how to digitize newspaper in China. The fragmentation of the development of newspaper digitization that happened in the early stage and the waste of human and capital resources due to not following standards should be avoided. At the same time, the technology of knowledge discovery and management should be considered. The digitization of newspapers can be processed successfully only after the above problems have been resolved.

4. Digitization of microfilmed newspapers

Digitization of microfilms is very important to digitize precious old newspapers. It has certain superiority comparing with direct digitization. The digital system for microfilms mainly comprises of computer, microfilm scanner and related software. One of the problems of digitizing microfilms is that not all microfilms can be recognized using OCR. For example, the Microfilming Centre has used OCR to recognize words in the microfilmed newspapers and rare books, but the accuracy rate is very low and not exceeding 70%. The reason for a low accuracy rate is mainly due to the long retention period of the original documents. The paper carriers deteriorate and they deform the characters, such as missing strokes of characters, broken characters, blurred characters and variant forms of characters. Another reason is related to the reduced ratio of microfilms. Microfilms with a low reduced ratio have relatively larger characters for easier recognition.

The Sun Yat-sen Library of Guangdong Province has introduced MicroDAX300, the electronic image system of MINOLTA of Japan, including MS3000 microfilm scanner and PowerFilm digital image processing software and TRS full-text retrieval system (made by TRS Information Technology Limited, Beijing) to establish the digital system for microfilms. After years of efforts, there are about 400 titles of microfilms published before 1949 have been digitized. Besides, many organizations such as the National Library of China, Shanxi Library, Zhejiang Library, Shandong Library, etc., have introduced digital devices for microfilms to start the digitization of newspapers.

5. Conclusion

It is clear that both microfilming and digitization have their own qualities and purposes: microfilming is well suited as a surrogating method for preservation. It is reliable and

relatively cheap. From the beginning, digitization is mostly applied in the field of access. It is relatively expensive, but it offered a number of new features, such as color, Internet browsing and search facilities. It is not considered as a useful preservation method because of the rapid obsolescence of maintaining and retrieving techniques. Also the durability of digital storage mediums has yet to be proven. Therefore, microfilming and digitization both can be embedded in the strategy of libraries and other heritage institutions. This demands a holistic approach. Ultimately, despite the seeming contradictions, microfilming and digitization complemented after all.

The microfilming and digitizing of Chinese newspapers have decades of history. Many cultural heritages are being preserved and protected, and resources can be retrieved and used. We realize that in the digital library era, we have to strive for satisfying the need of long-term access and users' needs by utilizing advanced technology. More importantly, we are of the opinion that the technological accomplishments are the basis, while policies and funds are the guarantee. The only way to preserve national cultural heritages effectively is to raise the government and the public's awareness, to establish national long-term preservation policy, to seek for the active participation of all levels of government, publishers, libraries and archives to form a complete preservation network.

Acknowledge contributions from my colleagues Ms. Wei Zhang, Ms. Meng Wu, Dr. Ben Gu and students Cheng Kit Ying, Tam Shuk Ying from University of HongKong
Acknowledge Jian Li, director of the Microfilming Centre and Mr. Hongbo Yang

WEB SITES REFERRED TO IN THE TEXT

1. YANG BIN Trend of Newspaper Digitization in China
<http://www.ifla.org/IV/ifla71/papers/156e-Yang-Bin.pdf> Accessed 2006.5.20
2. Newspapers http://www.nli.ie/co_newsp.htm Accessed 2006.5.20
3. Preservation Microfilming Guidelines: Newspapers
<http://www.srlf.ucla.edu/PI/techdoc/newspapersPrep.htm> Accessed 2006.5.20
4. Best of Both Worlds <http://webdoc.gwdg.de/edoc/aw/liber/lq-2-03/164-168.pdf>
Accessed 2006.6.27