Developing institutional repositories network: Taking IR Grid at Chinese Academy of Sciences as an example①

Zhongming ZHU[1]*, Dongrong ZHANG[2], Lin LI[2], Jianxia MA[1] & Xiwen LIU[2]

[1] National Science Library, Lanzhou Branch, Chinese Academy of Sciences, Lanzhou 730000, China
[2] National Science Library, Beijing Headquarters, Chinese Academy of Sciences, Beijing 100190, China

Received Oct. 17, 2011
Revised Dec. 19, 2011
Accepted Dec. 27, 2011

Abstract This paper introduces the current practice of building a network of institutional repositories (IRs) at Chinese Academy of Sciences (CAS), which is named CAS IR Grid. National Science Library (NSL) of CAS plays a leading role in the construction, promotion and implementation of CAS IR Grid. It aims to promote each institute of CAS to build IR of its own, and finally form the IR network of CAS institutes. NSL's experience is introduced in coordinating and supporting institutes' building of their respective IRs and promoting IR services by adopting collaborative and progressive development strategies. Achievements made during the development of CAS IR Grid are described and challenges for its future development are discussed. The authors aim to provide best practices for developing a network of institutional repositories in research institute settings, which can serve as a practical reference to other institutions engaged in the similar task.

Keywords Institutional repository (IR), Institutional repositories network, Knowledge asset management, Chinese Academy of Sciences (CAS)

1 Introduction

Institutional repository (IR) grew out of the open access movement and has been promoted as the fastest way to make open access a reality[1]. An IR can provide a variety of value-added services to benefit the scientist, the research institute, the research funder, and anyone who interests scholarly outputs[2]. Currently, an increasing number of universities and research institutes have built or are building their own IRs in order to capture, preserve and provide access to the intellectual output of their academic members. There is also growing awareness among these

CJLIS
Vol. 4 No. 3/4, 2011
pp 24–34
National Science Library, Chinese Academy of Sciences

① This work is jointly supported by the Knowledge Innovation Program of Chinese Academy of Sciences (CAS) and the West Light Foundation of CAS.
* Corresponding author: Zhu Zhongming (E-mail: zzm@lzb.ac.cn).

universities and research institutes of the important role of IR in institutional research distribution strategy[3–4]. The Registry of Open Access Repositories (ROAR) (http://roar.eprints.org) lists over 1,800 IRs worldwide, which are unevenly distributed, and mainly concentrated in some developed countries, including the USA, the UK, Germany, Japan and Australia. Although China Mainland was ranked the fifth in the number of IRs in the world, most of them are built by the institutes of Chinese Academy of Sciences (CAS). There are a few universities in China such as Xiamen University and Zhejiang University that have IRs operating. Some universities such as Tsinghua University, Peking University and Beijing Institute of Technology have started testing repository solutions and China Academic Library and Information System (CALIS) has planning work underway for accelerating the development of IR services in China's universities and colleges[5].

CAS is the largest research organization and a major producer of science, technology and medical (STM) information in China. For CAS, encouraging its affiliated institutes to build IR services is one of its main initiatives to promote open access to its research results. Under the aegis of CAS, National Science Library (NSL) of CAS launched a CAS IR Grid project at the beginning of 2008. Following a progressive multi-phased strategy, NSL assisted the institutes of CAS in the establishment of their local IRs via providing best practices guide, open source based IR software package, and consultation services and solutions and then helped the institutes populate the repositories with their intellectual assets. Keeping pace with the advancement of the promotion and deployment of IRs at CAS institutes, NSL set up a CAS IR Grid portal to harvest and aggregate metadata records from node IRs of the local institutes, thus an integrated CAS IR network was formed.

Although institutional repositories have sprung up at academic institutions across the world, and there is a steady growth in the number of IRs, the successful models and strategies for building IRs and achieving their sustainable development are far from mature. Therefore, we will describe our strategies and experiences in developing IR services in CAS and it is our hope that our practices can serve as a practical reference to other institutions engaged in the similar task.

2 Development strategies of CAS IR Grid

2.1 Background

CAS has 45,400 researchers and 48,000 graduate students, who produce over 55,000 research papers and electronic theses or dissertations (ETDs) per year[6]. However, such a large volume of research output, especially those high-level academic papers published abroad, are not made freely available for sharing and use by Chinese

researchers because of copyright and intellectual property protection issues, and they are separately recorded and hard to preserve and re-use. In addition, other types of works generated by researchers, such as conference papers, patents, research reports, etc., are even more difficult to share and re-use and are subject to the risk of loss. Therefore, spreading IR service within the confines of CAS will produce a network of scientific papers increasing at least at a rate of 50,000 papers annually. It will contribute to the realization of systematic management and preservation of knowledge assets within CAS, and bring about an influential and useful network of knowledge distribution and utilization within the scope of China as well. It will also provide a boost to the open access to scientific and technological knowledge nationally and the construction of open access repository at the national level.

2.2 Construction model

To some people, it seems natural to create a centralized institutional repository in CAS and require every researcher to deposit the products of his/her research in the repository. Nevertheless, one of the major concerns that arises from such a centralized IR is that CAS consists of more than 100 institutes geographically distributed in 28 cities across China and each carries out its research activities in a relatively independent mode and has its own library or information center. Establishing a centralized repository may inhibit each individual institute from giving the full play to its role as the producer of research output and from establishing its knowledge asset management system with its own characteristics. Therefore, a "hub-and-spoke" network model was proposed for the construction of CAS IR Grid. It envisages that each institute establishes its own local repository on the one hand, which is treated as a node of the Grid, and on the other hand, NSL constructs a centralized metadata repository via harvesting and aggregating metadata of academic resources stored in each institute' local IR and build an integrated interface for the aggregated resources to provide value-added services. The institute IRs mainly focus on providing institutes with tools and facilities for knowledge asset management activities, such as the capture, organization, preservation, dissemination, access and evaluation of the intellectual output. The centralized repository is to offer integrated services across repositories, such as data aggregation, search, virtual reorganization, content analysis, backup, etc. A collaborative approach, in particular through the provision of policy, mechanism and technical aids has been employed in the construction of CAS IR Grid. NSL and the library or information center of each institute both play significant roles during the process of the construction. The former plays the leading role in taking the initiative and shouldering the responsibility to advance the IR network construction and the latter plays the supporting role in building and

managing their own IRs. Compared with the consortium-based repository, which is popular in the academia of some developed countries and usually based on an existent research consortium established for a common cause and interest by a group of comparable universities and research institutes, CAS IR Grid is a repository network connecting headquarters and branches rather than "repository consortium" no matter in its nature or in its architecture.

2.3 Advancing strategies

The development of CAS IR Grid is apparently a tough task due to three considerations. Firstly, it is supposed to be a large-scale repository network in China. Secondly, some managers and researchers of CAS institutes are not clear about what IR exactly means and the importance of open access repositories and therefore they may not provide support or enough support for the construction of CAS IR Grid. Thirdly, the building process of IR itself is required to take into consideration various issues such as policy, technology, management, culture, etc. Being aware of the complexity in building CAS IR Grid, NSL adopted a multi-phased approach, which involves the following three phases, namely, 1) Pilot demonstration phase, 2) promotion phase, and 3) full coverage phase. The strategies involved in these phases are expounded as follows. The pilot demonstration phase concentrated on setting up demonstration IRs developed by several institutes and formulating policies and completing the development of IR software platform with required basic functions. The policy framework then is further optimized to be sample policy templates for references by all other institutes for building their IRs. In the promotion phase, built on work already successfully completed in the pilot demonstration phase, the focus will be placed on improving institutes' engagement with repository matters, demonstrated through increased participation of institutes in building IRs across CAS. The goal of this phase is to complete deployment of IRs in half institutes of CAS. At the same time, the CAS IR Grid portal is launched to connect and cover all IRs deployed at this phase. In the full coverage phase, the major task is to encourage the rest of institutes to build IRs, and bring them into CAS IR Grid to complete the overall goal of the Grid—covering all institutes of CAS.

The progressive strategy has also been used in the process of developing a list of types of content currently in repositories or planned for inclusion in repositories in the future, and also in the development of the IR platform as well. With regard to content development, those traditional and easily captured types of content, such as journal articles, conference papers, theses and dissertations, etc., are given priority over non-traditional types of content, including research reports, scientific data, learning objects, etc. The IR platform was developed following a way of from

supporting common functionalities of knowledge asset management (e.g. content collection, organization, preservation, dissemination, statistics collection and analysis, etc.) to advanced features of knowledge management or knowledge capacity management (e.g. association discovery, knowledge audit and knowledge capacity analysis based on semantic relations between content objects).

3 Promotion and implementation of CAS IR Grid

3.1 Establishment of sustainable support service mechanisms

As the organizer and facilitator of CAS IR Grid project, NSL has established a series of servicing mechanisms to ensure the promotion and implementation of IR services in institutes. They fall into the following five areas:

- The working procedures and guidelines for the promotion of IR services in CAS were formulated to ensure the promotional process run smoothly and efficiently. Some documents include *The Guidelines for Promoting IR Service in Institutes*, *The Agreement for Building IR Service*, *Concrete Measures and Rules for Promoting IR Service in Institutes*, and other project approval procedures and regulations. Following these regulations, an institute is eligible to submit an agreement-based application or a trial application to NSL for approval of buliding an IR. As for an agreement-based application, if approved, NSL will provide some support in terms of fund, expertise, and technology, and on the other hand, NSL will assess and review the IR construction results. With regard to a trial application, NSL just provides limited policy and technical support without any funding support, and there is no requirement for any assessment.
- The collaborative working mechanism was established and several teams were formed to work coordinately to attain the joint objectives. These teams involve subject librarian team, the policy team and the technical support team and they play different roles. The subject librarian team is responsible for concrete promotional tasks; each subject librarian serves as an institute's main liaison and also a facilitator for the deployment and implementation of IR in that institute, and is required to offer consultation and support services, involved in tasks such as planning, building and marketing a repository. The policy team coordinates their activities with the subject librarian team and takes responsibility for providing policy consultation and training services, in particular with regard to intellectual property issues and author rights. The technical support team collaborates with subject librarian team to provide technical support services,

such as development, deployment, training of IR system implementation, technical advices or responding to any other specific inquires or technical problems.

- By investigating IR policies and related intellectual property issues, NSL not only provided the institutes with policy templates to serve as best practice guidelines, but also offered consulting services concerning policies and intellectual property issues. These templates include a primary template and some supplementary templates. The primary template is about management measures and rules for building an IR. It provides reference in defining the responsibilities of a depositor and content types that will be included in the IR, access and distribution policies, evaluation and incentives measures, etc. Supplementary templates are some supplementary agreements on depositing and publishing specific content concerning journal papers written by members of the institute, articles published in institute-sponsored journals, papers presented in institute-sponsored conferences and workshops. By investigating journal publishers' policies towards open archiving, especially policies made by publishers of electronic journals subscribed by and used within CAS, NSL published a policy guide for open archiving to be used by the institute.
- A comprehensive and reliable consultation and technical support service system was set up to facilitate the promotional process. During the process, subject librarian team, technical support team and participants from the individual institutes collaboratively carried out IR construction needs analysis. Technical support team and subject librarian team developed a joint mechanism to provide the institutes with IR system and training services. In China, QQ is the most popular instant messaging service. A QQ group account was created to facilitate instant communication among institutes' IR builders, subject librarians, and technical support team in sharing experience or obtaining support services. A technical support service website (http://service.llas.ac.cn) was set up, dedicated to publishing and sharing experience and best practices in aspects such as IR building, releasing of updated IR software, etc. Support services were also provided via other flexible ways such as video conference, topic salon, on-spot technical support, to enhance communication with the institutes to solve all problems that they may encounter in building an IR and providing services.
- The IR software platform was continually improved and upgraded to meet incessantly expanding demands. A rapid prototyping approach, in particular a patchwork prototyping approach, was adopted to develop the IR platform[7]. Based on open source repository software package DSpace (http://www.dspace.

org), in-depth customization and extension were achieved to make the software platform able to meet needs and requirements of IR building of the individual institutes[8]. Two versions—CAS OpenIR Standard Edition and CAS OpenIR Improved Edition—were released successively. The Standard Edition focuses on providing basic functionalities, such as browse, search, customization, recommendation, usage statistics, knowledge inventories, etc., supporting collection and management of all conventional types of research output. It was used mainly at the early stage of the promotion phase. Based on previous edition, the Improved Edition offers extended functionalities in many aspects, including a featured adaptive and flexible framework to support multiple types of content, a multi-source content collection and integration framework, faceted browse and search, multiple dimensional knowledge asset statistics and analysis, researcher pages, etc. At present, with a few exceptions, most of the institutes' IR software has been upgraded to this new edition.

3.2 Organization and implementation of IR service in the institutes

According to above-mentioned promotional approaches, a general procedure for building an IR involves the following five steps: 1) An institute files an application to NSL for building an IR and proposes implementation plan; 2) the Department of Subject Information Service of NSL performs a preliminary review of the application in terms of major tasks, implementation conditions, schedule arrangement, management measures and assessment indicators; 3) the director of NSL who is responsible for IR management conducts a final review of the application which has passed a preliminary approval and then decides whether it should be approved; 4) the institute will receive a notice from the Department of Subject Information Service of NSL once its application has been approved and the related subject librarian of NSL will also be informed of the result. Then the institute is required to sign a formal project agreement and send an installation notice to the technical support team; and 5) according to the notice, the technical support team will assist the institute with IR operation.

The deployment of an IR system is just the beginning. From the perspective of the IR operation practices of most institutes, the reliable and sustainable development of IRs relies on the following safeguarding mechanisms:

- Putting the development of IR on the institute's agenda as a part of its informatization drive and endeavoring to obtain full and committed support from senior leadership. In fact, during the promotional process of CAS IR Grid, the libraries or information centers, as facilitators of their institutes' IR initiative, are working hard to get support from the leaders of their institutes,

and persuading them to sign IR building agreements with NSL. In a word, leadership and managerial supports are essential for these institutes to launch their IR services.

- Establishing a powerful team to push forward the IR development. In general, most IRs are operated by the institutes' libraries or information centers, whose directors are project leaders or principal facilitators, and besides, related metadata librarians and technicians are important participants. The departments of research management of the institutes are responsible for coordination of the project, and their department directors or designated staff who are in charge of research results management are major participants of the project (they may be assigned as project leaders in some institutes). Such an organizational structure is very conducive for overall service organization, formulation of IR policies, collection of achievement data, coordination among different departments, etc.
- Setting up reliable systems and formulating policies. They are not only major tasks of building an IR, but also assessment indicators according to the IR launch agreement signed between NSL and each institute. Regarding the IR policies on content submission, at present, the policies proposed by institutes are mainly encouraging authors to deposit their journal articles, monographs, theses and dissertations, patents, etc. However, as far as theses and dissertations are concerned, most institutes have formulated a mandatory policy to require graduates to submit their papers to their institute's library or IR. Some institutes are making a tentative effort to attract content for their repositories by allowing researchers to use the journal articles they deposit within the repository for their annual performance assessment, their title promotion and achievement declaration. This represents an important incentive measure that the institutes have adopted to populate their IRs.

Most institutes have adopted a content recruitment strategy of "library depositing as primary means, author self-archiving as supplementary means". There are two reasons for this. One is that it is difficult to carry out a mandatory policy at the early stage of IR building. Two is that self-archiving itself cannot guarantee the collection of all works of each and every institute of CAS due to such factors as staff retiring, graduate students leaving the institute, etc. Therefore, it is now popular for most institutes to collect content from other sources, for example, using a set of data import utilities specially designed and supplied by IR platform to capture related data from research management system, ETDs databases, or other relevant applications within the institutes' information environment, or to ingest the institutes' publication data, from Web of Science via its open web service interfaces, etc.

3.3 Current achievements

With the collaborative effort by NSL and the institutes, over 70 institutes of CAS have set up their IRs by the end of 2011. Among them 58 have succeeded in gathering a large amount of content in repositories and have begun to provide services to the public.

Total volume of items collected by all IRs has reached 230,915. While about 10 of them have collected more than 5,000 items, most of them have had more than a few thousand items. Of all items, 197,214 contain full-text and account for 85.4% of total items; 83,059 are papers written in foreign languages, which is 40% of the total.

Journal articles (71%), theses and dissertations (13%), conference papers (6%), and patents (5%) are primary content types in all IRs. Some IRs have also archived pre-prints, awards, books, presentations, research reports, multimedia objects, and other types of content to varying degrees.

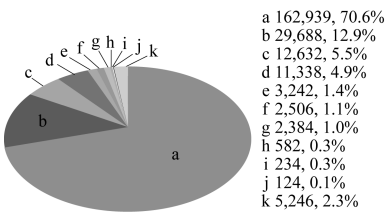

Fig. 1 Distribution of content types. Note: a, Journal articles; b, theses and dissertations; c, conference papers; d, patents; e, pre-prints; f, awards; g, books; h, presentations; i, research reports; j, multimedia materials; k, others.

All content covers a wide span of time. The earliest dates back to 1919, and the most current one is expected to be published in 2012, but most are knowledge outputs of recent 20 years (Fig. 2).

The access usage of CAS IR Grid shows a dramatic increase (Fig. 3). At present, it has 7,093,968 total views and 983,891 downloads.

As the institutes' IRs grew and reached a certain size, CAS IR Grid integrated portal (http://ww.irgrid.ac.cn) was made available online to the public in Apr., 2011. It harvests and aggregates metadata on a daily basis from 58 IRs with certain volume of content items. The number of harvested metadata records has totaled over 230,000. In addition, it has recorded total visits of 676,854 in a short period of time.

National Science Library, Chinese Academy of Sciences

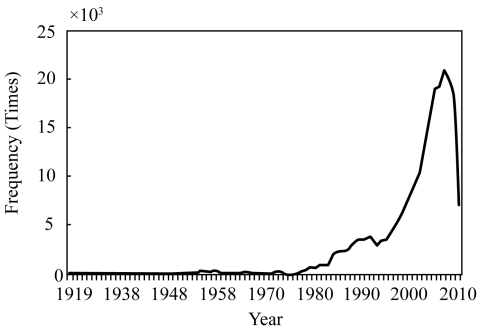

Fig. 2 Time distribution of content.

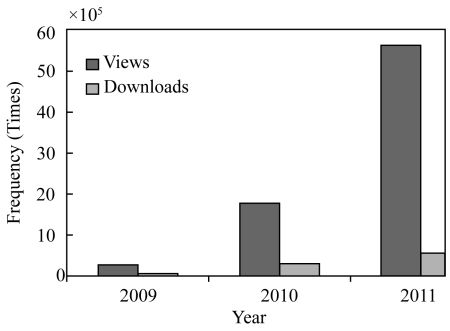

Fig. 3 Access and use of data in CAS IR Grid.

4 Conclusion and future work

Our efforts to implement CAS IR Grid services have resulted in significant progress over the past years. By setting up experimental and demonstrative IR applications to find out whether policy and technical issues need to be addressed, and by establishing a collaborative working mechanism between NSL and the institutes, the expected goal of promoting IR services across CAS has been realized. This result proves that the construction model and strategies of CAS IR Grid are feasible and effective. However, continuing in our efforts towards the full implementation of the CAS IR Grid services in all institutes of CAS, we will have to face some problems and challenges in the future. They include: 1) Setting up a long-term, effective system for the development of the institutes' IR services; 2) establishing feasible and reliable mechanisms, policies, and procedures involved in integrating non-traditional content types, in particular scientific data, into IRs; and 3) creating innovative value-added services and solutions to embed IR seamlessly into research processes and research management processes. Therefore, within the next few years, NSL will continue to collaborate with the institutes, and mobilize more institutes to join the CAS IR Grid. IR policy guides and templates, and corresponding implementation approaches and tools are supposed to be improved to make them serve their purposes better. NSL will also help institutes enhance their capabilities to collect and preserve various types of materials being housed in their IRs, with an emphasis shifting gradually from traditional types of research output such as journal articles and conference papers to non-traditional research output such as research reports and scientific data. Furthermore, NSL will conduct a pilot run of knowledge capacity management services via testing, demonstrating, and promoting knowledge asset audit service as well as corresponding application standards to facilitate the institutes to develop their knowledge management services.

References

1 Crow, R. The case for institutional repositories: A SPARC position paper. Retrieved on Nov. 3, 2011, from http://www.arl.org/sparc/bm%7Edoc/ir_final_release_102.pdf.
2 Hork, R. V., Mossink, W., & Proudman, V., et al. A DRIVER'S guide to European repositories—Inventory study of important DR related issues and good practices. Retrieved on Nov. 3, 2011, from http://www.driver-repository.eu/PublicDocs/D7.2_1.1.pdf.
3 Moody, M. K. Beyond the institutional repository: Campus research distribution strategies. Retrieved on Nov. 3, 2011, from http://scholarworks.boisestate.edu/lib_facpubs/5.
4 Hahn, K., Lowr, Y. C., & Lynch, C., et al. The university's role in the dissemination of research and scholarship—A call for action. EDUCAUSE Review, 2009, 44(2): 6–7.
5 Zhao, Y. C. An overview of the development of institutional repositories in Chinese colleges and universities. Retrieved on Nov. 3, 2011, from http://www.sal.edu.cn/2011/richeng.asp?luntan_id=1.
6 Chinese Academy of Sciences. Chinese Academy of Sciences Annual Report 2011. Retrieved on Nov. 3, 2011, from http://www.cas.cn/jzzky/nb/nianbao2011/.
7 Jones, M. C., Floyd, I. R., & Twidale, M. B. Patchwork prototyping with open source software. In Amant, K., & Still, B. (Eds.), Handbook of Research on Open Source Software: Technological, Economic, and Social Perspectives. Hershey: Idea Group, 2007: 126–40.
8 Zhu, Z. M., Ma, J. X., & Lu, L. N., et al. In-depth customization and extension of DSpace: Case model of developing an institutional repository system for institutes in CAS. In Su, Y., & Hakim, L. (Eds.) Proceedings of the 5th International Conference on Cooperation and Promotion of Information Resources in Science and Technology. Irvine: Scientific Research Publishing, 2010: 398–403.

(Copy editor: Ms. Lin PENG)