An investigative report on current long-term digital preservation situation among major Chinese libraries

ZHANG Mei*, LI Lin, ZHANG Xiaolin & LIU Xiwen
National Science Library, Chinese Academy of Sciences, Beijing, 100190, China

Received Nov. 21, 2008
Revised Dec. 9, 2008
Accepted Dec. 30, 2008

Abstract The paper reports a survey on the attitudes, arrangements, and operational model of more than 40 major Chinese libraries (CL) on long-term digital preservation. It reveals that digital preservation becomes an urgent concern for these libraries in our survey. Most of these libraries take a pro-active approach to the issue and most of them are in favor of participation in a certain collaborative preservation system, though a few still remain in a wait-and-see posture.

Keywords Digital resources, Long-term preservation, Libraries

1 Background

With the development of information technologies and Internet, digital resources have become more and more popular in the scientific and research fields. In fact, they are beginning to exceed the amount of printed resources holdings in many libraries. Some of them, especially the STM journals and proceedings, have taken the e-publishing as the main publishing medium. The monographs have also entered into the arena of digital publishing.

As the digital resources are becoming the main type of scholarly communications for scientists and scholars, most of the research and education institutions around the world have assumed the digital resources as one of the most essential service, if not the most essential service, provided by their libraries. Meanwhile, the academic libraries and special libraries abroad have been developing scientific and technological documentation system mainly for digital resources. For instance, Engineering Library of Stanford University will stop subscribing the printed resources in 2009[1], while in China, the major research and education institutions have also developed policies regarding digital materials as the main resources for their informational needs, and reduced the subscription of printed ones.

However, because of the highly dependency nature of the digital resources on the technical environment and acquisition techniques, its usability has been liable to natural disasters, technical obsolescence, economic loss and legal restrictions and

* Correspondence should be addressed to Zhang Mei (E-mail: zhangmei@mail.las.ac.cn). The authors would like to express their sincere thanks to Prof. Charles C. Yen for the language revision.
anyone of which may pose enormous challenge to digital preservation. Specifically in the first place, digital preservation encounters seriously technology difficulties, due to the instability of storage medium and the dependence on some special technologies which are constantly under rapid innovative changes and development, which may inadvertently shorten the life of the usability of digital information. Secondly, digital preservation faces severe archiving access problems. Currently, the common practice of acquiring digital resources is to gain access to data vendors’ databases through the Internet and to obtain the data vendors’ authorization on a fee-paid basis for users to gain remote access to view or to acquire their desired information resources, which separates the preservation issue from the subscription process. As a result, it is hard to obtain a long-term guarantee for the accessibility of the previously subscribed digital resources if the subscription is canceled at a later time. Thirdly, digital preservation has to deal with the complicated issues of usage and appropriate control mechanisms. It is a well known fact that digital preservation involves with complicated intellectual property right issues and especially with various stakeholders. The group subscription model makes it difficult to build up an effective and efficient preservation mechanism for the resources providers, the purchasers and their authorized users. Finally, digital preservation also needs to address the perennial hidden risks involved. As such, digital preservation requires complex and multifarious considerations for procedural, matters as well as for financial, technological and administrative support so as to insure its successful operation.

Nowadays, the long-term usability of digital resources has become increasingly a serious and universal concern of the society at large. Some of the developed countries or organizations have already taken strenuous efforts working on this issue. In China, related research on and experimental practices for digital resources preservation have been conducted by several major libraries. A few of them are currently working on the constructions of national digital preservation system.

There are basically three types of operation model in preserving digital resources, namely, institutional independent preservation, third-party preservation and co-operative preservation. In the institutional independent preservation model, the institutions should obtain the right to use the digital resources in the first place. Once the preservation operation is authorized by the rights holders, such as publishers, the digital preservation institution can then open up their door for business. Take British Library (BL) as an example. As a legal deposit library, BL sets up its digital preservation policy to ensure the perpetual access to the digital materials in its archived collections in accordance with their collection archiving policy[2]. The digital publications have been included in BL’s preservation scale, and the preserved materials are allowed to be used on-site only. The National Library of Netherland (Koninklijke Bibliotheek, KB) adopts this model in its e-Depot preservation service as well[3]. As for the third-party model, the preservation of digital resources is

---

performed by the third parties instead of the resource providers or its subscribers. This type of service is usually provided by large research libraries, data centers or commercial sectors[4]. The resource subscribers could reduce their long-term investment cost by tapping the expertise and facilities of the third parties, like Portico, a non-profit U.S. based firm for archiving digital materials launched in 2005. In the case of Portico’s business model, the publishers are required to sign an agreement granting Portico the license to ingest, archive, migrant and provide authorized access to their archived digital materials. Those libraries that have subscribed Portico’s service also need to comply with the Portico Journal Archive Licensing requirements in order to ensure the campus-wide access to the preserved materials under specific trigger events[5]. Besides Portico, Arts and Humanities Data Service (AHDS)[6] is another third-party preservation service. As for the co-operative model, different stakeholders co-operate on resolving the issues of risk and cost sharing related to digital preservation. LOCKSS program initiated by Stanford University Libraries adopts the co-operative model. The participating members use the LOCKSS software to build up their own box as a cache for collecting the resources from target publishers. The integrity and correctness of the content preserved locally is continuously compared to the same content collected by other LOCKSS boxes[7]. Kopal, a co-operative development of a long-term digital information archive in German, is another example of co-operative preservation model[8].

Researchers have been studying the various viewpoints toward the digital preservation by library administrators and their choice among different preservation models in recent years. In 2006, the Council on Library and Information Resources (CLIR) and Association of Research Libraries (ARL) conducted a survey of the directors of academic libraries in North America to investigate their attitude towards digital preservation by means of sending out questionnaires and conducting telephone interviews. They found out that the current licensing arrangements were inadequate to protect the library’s long-term interest in accessing electronic publications. Individual library cannot address the needs of e-journal preservation on their own. Many scholarly e-literatures were not covered by archiving arrangement. While e-journal archiving programs were becoming available, no comprehensive solution for the permanent protection from loss of such materials has emerged yet[9]. In January 2008, Portico and Ithaka developed a survey for library directors at four-year colleges and universities in the United States to examine their community’s current perspectives on the preservation of e-journals. The web-based survey was sent to 1371 library directors and received 186 returns. It revealed that while the academic libraries at large believed that the digital preservation was important, there was still significant confusion about how best to pursue it and how urgent it was. Many libraries seemed to be taking a wait-and-see approach, with some institutions relying on the result of actions taken by other libraries for the time being[10]. In the early 2008, Portico and the Association of Learned and Professional Society Publishers (ALPSP) carried out a survey covering more than 400 publishers to understand their current digital preservation strategies and practices. The response rate exceeded 65%. The key findings included that the majority of publishers realized the importance of digital preservation and they were strongly motivated to tackle
with the issue. Almost all publishers presented a strong desire for developing the best practice and industry standards. A significant majority of publishers indicated that they would in fact prefer other institutions to take the responsibility and national libraries in particular were a popular choice. While most publishers were clear about the distinction between issues of long-term access and long-term preservation, the majority of whom believed that they had an inherent obligation in securing long-term access for their clientele\textsuperscript{[11]}.

2 Aims, scope and methodology of our study

In order to promote effectively the development of a national long-term digital preservation system, it is essential to have an accurate understanding from the very beginning about the attitudes, principles, strategies and practice, including the preferred ways of handling long-term digital preservation matters of those information organizations. Accordingly, these authors have taken a survey, funded by National Science and Technology Library (NSTL), on those libraries that have relatively a larger budget for their programs of digital resources acquisitions, subscriptions, accessing and servicing.

Our survey was carried out from August to October 2008. The libraries that we included in our survey were of the following categories: the National Library, public libraries with rich digital resources, 985 academic libraries and some major special libraries. Although the survey was a non-random sampling, it covered the main libraries in China and therefore its findings could be revealing. The survey was sent to 63 library directors and received 44 returns from them (see Table 1). The response rate was 69.8%. All the returned questionnaires were valid.

<table>
<thead>
<tr>
<th>Type</th>
<th>Amount</th>
<th>Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special libraries</td>
<td>13</td>
<td>29.5</td>
</tr>
<tr>
<td>Public libraries</td>
<td>8</td>
<td>18.2</td>
</tr>
<tr>
<td>Academic libraries</td>
<td>23</td>
<td>52.3</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
<td>100</td>
</tr>
</tbody>
</table>

Our survey focused on the following four areas, namely, developing status of digital resources, arrangement alternatives of long-term preservation for digital resources, requirements of long-term preservation models for digital resources, practice of long-term preservation for digital resources. The face to face or telephone interviews were also conducted with the directors of 6 major libraries during this process of fact-gathering.

3 Key findings

3.1 Developing status of digital resources

The emergence of long-term preservation issue for digital resources is closely related to the developing status of digital resources in a given library. The survey investigated
the current status and trends of the digital collection development in those 44 libraries surveyed. The result is shown in Table 2. As for the main format types of library collections, more than 90% of libraries (41 libraries representing 93.2% of the total 44 libraries surveyed), had an equal amount of resources in print as it was in digital format, whereas the remaining three libraries had predominantly a collection in printed format. Also, there were 20.9% libraries that would develop digital collection first, whereas only one library would assign the development of printed materials as their first priority. From the perspective of different types of libraries, all the respondents which chose to develop digital resources first were from special and academic libraries. The Table 2 shows that a great majority of the libraries preferred a balanced collection development strategy for both printed and digital resources.

<table>
<thead>
<tr>
<th></th>
<th>Special library</th>
<th>Public library</th>
<th>Academic library</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amount</td>
<td>Ratio (%)</td>
<td>Amount</td>
<td>Ratio (%)</td>
</tr>
<tr>
<td>Current status of collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Printed mainly</td>
<td>2</td>
<td>15.4</td>
<td>1</td>
<td>12.5</td>
</tr>
<tr>
<td>Printed + Digital</td>
<td>11</td>
<td>84.6</td>
<td>7</td>
<td>87.5</td>
</tr>
<tr>
<td>Trend of collection development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Printed first</td>
<td>1</td>
<td>8.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Printed + Digital</td>
<td>8</td>
<td>66.7</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Digital first</td>
<td>3</td>
<td>25.0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The survey also finds out that, 66.7% of the libraries which would develop digital resources first have taken digital preservation actions (as compared to the 59.1% of all those who have taken actions on the long-term digital resources preservation). Owing to the insufficient samplings, we cannot make any conclusive remarks in statistical terms. However, the data above do suggest that the libraries which had paid more attention to digital resources would be more inclined to take pro-active measures in dealing with the issue of long-term digital resources preservation.

3.2 Arrangement alternatives of long-term preservation for digital resources

3.2.1 The scope of long-term preservation for digital resources

Owing to the fact that libraries may adopt different preservation strategies for different categories of digital resources, our survey divided the digital resources into the following three categories for the clarity of our elaboration, namely, 1) self-built resources, 2) subscribed resources, 3) web-harvested resources. All the responding institutions took the view that the self-built digital resources should be preserved in their entirety. As for the subscribed digital resources, 25% of surveyed institutions would preserve materials in this category as a whole, while 70.5% would preserve
selectively, and the rest two libraries would not preserve any of them. As for the web-harvested resources, 70.5% would preserve selectively. Among the remaining 13 responding libraries, there were 11 of them indicating that they would not preserve and the other two did not indicate their preference either way.

3.2.2 Preferred institutional leadership role in the undertaking of long-term preservation for digital resources.

More than half respondents (52.5%) choose the library to take the responsibility for digital preservation. One respondent pointed out that the library here should refer only to those non-profit libraries at the national level. There was 7.5% of those surveyed institutions regarding the preservation of digital resources should be the responsibility of publishers, while 27.5% of them preferred some third parties rather than libraries or publishers to undertake the preservation task. Besides, there were 12.5% of those surveyed institutions who had no clear idea as to where the responsibility for long-term preservation of digital resources should lie. From the perspectives of all types of responding institutions, 75% of public libraries preferred that the library should assume such responsibility, whereas academic libraries held an evenly divided view on this preservation issue. There was an evenly distributed 42.9% institution that opted equally for either the library or a third party to take the helm for such professional undertaking.

3.2.3 Arrangements for archiving access rights

Among those libraries surveyed, 54.5% of which would insist on obtaining the perpetual archiving access right, and 13.6% of them would like to be granted the perpetual copyright, while 40.9% of them preferred to acquire the perpetual copies of the resources and the metadata necessary for providing professional services. A small amount of libraries (11.4%) was in the process of giving it a thorough consideration.

As for the issue of acquiring the archiving access rights mentioned above, many libraries (75.0%) placed such requirement into their subscription agreement with the publishers. The percentage of the libraries which preferred to sign an independent archiving access right agreement or gain an informal license to acquire the expected archiving access rights was less than 10%; the rest 18.2% do not give a clear indication about their preferred arrangement for this issue of archiving access rights.

3.3 Requirements of long-term preservation models for digital resources

3.3.1 Preference of preservation models

The survey asked the respondents to rank the three types of preservation model, namely, self-managed preservation model, third-party preservation model and co-operative preservation model, in accordance with their preference. The results diagramed in Fig. 1.
An investigative report on current long-term digital preservation among major CL

ZHANG Mei et al.

Research Papers

National Science Library,
Chinese Academy of Sciences

The above reveals that, Chinese libraries preferred the co-operative preservation model. However, public libraries preferred the self-managed preservation model slightly more than the other two models, whereas the co-operative preservation model was the first choice among the other two types of libraries.

Moreover, as for the question regarding which organization should play the third-party role in digital long-term preservation, 84.5% responding libraries selected a national preservation system. In contrast, there were only a few libraries which had chosen the commercial service sector or an organization within the same institution, such as the information center in the same institution to be the third-party digital preservation organization.

3.3.2 Library requirements and concerns for different operational models

In view of different digital preservation models having different requirements and problems related to their service capacity, the authors have factored these elements in the context of institutional service into our study.

Fig. 2 and Fig. 3 below are the requirements and concerns of the libraries for the self-managed preservation model.
The first two need-requirements of self-managed preservation model are the adequate funding for operational cost & personnel support (90.5%), and a clarification of content and scope of the digital preservation project (88.1%). The main concerns about the self-managed preservation model include but not limited to the following, such as the lack of sustained and adequate funding (83.7%), intellectual property rights attribution (67.4%), and difficulties in gaining the necessary authorizations from data vendors (67.4%), and so on.

Fig. 4 and Fig. 5 below are the institutional requirements and concerns of the third-party preservation model.

For the third-party preservation model, almost all the institutions which participated in our survey emphasized the importance of the clarification of mutual rights and obligations in any signed bilateral agreement for long-term digital preservation undertakings (97.6%). A second institutional requirement was the demand for a sustainable digital preservation policy for preventing the loss of data from the digital preservation system when it ceased to operate (85.7%). There was also a grave concern among responding institutions about the possible interruption or stoppage of service when the digital preservation is administered by a third-party institution.
An investigative report on current long-term digital preservation among major CL (90.2%). As for the issue of intellectual property rights, it was not as remarkable as compared to other issues associated with the third-party digital preservation model. Less than half of the surveyed institutions expressed their concerns on this issue (41.5%).

Fig. 6 and 7 show the survey results of the co-operative preservation model for digital resources.
3.4 Practice of long-term preservation for digital resources in China

In order to learn the real situation of long-term preservation for digital resources in China, the authors first calculated the number of participating institutions with the practice of digital preservation. The results revealed that 59.1% participants (44 respondents) have embarked on the task of digital preservation (Fig. 8). According to our survey finding, most of academic and public libraries were involved in varying degrees for the preservation of digital resources. However, most of the special libraries have not yet to develop such practice. The situation in comparative perspective is diagramed below.

![Fig. 8 The situation of digital preservation practice in the various types of libraries in China](image)

3.4.1 Institutions without digital preservation undertakings

Institutions without having a program of digital preservation have different reasons of their own. Some (66.7%) participants took the view that a period of time was necessary for them to understand the issue more thoroughly before making a move on the issue of digital preservation. This was due to insufficient practical experience in this area among domestic information organizations. Some (38.9%) participants considered the budget shortage situation would be compounded by the fact that “electronic resources should have the relative print versions as the backup” so much so as to make such preservation practice untenable. Several surveyed institutions (16.7%) gave another reason for their hesitation in assuming such preservation practice. That was because they were adversely affected by the similar undertakings...
of some other libraries that had already been in such operation for sometime. Besides, one participant pointed out that there was a lacking of a mechanism to safeguard such digital preservation endeavors.

As for the outlook for the task of digital preservation in the near future, 66.7% of the surveyed institutions which do not yet have an on-going digital preservation operation could not decide whether they would start the preservation task in the next two years; 37.5% will take some actions in this regard; 3 institutions did not think they would take any action for a long-term digital preservation project in the foreseeable future.

### 3.4.2 Institutions engaged in practice for digital preservation

Based on the perspective of digital resources categories of those libraries that have an on-going long-term digital resources preservation operation, the self-built digital resources are the ones that all libraries would preserve for a long period of time. There were 2/3 surveyed libraries who would preserve part of the subscribed resources; 29.6% surveyed libraries indicated that they would preserve web-based resources.

From the perspective of publishing types of resources, most participating institutions (74.1%) would preserve digitized journals in science and technology; 44.4% participants would preserve digitized archiving materials; 29.6% participants would preserve web pages and images. Furthermore, other resources types might also be included by these surveyed libraries such as thesis and dissertations, ebooks, videos and images, newspapers in PDF format, databases and scientific report, etc.

From the perspective of preservation models, all the surveyed institutions selected the self-managed preservation model. Basically based on such a premise, 4 surveyed libraries also took part in an additional co-operative preservation system in order to preserve certain resources with an added measure of security.

From the perspective of preservation operation systems, Dspace was adopted by 4 surveyed libraries, and there were also other preservation operation systems that had been adopted such as dILAS/MagicStore, TRS, TPI and hardisk backup systems, etc.

From the perspective of whether institutions had developed a specific policy for the long-term preservation of digital resources, 17.9% of the surveyed institutions made such a policy, and another 50% were in the decision-making process. Of those policies that were currently in effect, 22% contained the items of rights and obligations of digital preservation; 18.5% contained the item of selection standards for digital resources for long-term preservation; 11.1% defined the source of funding for digital preservation operation; 25.9% described the workflow.

From the perspective of those impact factors on the outcome of a digital preservation operation (Fig. 9), it is clear that a well-framed digital preservation policy brings satisfactory result for such undertakings. Moreover, 2 surveyed libraries
indicated that the degree of importance that the library administrators assigned to such an undertaking and the wide collaboration with other institutions would also be the impact factors to leverage the effect of the digital preservation undertakings.

As for the evaluation of preservation effect, the main method by the libraries surveyed used (42.9%) was to develop standards for digital long-term preservation and to have their performance monitored periodically by either the operating unit itself or by a superior level of management body. 17.9% surveyed libraries would have a third party to do the outcome assessment based on their established standards for long-term preservation of digital resources; 53.6% surveyed libraries have not yet done such a preservation evaluation.

3.5 Suggestions for the development of a national long-term preservation for digital resources

At the end of our questionnaire, there was an open invitation for the participants’ suggestions on the development of a national long-term preservation for the digital resources. 50% participants responded to our request and the article summarized their answers as follow.

3.5.1 Improving the awareness of digital preservation.

We should not only improve the preservation awareness of the librarians, but also strengthen the importance and need to the upper management team. To realize such purpose, it is recommended that the following efforts need to be taken, such as studying the national and international status and policies of preservation, proving the needs and feasibility of preservation, providing evidence that libraries are able to complete the issue of preservation, and translating the responsibility of making long-term preservation of digital resources into the Chinese library law now being drafted.
3.5.2 Constructing the national preservation system for digital resources.

It is much easier for a national long-term preservation system of digital resources than for an individual digital preservation system to solve the issues of standards, administration, security, key technical problems, redundant investments and duplicated efforts for the sake of maximize the benefits derived from digital preservation undertakings. It is also necessary for the central government to give strong support to the regional governments and professional organizations in their either joint or separate endeavors for the preservation of relevant digital resources.

3.5.3 Formulating policies for a national long-term preservation system of digital resources.

A policy or guidelines for long-term preservation of digital resources should be made at the national level. It is the foundation upon which by-laws are made for the coverage of the following operational areas such as participants’ membership rights and obligations, standards for technological equipment and its maintenance, technical support, professional licensing requirements, and outcome assessment for digital preservation undertakings etc.

3.5.4 Applying different preservation models for different categories of digital resources.

Original self-built digital resources may be preserved by the institution that produced them in the first place. When the institution is not able to preserve such digital resources, the third party preservation model becomes applicable. Subscribed digital resources are more suitable for the government-designated third parties to preserve. Co-operative preservation system may be delegated to handle the preservation of the web-harvested resources.

4 Conclusion

The survey describes the basic attitudes of the domestic libraries with regard to long-term preservation of digital resources. Such survey is similar to those made by scholars in foreign countries. The authors investigated in some depth about issues related to the co-relationship between the origins of digital resource and the resulting digital preservation arrangement. In addition, preservation models, institutional requirements and concerns about long-term preservation of digital resources are also examined to some extent.

From the survey findings, it is learned that the collection size of the digital resources and the print resources in most of our surveyed libraries are just about the same. Moreover, the majority of the institutions have adopted a balanced collection development policy for both printed and digital resources. Some institutions, however, give priority to the development of the digital resources. It is quite evident that a commitment to the long-term preservation for digital resources has dawned upon us as an urgent and pressing issue for libraries large and small in our society.
today. Most of the surveyed libraries would preserve their self-built digital resources by themselves. A great majority of the surveyed libraries would preserve or selectively preserve their subscribed digital resources in one way or the other. More than half of the surveyed libraries would prefer the national libraries to take the leadership and responsibility for long-term digital preservation. The survey reveals that the Chinese libraries have taken a pro-active attitude towards the issue of digital preservation.

The survey also indicates that most of the institutions have already taken up the task of digital preservation. The majority of the surveyed libraries are willing to adopt the co-operative preservation model, especially for the academic and special libraries. When in consideration for using the third-party preservation model, most of the surveyed libraries preferred to delegate such matter to national libraries. Most of the surveyed libraries expected to get necessary funding and staff training through their active participation in a national digital long-term preservation system. All of the above mentioned library understanding and practice are relevant and essential to the further development of a national long-term preservation system for digital resources in the near future.

The survey divides the participating libraries into three categories; namely, public libraries, academic libraries and special libraries. It reveals that libraries in different categories would adopt different strategies towards digital preservation; even the ones within the same category may act quite differently. Such differences may be explained from their different missions, their personal and financial resources and/or from a host of other factors. Nevertheless, for the limited space of the questionnaire, the authors have not delved deeply into those factors that affect the librarians’ attitude which ought to be explored at a later writing opportunity.

The survey also has revealed a series of library operationally related problems, which need to be addressed quickly in one way or the other by library leadership, by professional organizations, by governments’ ordinance and/or through legislative process at all levels. These include but not limited to the following: inadequate and unstable library funding, outdated administrative structures, limited funding for technological equipment and their maintenance, inadequate technical support, undemocratic policies and procedures regarding recruitment, retention and rank advancement for library professionals, etc. In short, libraries are facing unprecedented challenges in many, if not all of their traditionally operational areas as well as a new exiting vista for their splendid service opportunities. This is the new library reality under the impact of the rapid development of information technologies and digital resources. The issue of long-term preservation of digital resources is but a microcosm of this new information age of ours. People should continuously explore it in a meticulous and collaborative way so as to gain a through understanding about all aspects of this long-term digital preservation issue within the context of the overall professional undertakings. Unless and until we have accomplished this task, our mission of rendering efficient and effective library service to the reading public will remain to be haphazard and under-fulfilled.
Reference