2-1-2008

Analysis and Digital Processing of the 1911-1949 China Literary Collection

Meng Zhan
Fei Yu

Follow this and additional works at: https://scholarsarchive.byu.edu/jeal

BYU ScholarsArchive Citation
Available at: https://scholarsarchive.byu.edu/jeal/vol2008/iss144/5
ANALYSIS AND DIGITAL PROCESSING OF THE 1911–1949 CHINA LITERARY COLLECTION

Meng Zhan
Wuhan University Library, China

Fei Yu
School of Water Resource and Hydropower, Wuhan University, China

Summary: This paper analyzes the properties of print publications dating from the period of the Republic of China, 1911–1949. It discusses aspects of paper, binding, publication, contents, the present situation of use, preservation, and their distribution in mainland China. Then it discusses the digital processing aspects of the Collection. This paper also elaborates work flow management and quality control methods in digital processing. The research is part of the China-America Digital Academic Library Project (CADAL).

1 The Collection Published between 1911–1949 and analysis of its properties

1.1 The period of Republic of China (1911–1949) and the Collection Published between 1911–1949

The 38-year period from the Revolution of 1911 to the founding of People’s Republic of China in 1949 is defined as the Republic of China. This period, which ended imperial power and led to a republic, spanned ancient China and modern China. This historical period of transformation and reform saw much conflict between old and new ideas brought about by cultural exchange between home and abroad, and a wealth of literary and historical writings were produced. (Li Fang, 2005, p. 109) Additionally, due to cultural reform and innovations in printing technology, many useful and important publications sprang up that made use of new printing materials and technology. Many printing presses, newspaper offices, and libraries were established, as well as library special collections and archives. For these reasons, publications from 1911–1949 are valuable not only for their historical and literary content (Wang Xiangfeng, 2005, p. 63), but also as examples and artifacts of the revolutionary improvements in the history of printing such as papermaking, printing, and binding techniques. It is reported that over one-hundred thousand varieties exist of some kinds of books published in this period. (Li Fang, 2005, p. 109). Other types of publications such as periodicals, newspapers, magazines, pamphlets, and political tracts issued by government and nongovernmental publishing companies also flourished. All of these played a significant role in the development of Chinese culture during the Republican period.

According to the definition given in "The Catalogue of the collection published in 1911–1949," books published from 1911 to 1949 belong to the 1911–1949 Literary Collection (Li Fang, 2005, p. 109). From this, it can be inferred that the only criterion used to define the Collection published from 1911 to 1949 is "publishing time," because this period was a turning point from the ancient to contemporary in China. The literature in this period was both similar to Chinese ancient books while at the same time exhibiting great differences. For instance, as to paper, it not only had the Xuan paper produced with ancient traditional technology, but also had strongly acid paper produced with new technology; as for types of edition, there existed both handwritten editions and printed editions; as to binding, there coexisted thread-bound editions and machine pack editions; as to content, the writing in classical Chinese and vernacular were both widely accepted. The co-existence of ancient and modern features in this Collection exemplifies the extraordinary social conditions during the transition from old China to modern China.

1.2 Properties of the Collection

1.2.1 Properties of paper

At present, publications from 1911–1949 preserved in libraries are liable to be old, hard and brittle. Wuhan University Library, as part of the CADAL preservation project, processed more than 17,000 periodicals published during this period. Before being processed, the paper of these periodicals was generally brittle, hard, and yellowed. More than half of them appeared to be worn to some extent. Compared with the ancient books preserved in our university library under the same external circumstances, the damage to these Republican publications was more serious. Why would these much
younger books suffer more serious damage than the much older ancient books? As indicated in the findings by the National Library through the study “The Investigation and Analysis on the Paper’s Acidity of the Library-stored Literature and Their Present Preservation,” the Republican materials mostly used mechanical mill paper produced with wood pulp, which resulted in an acid paper of poor quality and short life (Jiang He, 2005, p. 10). In addition, the paper was also hard and crisp, and tended to be brittle to the touch. In contrast, paper used in ancient books was produced by Chinese traditional papermaking technology, handmade using extruded fiber and the pliable and leathery parts of plants such as hemp and bark. Paper made with these materials and methods tended to be neutral and alkaline, soft but durable; as the saying goes, “the life span of paper is up to million years.” For example, although the books in Song and Yuan dynasty have lasted as long as one thousand years, they are still readable.

According to a study done by the National Library, we know that storage life of newspapers and books published from 1911 to 1949 is about 50-100 years to 100-200 years (Jiang He, 2005, p. 10). Based on these figures, it can be assumed that newspapers of the Republican era are approaching the edge of natural extinction, and books of this period are also rapidly aging. During the course of processing the collection, we took care to prevent the items from being destroyed. For example, we used a vertical scanner that did not touch the surface and emphasized light while moving the item. Even so, 70% of the items still needed repair after digital processing. These publications are very brittle, and their life span is limited.

1.2.2 Properties of Binding

Through our daily experience at our university library, we learned that the ancient books were completely sewed by hand, while modern periodicals are bound by machine. In the Republican period, the technology of mechanical papermaking and printing imported from the West (in that time called “international technology”) was more and more used, and the binding of books was also transferred from ancient handwork to machinery. During the early years of the Republican era, publications using Xuan paper in general maintained the technology of hand sewing, now called thread-binding.

In the later stages of this period, all publications were machine-bound. Therefore Republican publications include both hand-bound and mechanically-bound items. The technology of mechanical binding in the initial stages of the Republic of China was simple, and the binding materials were inferior. Hence, binding quality was poor, and pages pulled loose from the binding over years of use. During the course of digital scanning in the CADAL project carried out by our university library, this sort of damage (called “scattered pages”) affected up to 20% of the materials, and six volumes of this amount were books “broken up when touched,” that could not be restored at all. This phenomenon of “scattered pages” occurred both due to poor paper quality and also due to poor binding.

1.2.3 Properties of publishing

The print record records incidents, culture and ideas of every historical period. In this special time period of the Republic of China, publications developed in unprecedented ways. Publications dating from the Republican era display the following characteristics:

Firstly, publishers were diverse. They were not only formal publishers such as press and print houses, but also informal publishers such as organizations, political parties, and the government in the broader sense. Additionally, there were offprints, printed books and handwritten editions of the works of well-known authors. This latter type of publication, published by informal publishers, makes up the majority of publications from the Republican period.

Secondly, because they made use of the wealth of new publishing technology available during the Republican era, editions from this period are characterized by use of a diversity of printing methods, such as mimeograph editions, letterpress print editions, handwritten editions, woodcarved editions, lithographic editions and ormolu (imitation gold leaf) editions. By the later period, letterpress printing accounted for a large proportion of editions.

1.2.4 Properties of content

The Republican period was an exceptional period in Chinese history. An exchange of culture between home and abroad contributed to the collision of ideologies and resulting innovations. This period was a
time of cultural ferment in China, just as the Spring and Autumn and the Warring States Periods were. Books of different political opinions, academic viewpoints and ideas were published in this period, some advocating maintaining the old social order, while others promoted democracy, revolution, and a new culture. The most typical schools of thought were the advanced ideas of the New Culture Movement advocated by Lu Xun, the New Democracy advocated by Sun Yat-Sen, and Communism advocated by some revolutionary representatives such as Mao Zedong. Additionally, all kinds of publications were produced that expressed the politics and ideological principles of the government of the Republic of China, including new technology and scientific publications as well as Western-oriented publications. Clearly, Republican-era publications were not only various in type but also reflected the diverse society of the period.

2 Analysis of the Status Quo of the Collection

2.1 Status of the Distribution of the Collection

Publications from the Republican period were widely distributed and may be found in archives, libraries, historical archives, and literature and history research institutes. Some individuals also collected publications from this period. The Second Archives of China holds the most volumes of publications, which vary in content, from the Republican period (Wang Xiangfeng, 2005, p. 63). The National Library of China, public libraries in different provinces, and archives also have extensive collections. University libraries such as the libraries of Peking University, Nanking University, Fudan University and Wuhan University are also rich in books from Republican China. In addition, unpublished theses, usually handwritten, by students of these universities during the decades of the Republican period have also been preserved. These writings have their own unique historical and societal value.

2.2 The Dilapidated Condition of the Collection

The high acidity of the paper of Republican-period publications contributes to their aging and deterioration, and papers have even lost their mechanical strength, tending to crumble when touched, and many are not readable. The following chart gives enlightening figures.

<table>
<thead>
<tr>
<th>Name of library</th>
<th>Total # of 1911-1949 volumes</th>
<th>% damaged</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Library of China</td>
<td>670,000</td>
<td>90%</td>
<td>100% of earliest vols. damaged</td>
</tr>
<tr>
<td>Jilin Library</td>
<td>160,000</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>Nanking Library</td>
<td>700,000</td>
<td>60%</td>
<td>Books crumble to the touch.</td>
</tr>
<tr>
<td>Chongqing Library</td>
<td>170,000</td>
<td>50%</td>
<td>Extensive damage</td>
</tr>
<tr>
<td>Wuhan</td>
<td></td>
<td></td>
<td>50,000 nearly lost</td>
</tr>
</tbody>
</table>

As this chart makes clear, most publications dating from the Republican era that are still extant are in a sorry state of preservation. The low quality of the paper used makes them subject to serious damage. It is high time that we adopt methods to rescue and protect these publications.

2.3 Accessibility of Republican-era publications

For various historical reasons, books published between 1911 and 1949 in the mainland of China are not all available to readers. Only some are accessible in rare books collections. Others are restricted to consultation by authorized researchers. The policy of “rare books” was first proposed by Mr. Zhao Wanli, who was the director of the Department of Reliable Texts of the National Library in 1953. Criteria for inclusion in the Republican period collection were the ideological, historical, and artistic properties of the text (Zhao Changhui, 2004, p. 22). Although the publications included in the “rare books” category covered the whole period of the Republic of China (time of publication), all libraries still considered the revolutionary nature and ideological content as criteria in collecting. Consequently, only some of the books in the collection are accessible to readers. Some libraries that have not yet established rare books reading rooms now place some books which are in accordance with this standard in modern literature reading rooms to be accessible to readers.

3 Research into digitization of Republican-era publications

Given the special nature of publications dating from the Republican period, we can say that accelerated
digital processing will facilitate the preservation of literature and increase use. But important questions remain: what standards should be adopted for processing of these publications, and how can we control the quality? Our university library took part in the development of the project under CADAL, and during this time, we also conducted some research on these publications, including how to determine the requirements of the digital process, processing standards, and methods of quality control.

3.1 Estimation of the requirements of the Digital Process

3.1.1 Estimation of the Digital Process

There are two points related to the digital requirements of 1911-1949 publications: the first one is digital preservation, and the other is users’ network. Take our university library as an example. The library has nearly 30,000 volumes published between 1911-1949, only 85% of which can be put into the reading rooms for readers to read. Because Wuhan University specializes in the disciplines of literature, history and philosophy, there is heavy use of Republican era publications from both teachers and students. But because of paper and binding, this literature has been seriously damaged, and cannot be placed in open access to readers. Therefore it is necessary to do digital processing so as to preserve them in closed shelves. Additionally, Republican-era publications are located not only in the library, but also in reference rooms of relevant colleges and departments. Therefore it would be beneficial to carry out digital processing that could be widely shared through the campus network.

3.1.2 Estimation of the Resources increment Service

In the library field, collection development policy considers increasing usage. For books published from 1911-1949, digitizing and online retrieval can offer more services than just increasing usage. Sometimes readers can also read the full-text contents through the network if necessary, thus providing increased access. The library's and readers’ efficiency are both improved.

3.2 Resource Selection Standard

We established a resource selection standard as follows:

Resource selection scope:
All print literature located in universities (including libraries, archives and reference rooms of every school and department etc.) are the resources of the database. Such a scope contributes to the construction of a sharing system of the Collection throughout the university.

Resource selection standard:
According to the criteria described in the “Bibliography of Republic Period of China”, it is known that books in any form of binding published between 1911 and 1949 are technically candidates for inclusion in the digital Republican Era Collection. Our university library set the resource selection standard to be based on the year in which the book was published. Therefore the collection includes not only traditional thread bound books, but also paperbacks; not only the books typed vertically, but also the format of horizontal writing; not only the books written with ancient text and complex Chinese characters, but also the ones written with modern words and simplified Chinese characters.

3.3 The Application of Digitalization Standards

The management center of CADAL drew up a series of criteria for digitization of resources in the collection. It also adopted some open electronic book standards for creating and conserving digital resources. 600dpi was accepted as the scanning resolution, XML was exerted in catalog navigation, and image files were saved in TIFF and DjVu format. This way of adopting current standards is different from earlier projects. For example most early domestic electronic books needed special readers. Also they were produced with low precision. In contrast to this, objects of the Republican Digital Collection are produced according to current, convenient standards. These standards can also facilitate the conservation of digital resources and its further development.

The management center of CADAL has established a metadata standard that adopted Dublin Core (DC) standards for digital processing. Our University amended this within the framework of CADAL, and applied it to digitizing the Collection Published from 1911-1949. For example, we increased the repeatability of
some useful elements in the set of metadata, and augmented the description of the title (for example, how to deal with the title of bound volumes, book series, and books in different languages, etc.). We also made specific regulations for “scope of space-time” and “punctuation.” These revisions not only utilized the framework of standardized files in CADAL, but also allowed for the particularity of the data among publications of the Republican era. This sort of metadata will improve the effectiveness and quality of data exchange and shared services in the future.

3.4 Management of Digital Processing and Quality Control

3.4.1 Management of Digitizing

During the course of digital processing of the Collection, there existed two work flows, namely (1) scanning and metadata labeling through a professional company and (2) quality control and classification by the library staff. Processors in the company are very specialized and efficient in scanning and conducting digital processing (such as error checking and removing unwanted information), while our university library staff are experts in the aspects of quality checking (textual and metadata). This work flow ensures that the specialties of different organizations are fully used. It not only increases the scanning rate and quality but also guarantees the quality of metadata.

3.4.2 Process quality control

Scanning quality control: As we have outsourced scanning we have developed a “digital processing standard” (including the requirements on scanning precision, image decontamination and scanning integrality) and made this a part of the legal agreement between the company and the University Library. When the company finished scanning and execution, the examiners of our university library checked the processed materials in detail. If an error was found, the company was asked to reprocess the data. Because of these strict measures, each publication appears very clear on a computer monitor, sometimes clearer than when it is seen by the naked eye.

Metadata quality control: Because they directly affect document retrieval, special attention should be focused on creating metadata. As mentioned above, our university library amended the metadata standards stipulated by CADAL to record metadata for the Collection. We extended some useful fields and changed necessities (such as making the keywords field essential and setting the minimum number of keywords at 3). The Metadata marking work flow has 2 levels: the first one is metadata markup by a digital processing company according to standards (the main key words are not required), and the second one is checking up and marking some keywords on the processed metadata by the cataloging department of our university library. In order to control quality of the processing company’s metadata, our university library has drawn up “testing regulations of the metadata” for the Collection Published in 1911~1949, which specify not only technical requirements but also establish an error-based payment subtraction formula. In this formula, different fields have different subtractive requirements, and when the total value of the subtractive points reaches a certain point, deductions will be made from the company’s processing fee. Under these double constraints of technology and finance, quality of the Collection metadata processed by our university library is well secured.

4 Conclusions

Publications printed from 1911 to 1949, during the Republican Period of China, are the product of a specific historic period of China. Because of limits on publication quality of the period, their paper is close to natural self-destruction, so it is urgent for us now to protect them by means of digital management. The work of protection not only has useful practicability for global information sharing but also has historic significance in protecting the human cultural heritage.

Properties of these publications raise some study questions for literature preservation departments (such as libraries) to think about: how to draw up collection development standards, how to determine rational work flows and how to control scanning and metadata quality in the course of digital processing.
References: