Academic institutional repositories in China: A survey of CALIS member libraries

Yongchao ZHAO1†, Xiaoxia YAO1,2 & Chengfu WEI2

1 CALIS Administrative Center, Peking University, Beijing 100871, China
2 Peking University Library, Beijing 100871, China

Abstract

Purpose: China Academic Library & Information System (CALIS) planned to launch an institutional repository (IR) project to promote IR development and open access at colleges and universities in China. In order to get to know the current state of IRs in academic institutions, with the help of Peking University Library, CALIS Administrative Center conducted this survey.

Design/methodology/approach: We conducted an online survey of CALIS member libraries.

Findings: Firstly, the development of IRs at China’s colleges and universities is still in its infancy. Secondly, the Chinese colleges and universities have reached a consensus on the objective for having an IR. Thirdly, they are having high expectations of IR functions. Fourthly, they prefer to establish a centralized IR system at a minimum cost. Finally, there are both similarities and differences between the Chinese academic institutions and their counterparts in other countries in the state of IR development.

Research limitations: The questionnaire needs to be improved because there is a lack of enough questions for those who do not plan to build an IR. Comparatively lower rate of valid questionnaire return can affect the accuracy of the results. It is hard to go into an in-depth discussion only based on the data collected from this questionnaire survey, and consequently, the findings from the survey can hardly present an accurate and comprehensive picture of the current state of IR development in the academic sector in China.

Practical implications: The survey results provide essential foundation for CALIS IR project, and meanwhile the research can serve as a reference source for the future studies of the development of IRs at China’s colleges and universities.

Originality/value: It is the first national survey focused on the development of IRs in academic institutions in China.

Keywords Institutional repository (IR); Questionnaire survey; China Academic Library & Information System (CALIS); Academic libraries

† Corresponding author: Yongchao Zhao (E-mail: zhaoyc@calis.edu.cn).
1 Introduction

An institutional repository (IR) is a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and accessible to end users both within and outside of the institution, with few if any barriers to access. In other words, the content of an IR is institutionally defined, scholarly, cumulative and perpetual, and open and interoperable[1]. Not only is it an indispensable tool for the management of an organization’s knowledge assets, but it also represents an important mechanism for developing knowledge management capabilities of the organization[2].

So far, the Registry of Open Access Repositories (ROAR, http://roar.eprints.org/) lists about 3,000 IRs worldwide. In terms of the number of IRs established in a country, the top five are the U.S., the U.K., Germany, Japan and Spain. China ranks 8th with 79 IRs. However, 65 of them are built by different institutes of Chinese Academy of Sciences (CAS)[3]. Certainly, ROAR’s list does not cover all the IRs built by the Chinese colleges and universities. Because of the metadata standard, resource openness, and other factors, some IRs are not registered in ROAR. Undoubtedly, CAS institutes have been pioneering the IR initiatives in China in both the number of IRs and the objects stored in IRs[4]. An integrated CAS IR network has been formed through CAS IR Grid portal, which harvests and aggregates metadata records from node IRs of the local institutes[5]. The Chinese colleges and universities may follow best practices of CAS institutes in developing IRs at their institutions.

China Academic Library & Information System (CALIS), approved by the State Council, is one of the public service systems of the “211 Project”© for higher education and the largest library consortium in China. With the purpose of resource sharing, CALIS has been striving to provide various sustaining services of great quality for higher education to stimulate scholarly communication, and now it has become a vital component of China’s information infrastructure. In order to encourage the member libraries to provide IR services for better managing, preserving, displaying and sharing their academic output on the one hand, and raise the awareness of open access among colleges and universities on the other hand, CALIS planned an IR project for Phase III (2009–2011). Before the implementation of the project, we need to investigate the current state of IR development in the Chinese higher education institutions. For example, how many colleges and universities have built an IR? what are the difficulties that they have come across

© The “211 Project” is a project in which the Chinese government concentrates central and local funding to construct about 100 universities and a number of key disciplines to reach the level of world-class university in the 21st century and to meet the challenges of the world’s new technological revolution[6].
in building IRs? What are the essential functions of an IR system? Thus, with the help of Peking University Library, CALIS decided to conduct a national survey of its member libraries.

We learned from other IR projects both inside and outside China that the questionnaire was the most popular way for researchers to collect data, and their questionnaires generally included the number of IRs, objects stored in IRs (type and format), software used, budget, content recruitment, strength and weakness of an IR, policy and standards, etc. We may summarize their research findings as follows:

- Librarians are leading IR development and implementation, serving on planning and advisory committees, pilot-testing software, recruiting content, identifying early adopters, etc.[7–10].
- To manage, preserve and display an institution’s academic output is one of the benefits that an IR can bring to both academic institutions and society[11,12].
- There can be undeniable difficulties in generating content, especially in the beginning. Often, academics are unwilling or lazy to deposit their research papers, and another impediment to run an IR is the intellectual property right issue[7,8,10,12–14].
- Academic institutions in Asia are not so interested in open access, and the development of IRs has been still in its infancy so far[11,15].
- China’s academic institutions are making slow progress in developing IRs. Although some colleges or universities have IRs operating, there is a lack of objects stored in IRs and the content has not been made available to the public. Some universities with rich digital resources even restrict IRs’ open access[4].
- IR consortium represents one of the trends in the future, which can stimulate resource sharing and usage among different institutions[16,17].
- From the user’s perspective, the IR’s biggest advantage is the long-time preservation and the usage statistics of the objects stored in IRs[13]. Presently, most of IR systems provide only some basic functions, but the users are expecting more customized services, such as those based on Web2.0 technologies[7,18].

2 Methodology

2.1 Respondents

CALIS, as the largest academic library consortium in China, has more than 1,300 member libraries. According to incomplete statistics from CALIS, there are 2,400 universities in China[6]. In order to have a larger sample size, we decided to conduct a survey of all member libraries of CALIS by inviting librarians who are leading
the IR projects at their institutions to fill in the questionnaire. When the research was carried out, nearly 1,000 member libraries had activated their accounts at CALIS, and they were from 31 provinces, municipalities and autonomous regions of China Mainland, except the Tibet Autonomous Region. Various types of colleges and universities participated in our survey, including regular higher education institutions (HEIs), key universities such as 211 Project/985 Project evaluated universities\(^2\), adult HEIs\(^3\), independent colleges(state-owned, private sector-run), postsecondary vocational institutions and other types of colleges. Thus, this research sample is representative of the population that we intend to investigate.

2.2 Survey method

The survey was conducted at the beginning of June in 2011, and lasted for a month. We designed an online questionnaire through the open source software LimeSurvey, and then invited librarians leading IR projects at their institutions to fill in the questionnaire online. Member libraries can be informed of the survey via CALIS three-tier administration system—national administrative center/provincial information centers/member libraries. More specifically, CALIS Administrative Center sent an e-mail invitation to the contact persons of 31 provincial centers (usually only one contact person for every provincial center), and asked them to help inform local CALIS member libraries of the survey. As we know, most provincial information centers continued to notify contact persons of local member libraries (usually only one contact person for every member library) by e-mails, and they would also post a notice on the homepage of their websites inviting librarians to participate in the survey. The respondents could just click on the link, fill in the questionnaire and submit it.

2.3 Questionnaire structure

According to our literature investigation and on-the-spot investigation, we found that China’s academic institutions could be classified into four types in terms of IR construction: 1) Already having an IR operating, 2) building an IR, 3) planning to build an IR, and 4) not having an IR at all. Therefore, we included three groups of questions in the questionnaire, which were named as “Yes/Building”, “Planning” and “Not yet”.

\(^2\) 985- and 211-Project evaluated universities are normally famous universities with high reputation in China. At present, there are 112 “211 Project” universities, among them, only 39 are “985 Project” universities\(^19\).

\(^3\) Adult higher education takes forms of full-time adult higher education, TV& Broadcasting universities, web-based higher education programs, and self-taught examinations. Graduates from adult higher education section are eligible to get state-recognized graduate diploma and degree certificate if they reach the requirements\(^20\).
Each questionnaire contained an instruction part and a main-body part. The former consists of a brief introduction of the IR definition, the survey objectives and greetings; the latter, 35 questions, including one question to ask whether respondents have an IR operating, 19 questions for those who choose “Yes/Building”, 7 questions for those who choose “Planning”, and 8 questions for those who choose “Not yet”. In terms of the format of questions, both compulsory questions and optional ones are included. The questions cover various aspects of an IR, for example, the benefits of an IR, software used, content recruitment, content types and formats, reasons for building an IR, the difficulties of operating an IR, etc. At the end of the questionnaire, we required the respondents to write down their names and e-mails, and most respondents did so.

2.4 Questionnaires returned and the rate of valid return

Till the survey was closed, we had received about 800 replies. The following are the steps involved in judging whether a returned questionnaire was valid or not.

At first, we selected questionnaires in which respondents answered all the compulsory questions and we got 384 questionnaires. Then, we found that some questionnaires came from the same institution. Although we had explained the rule clearly in the notice that only one questionnaire would be accepted from the same institution, we still received questionnaires that were filled in by different respondents from the same institution. In this case, the questionnaire in which the respondent answered both the compulsory and optional questions and wrote down his or her name and e-mail address was selected. A total of 349 questionnaires were then selected among the 384 questionnaires.

The last step was to eliminate invalid questionnaires by our definition of an IR. There were three kinds of databases that need to be explained. The first one is the research output database, an in-house database collecting all the research output of a college’s or university’s faculty members, administrators, students or alumni, which was always considered as the “quasi-IR” because it could provide the basic academic resources for an IR[21]. The other two are thesis and dissertation database and specialized subject database, which normally collect and preserve only one kind of digital resource and the full texts are often not accessible[14]. In this paper, we counted a research output database as an IR, but not thesis and dissertation database or specialized subject database. Fifty-four respondents of 349 institutions stated that they had an IR operating. According to the website addresses (URLs) of the IRs provided by each respondent, we checked their IRs one by one. In the end, there were 315 valid questionnaires, with a rate of valid return of 39.4%. In addition, we got 137 pieces of advice and suggestions.
3 Findings

3.1 Status of IRs

Among the 315 respondents, only 20 responded that they had operational IRs while other 25 were building one. It means that the development of IRs at China’s colleges and universities is still in its infancy. The detail is presented in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Regular HEI</th>
<th>Adult HEI</th>
<th>Independent college</th>
<th>Private university</th>
<th>Vocational college</th>
<th>Others</th>
<th>Total</th>
<th>211/985 Project evaluated university</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>17</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Planning</td>
<td>79</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>91</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Not yet</td>
<td>136</td>
<td>2</td>
<td>16</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>179</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>257</td>
<td>2</td>
<td>20</td>
<td>12</td>
<td>20</td>
<td>4</td>
<td>315</td>
<td>66</td>
</tr>
</tbody>
</table>

3.2 Types of objects

Fig. 1 shows the types of objects stored in IRs. It can be observed that various categories of objects are deposited in IRs, and the main focus of the holdings is on theses and dissertations (85.71%) followed by journal articles (84.13%) and courseware (78.41%).

Because of previous projects, such as the project carried out by CALIS to build thesis and dissertation databases, colleges and universities in China have accumulated a great number of academic resources. For academic institutions, it is comparatively easier to populate their IRs with journal articles. Surprisingly, pre-prints got the
lowest support. Here are two possible reasons. The first one, although the copyright of pre-prints belong to the authors and they can self-archive a copy of the pre-print version on IRs, some publishers may insist that self-archived pre-prints are removed once the paper has been accepted for publication. To avoid the risk of intellectual property rights, authors often remove their pre-prints from IRs[22]. The other one, it is to assure the quality of content being housed in the repository. When the pre-prints are submitted into an IR, their quality is not checked, and therefore it is hard to guarantee their quality[23].

3.3 System functions

As shown in Fig. 2, most of the 12 functions have got a support rate above 60%. Search by subject/author (94.92%), content submission & editing (91.75%) and full-text downloads (89.84%) are the top three functions. Moreover, respondents hoped that an IR can provide more functions, such as discussion, an application programming interfaces (APIs) that can interact with other Web services, data analysis, etc. Overall, the respondents are having high expectations of the IR system functions.

![Fig. 2 Functions of an IR system (N = 315). Note: a, DOI registration; b, full-text search; c, search prompt automatically displayed; d, supporting multiple languages; e, preview of full text; f, RSS; g, the latest uploaded content; h, download ranking; i, visit counts; j, full-text downloads; k, search by subject/author; l, content submission and editing.](image)

3.4 Ways to build an IR

Fig. 3 illustrates how the academic libraries are building an IR. Surprisingly, the majority of libraries preferred to contract with a third-party provider (57.46%), about one third (34.92%) respondents chose to use open-source software and only 7.62% of the respondents would establish their IRs independently. What is more, their suggestions demonstrated that the open-source IR system cannot satisfy users’ requirements. In addition, building an IR by using open source software or by
developing their own software independently will pose technical challenges to most colleges and universities due to availability of technical support. As a result, more colleges and universities prefer to contract with a third-party provider.

In addition, among the 45 universities either having an IR operating or building an IR, 10 selected DSpace open source software, 14 developed their own IR system independently, and the rest 21 institutions contracted with a third-party provider.

![Fig. 3 Ways to build an IR (N = 315).](image)

### 3.5 Reasons for building an IR

As shown in Fig. 4, most respondents agreed that IR is a feasible way for management, display and long-time preservation of an institution’s academic output.

![Fig. 4 Reasons for building an IR (N = 315). Note: a, Without special reasons; b, influenced by the upsurge of interest in building an IR; c, long-time preservation of an institution’s academic output; d, displaying an institution’s academic output; e, preservation of an institution’s academic output.](image)

### 3.6 File formats of objects

Table 2 illustrates that various formats of content have already been included. Obviously, PDF (77.78%), DOC (60%) and JPEG/GIF, etc. (48.89%) are the most popular ones.
Table 2  File formats

<table>
<thead>
<tr>
<th>Formats</th>
<th>Yes/Building (n=45)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF</td>
<td>35</td>
<td>77.78</td>
</tr>
<tr>
<td>DOC</td>
<td>27</td>
<td>60</td>
</tr>
<tr>
<td>PPT</td>
<td>20</td>
<td>44.44</td>
</tr>
<tr>
<td>TXT</td>
<td>15</td>
<td>33.33</td>
</tr>
<tr>
<td>HTML</td>
<td>15</td>
<td>33.33</td>
</tr>
<tr>
<td>JPEG, GIF, etc.</td>
<td>22</td>
<td>48.89</td>
</tr>
<tr>
<td>MP3</td>
<td>7</td>
<td>15.56</td>
</tr>
<tr>
<td>WMA</td>
<td>8</td>
<td>17.78</td>
</tr>
<tr>
<td>RMVB</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>AVI</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>MPEG</td>
<td>5</td>
<td>11.11</td>
</tr>
<tr>
<td>DVD</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SWF</td>
<td>3</td>
<td>6.67</td>
</tr>
</tbody>
</table>

3.7  Content recruitment

Fig. 5 reflects different ways to recruit content. Obviously, most authors submit content via the Web form. Only a few institutions harvest repository data using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

Fig. 5  Ways to recruit content (N = 45). Note: a, Data batch imports via back-end commands; b, content submission via the Web form; c, data batch imports via special tools; d, automatic content submission via tools; e, metadata harvesting using OAI-PMH.

3.8  Ways for the deployment of IR systems

As shown in Fig. 6, the majority of 45 respondents who have already built an IR or are implementing an IR chose to deploy a centralized IR system. They preferred to
enable different users to have different digital right management controls for content submission and management. Overall, the construction of IRs in China’s colleges and universities is still in its initial stage. Just as Chen\textsuperscript{24} said, as the implementation of an IR involves various aspects, such as policy, staff, technology, service, funds and promotion and publicity of the benefits of the repository, it is an issue too complex to be handled by an academic library. After all, a centralized IR system can be much helpful to the construction and sustainable development of an IR.

Fig. 6 Ways to deploy IR systems ($N = 45$). Note: a, Deploying only one centralized IR system; b, deploying distributed local IR systems and the centralized IR system harvesting metadata; c, other ways.

3.9 Access control

With regard to access control of IRs, 10 out of 45 who have an IR operating or building an IR chose open access, which merely accounted for 22% of the total (Fig. 7). Although IRs are means to promote open access, due to the risk of intellectual property rights, most colleges and universities in China still preferred a limited access, such as allowing access to the IR system for only computers having an on-campus IP address. Meanwhile, as to access to the content stored in the repository, 30% of respondents stated that they would allow open access, and users can browse and download the full-text files from the repository freely. The rest respondents preferred limited access and they would allow the person who submits content to IRs to determine users’ access rights.

Overall, we still have a lot of work to do to promote open access at the Chinese colleges and universities. To promote open access, the first batch of 26 academic libraries who will participate in the IR project of CALIS Phase III are required to enable the public to access their content stored in their IRs. They will also be required to provide more than 2,000 records in at least 3 different formats in each IR, and meanwhile they will allow IR users to browse and download full-text files. Indeed, considering the protection of intellectual property rights and the willingness and needs of the participating libraries, we have also configured our system to enable user-level digital right management (DRM) controls for content access and management.
3.10 Difficulties in running an IR

As shown in Fig. 8, the major challenges for these colleges and universities to run an IR are the lack of dedicated staff responsible for repository work, content recruitment, and promotion and publicity of the benefits of an IR. Like in other countries, the Chinese academic libraries are active IR facilitators. However, it is hard for an academic library to implement IR services without the cooperation and support of the other academic departments, especially in recruiting content. Undoubtedly, content recruitment for an IR is a major challenge worldwide.

Fig. 7 Access control of IR system and IR content (45 respondents).

4 Discussion

The survey shows several characteristics of the IR development at the Chinese colleges and universities. Firstly, the IR construction of China’s colleges and universities is still in its infancy. Only a pretty small percentage of the Chinese academic institutions have built an IR. 20 out of 315 respondents stated that they had built an IR, accounting for 6% of the total. Many academic libraries were not
Academic institutional repositories in China: A survey of CALIS member libraries

Yongchao ZHAO et al.

Research Paper

clear of the definition of an IR and they regarded thesis and dissertation database and specialized subject database as IRs. 30 out of 54 respondents claimed having an IR operating. However, their IRs were actually thesis and dissertation databases or specialized subject databases. Moreover, there is still a lot of work to do to promote open access at China’s colleges and universities.

Secondly, these colleges and universities have reached a consensus on the objective for having an IR. As the survey result shows, most academic institutions chose to build an IR because it is an ideal way to preserve and display an institution’s academic output, in line with the purpose of establishing an IR. To our surprise, pre-prints got the lowest support as a type of content for an IR, and merely about 44% of respondents agreed that they should be stored in an IR.

Thirdly, China’s colleges and universities are expecting an excellent IR system. The ideal IR system is expected to provide 15 powerful functions, capable of catering for different types of content and formats, or, more specifically, about 30 content types and 20 formats. It is obvious that the current open source IR software cannot meet their requirements. Therefore, this definitely demonstrated why nearly 60% of respondents were inclined to build an IR through a third-party provider.

Fourthly, China’s colleges and universities prefer to establish a centralized IR system at a minimum cost. The survey has found that a significant impediment to run an IR is the lack of dedicated staff responsible for repository work and most universities tend to deploy one IR system, which enables different departments to have DRM controls for content access and management. The plan of building a CALIS IR system with major functions was well received by 90% of the respondents, for it could save much manpower and reduce the development cost for them. Likewise, 80% of respondents were in favor of the idea to establish a CALIS IR integrated portal.

Finally, we made a comparative analysis of the overall growth of IRs between the Chinese academic institutions and their counterparts in other countries. Similarly, we found that academic libraries are acting as the facilitators of the development of IRs, and the major challenge that the institutions faced in running an IR is content recruitment worldwide. Our survey also proved that it is a promising way to preserve an institution’s academic output via IRs. By comparison, the construction of IRs in China’s colleges and universities started relatively late, just in its beginning stage. At present, the most urgent issue is how to establish an IR system, and a centralized IR is preferred. The Chinese colleges and universities did not pay enough attention to such factors as staff and expenditure involved in the construction of IRs. Overall, China’s colleges and universities are enthusiastic about the development of IRs and confident in a promising future for IRs.
This study has several limitations. At first, we did not describe the definition of an IR and its scope in the instruction part of the questionnaire very clearly. As a result, some returned questionnaires turned out to be invalid, and some extra work needs to be done to confirm whether or not an institution’s IR is an IR according to our definition. Secondly, the questions are mainly designed for universities who have an IR running or planning to build an IR, but we did not devise enough questions for those who have not had an IR yet. As a result, we do not know why they have not built an IR, and what are the chances that they may build an IR in the future. Thirdly, we received several responses from the same institution, which was mainly due to the distribution of questionnaires. Because this questionnaire survey was carried out nationwide, and the survey notice was sent by e-mail, or posted on websites, or member libraries may be informed of the survey by telephone, the whole distribution process was quite complex and not easy to control. Finally, the final questionnaires returned which can be regarded as valid were not as many as we expected, though we conducted the survey based on about 1,000 CALIS member libraries nationwide. There are three possible reasons: 1) Not all the member libraries had got the survey notice; 2) the deadline had passed when some respondents got to know the survey; 3) some libraries may not be interested in this survey.

Since the development of IRs at the Chinese colleges and universities is still in its elementary stage, academic librarians, who are leading the IR initiatives and developments at their institutions, may have a vague understanding of how to promote IR construction. It is hard to go into an in-depth discussion only based on the data collected from this questionnaire survey, and consequently, the findings from the questionnaire survey can hardly present an accurate and comprehensive picture of the state of IR development in academic institutions in China. CAS institutes have been spearheading the IR initiatives in China and they have made much progress in related policies, mechanisms and services. The Chinese colleges and universities can learn from best practices of CAS institutes in their endeavors to develop IRs at their institutions.

5 Conclusion

Our survey reveals that the development of IRs in the Chinese institutions of higher learning is still in its infancy. The Chinese colleges and universities have reached a consensus on the objective for having an IR. They are having high expectations of IR functions and they prefer to establish a centralized IR system at a minimum cost. Moreover, we also found that there are both similarities and differences between the Chinese academic institutions and their counterparts in other countries in the state of IR development.
To sum up, we have achieved our anticipated goal via this national survey, and we finally determined the detailed outline of CALIS IR project based on the survey results. So far, the project has been basically accomplished. The CALIS integrated IR portal (http://ir.calis.edu.cn/) and two local IR systems, one built on open source software and the other one developed independently, have been officially launched since this May, and they can be used by all of the Chinese colleges and universities free of charge.

References