

Table of Contents

Cluster 1 – Cross-Domain Information Services

Session 1 – Document Delivery Services

China Direct: Successful Practice of Cross-Domain Information Service between China and Germany PING JIA, SUQING LIU, SUNZHONG CHEN	3
Sino-German “Digital Information Provision”, subito China Direct: Cross-Domain Information Services TRAUTE BRAUN-GORGON	10

Session 2 – Internet Resources

A Simple Search Engine Based on Agricultural Ontology JIEZHENG PI	16
Fields of Co-Operation between AII/CAAS and ZB MED: Present and Future TANJA MEYER, ULRICH KORWITZ, NENSKE GRONE	21

Session 3 – Portals

vascoda – Co-Ordination of International Activities of German Subject Portals ANNE BARCKOW	29
The Structure and Construction of Intellectual Property Network of CAS WEN YI, FANG SHU, ZHANG DENGSHUO	34
Design and Implementation of the Sino-German Humanities and Social Sciences Portal ZHOU CHEN, WANG ZHIGENG, QU YUNPENG	43
From a Virtual Library to a Virtual Research Environment: CrossAsia – CrossAsia Campus MATTHIAS KAUN	50

Cluster 2 / Session 4 – Non-textual and Digital Curation

Competence Centre for Non-Textual Material (CNM) MARGRET PLANK	57
Research and Practice on SIP Ingestion Based on Trusted Workflow Management WU ZHENXIN, LIU JIANHUA, GAO JIANXIU	63
Assessment of Trustworthiness of Digital Archives GREGOR HORSTKEMPER, TOBIAS BEINERT, SABINE SCHRIMPF	71

Cluster 3 / Session 5 – Information Literacy and OA

Information Literacy in Tsinghua University JIA LIN, XIAOMU ZENG	81
Developing Information Literacy Skills by Informal Learning MARGRET PLANK	90
Open Access and the Collaboration with Publishers NORBERT LOSSAU, BIRGIT SCHMIDT, BARBARA BAYER-SCHUR, DALE PETERS	95

Cluster 4 / Session 6 – IR and National Licensing

Progress in Developing CAS IR Grid ZHONGMING ZHU, XIAOLIN ZHANG, DONGRONG ZHANG, JIANXIA MA, LINONG LU	107
Building Repository Networks with DRIVER: Joint Explorations with the IR Grid in China WOLFRAM HORSTMANN	116
Explore New Model of National License in China MENG LIANSHENG, ZHENG JIANCHENG	118
German National Licensing Activities: A TIB Perspective MARKUS BRAMMER	122

Progress in Developing CAS IR Grid

ZHONGMING ZHU¹, XIAOLIN ZHANG², DONGRONG ZHANG², JIANXIA MA¹, LINONG LU¹

¹ Lanzhou Branch of National Science Library, CAS, Lanzhou, P. R. China,
{zxm,majx,luln}@lzb.ac.cn

² National Science Library, CAS, Beijing, P. R. China, {zhangxl, zhangdr}@mail.las.ac.cn

Abstract In this paper, we introduce the current practices of building Chinese Academy of Sciences institutional repositories grid (CAS IR Grid). Firstly, we outline its vision and implementation strategies. Then we describe the status quo of spreading IRs across CAS institutes, which covers mainly the release of a formal version of IR building package based on open source software DSpace, establishment of long-term working mechanism of promotion IR service, and the achievements of spreading activities. We also present the development of CAS IR Grid service portal. Finally, we put forward some considerations on future development of CAS IR Grid are explained.

Keywords Chinese Academy of Sciences, institutional repositories, repositories network, repositories grid

1. Introduction

Planning and building an institutional repository (IR) service has become a popular knowledge management practice in institutions. IR is an institutional knowledge asset management platform to help institution set up a centralized space to collect, organize and preserve its knowledge outputs in a systematic manner, thus to avoid the most likely risk of losing them, which now are scattered among various groups and administered by researchers themselves. IR also provides new means and channels for showcasing and disseminating institution's research outputs, which will increase their awareness and rate of use, bring growing scholarly influence and prestige for researchers and institutions [2]. In addition, it can be used to develop practical solutions to enable research assessment activities [3]. Hence, IR is also an important facility and mechanism for institution to develop knowledge capacity and disseminate its knowledge as a whole.

Chinese Academy of Sciences(CAS) has more than one hundred institutes in 28 cities across China and more than 25,000 researchers and 40,000 graduate students (50 percent are doctoral) organization [7], it plays a very important role in Chinese science research and is a major producer of STM information in China. So as a pragmatic strategy for advancing knowledge management practice and advocating open access activities in CAS and China, the CAS IR Grid was brought forth by NSL in 2007. It envisaged helping each institute establish its own local repository as a node of the Grid, and NSL will construct a centralized metadata repository via harvesting and aggregating metadata of academic resources stored in distributed institutes' IRs, and will keep an integrated interface for the aggregated resources and provide other value-added services. At present, CAS IR Grid is progressing smoothly, there are nearly half of institutes having started IR service, and a pilot portal for CAS IR Grid has been launched as well.

2. Visions and Strategies

As strategic actions of pushing open access activities and facilitating actual knowledge management practices, CAS IR Grid bears visions as follows:

- Develop a knowledge management infrastructure to facilitate capture, access, preservation, dissemination of CAS-wide knowledge attainments, and to be an indispensable component of integrated information and knowledge environment across CAS.
- Shape a sustainable knowledge capacity building mechanism for institutes across CAS.
- Bootstrap and foster a culture of open access in CAS and China.
- Facilitate national and international collaboration in development of high-level knowledge repositories network service.

To realize above visions, we take a two-stepped strategy to build CAS IR Grid service progressively (see Figure 1). The overall process is divided into two successive steps, and evolutionary principle is followed within each step as well.

The first step is to build IRs in institutes. In this phase, NSL plays a leading role to provide consultation and support service, and institutes libraries assume major implementation tasks in planning and advancing their own IR service in the help of NSL. Support services provided by NSL are involved in the whole process of implementing IR in an institute, including offering consultation on planning, helping devising policy mechanism and framework, providing IR software packages and technical support service, etc. Meanwhile, we also adopt a demonstration strategy to push work of spreading IRs forward effectively. We select some institutes as early adopters to set up examples, to show demonstration effects and to attract more institutes to follow actively.

The second step is to develop and deploy CAS IR Grid service portal running by NSL. It is some kind of an implementation of an OAI-PMH service provider, can harvest metadata exposed by IRs deployed in institutes via their built-in OAI-PMH data provider interface. Of course, CAS IR Grid service portal will not just act as an OAI service provider, in some extent, it is a grand IR at CAS-wide level, therefore, is necessary to provide advanced browse and search services and other enhanced services.

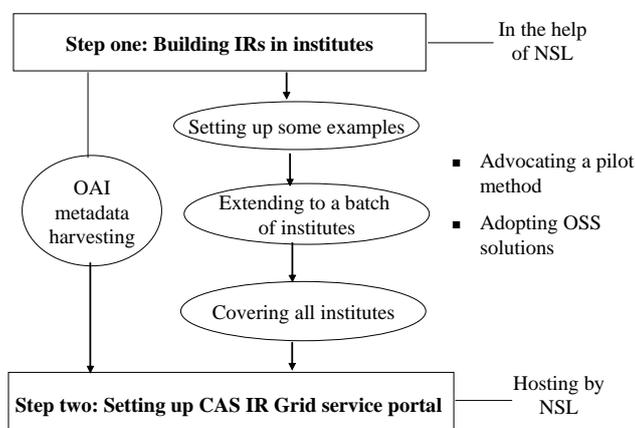


Figure 1: Strategies for developing CAS IR Grid

Concerning technical aspects of implementation of CAS IR Grid, the advantages and disadvantages of making use of proprietary vs. OSS software are hotly debated [1], however, CAS IR Grid is determined to take a strategy of embracing and making use of OSS applications actively.

3. Status Quo of Spreading IRs in Institutes

3.1 Finished Setting up an Exemplar IR for Others to Follow

Best practices are very important for implementation of IRs in institutes. By conducting a pilot deployment of IR, firstly, it will help get experiences in building IR; secondly, it will serve as example thereby stimulating and attracting other institutes to participate actively in the community of building IRs; thirdly, it will help understand accurately the requirements of institutes in building IRs. Therefore, we are very careful in choosing the institute for setting up an example and laid out the principles for doing this work as follows:

- The candidate must have urgent needs of building an IR;
- It can get strong support from the head of the institute;
- The institute library has clear understanding of IR;
- It can guarantee necessary inputs to IR implementation;
- It has comparatively rich experiences in implementation of information applications.

After evaluation, Institute of Mechanics, CAS became our choice. In the collaboration with the institute: (1) we got a clear understanding of needs and requirements from institutes, which laid a basis for later selection, customization and extended development of IR building software. (2) We also have put forward a set of recommended policies framework for referencing and guiding institutes' IR building practice. (3) An exemplar implementation of IR has set up for other institutes to follow.

Figure 2 shows home page of this demonstration IR site (<http://dspace.imech.ac.cn>).



Figure 2: Home page of IMECH IR

3.2 Released an Internal Version of IR Building Package Based on DSpace

As stated above, we are in favor of using OSS solutions in construction of our IR platform. After some comparisons and evaluations, we choose DSpace [5] as a prototype system to build CAS IR platform via a way of extending and optimizing. Main reasons for choosing it are including: (1) it has the largest community of users and developers worldwide. (2) It can theoretically manage and preserve all content types. (3) It has critical core functions we need. (4) It has a layered architecture and clearly defined API, by which can it be easily customized and extended. (5) Using Java platform makes it very convenient to integrate plentiful OSS packages or tools, when needed, in Java world. In general, we think using DSpace is a cost saving and efficient selection.

In a concrete case, we use DSpace 1.4.2. To make it well adapt to meet our needs and requirements, we have done many optimizing and extending development work. Following is a list of some major customizations and extensions:

- It has been highly localized to support Chinese application, not only providing Chinese interfaces, but also handling names of people, order of sorting, search queries, etc to conform to Chinese customs.
- The default metadata schema has been extended to meet the needs and requirements for describing special content types, for instance, patent.
- Additional quick submission workflow is extended to be default alternative to replace its built-in complicated submission workflow. Now it contains just two-steps (pages) to finish one submission in a few minutes, therefore, making submission an easy-doing job.
- The import function of DSpace is enhanced to support bulk import particular formatted data, for example, those legacy data accumulated in excel format.
- An extended function is implemented to support knowledge asset statistics. It can generate knowledge inventories at institute-level, research unit-level, or individual-level, and with combined conditions of researchers' position, status, and publication date of item. The statistics results can be exported with an excel-formatted file to be saved or printed.
- The default installation of DSpace is based on a compiling method containing many steps to be executed manually. This makes the deployment of DSpace-based application a laborious work, and usually needs some technical skills. For deploying an IR application in many institutes easily, we improve its installation process and create one-click installation package, therefore, any one with little computer skill can install the IR package in a few minutes.
- There are other major or minor optimizations, enhancements, or extensions having been done concerning user interfaces, browse and search, knowledge organization, user management, access control, etc.

3.3 Established a Long-Term Work Mechanism for Promoting IR Service

3.3.1 Organizational Structure of IR Promotion

Figure 3 presents an overall organization structure for promoting IR service in CAS.

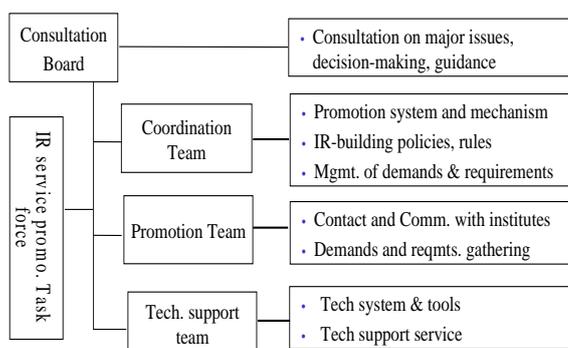


Figure 3: Organizational structure of IR promotion

The consultation board consists of leaders from NSL. Its main tasks are to guide the direction of promotion service, make important decisions involved in overall promotion arrangements, and provide regular consultation service on major issues concerning the promotion activities.

The coordination team is mainly responsible for overall organization, coordination and collaboration tasks involved in the process of promoting IR service. They include proposing a set of policies templates to be referenced by institutes; capturing, analyzing and allocating efforts to meet service needs from institutes; and establishing work mechanism for promotion of IR service.

This team is composed of members from NSL and institutes. Those from NSL are usually head of coordination groups of NSL. They coordinate efforts to meet the requirements that need their support. In addition, members from institutes are usually directors of libraries of institutes as representatives of institutes to propose requirements.

The promotion team is composed of subject librarians from department of subject information service of NSL, and is responsible for concrete IR promotion tasks. Each member of this team will take charge of several institutes' IR promotions. They are required to provide following services:

- Pre-consultation service: it may cover any issues concerning IR building.
- Help the institutes plan an IR service.
- Help the institutes formulate an adapted policy framework based on what policies templates proposed by the coordination team.
- Help the institutes work out feasible work procedures for running IR service.
- Help the institutes install CAS IR package.
- Provide training in using and maintaining IR system.
- Help the institutes prepare legacy content and transfer them into IR system.

The technology support team consists of IT engineers from IT department of NSL. Their responsibilities concerning following two aspects:

- Develop software package and tools in support of building IR. As described in section 3.2, we need to customize, improve and extend DSpace to be a more suitable application for us. In addition, we also need to add new features constantly to keep up with the development of IR itself.
- Provide technical support service involved in IR promotion and application activities. Regular technical support services may covers a range of services, such as installation

of IR package or tools, training, guiding the use of application, customization, distribution of software patches or updates as required to resolve specific issues/problems, providing technical advices or resolving any other specific inquires or technical problems. These services usually are required to be provided in a timely manner via telephone, email, instant messages, support site, or remote desktop technology.

3.3.2 *Basic Working Procedures for IR Promotion*

Organizational structure clearly defines division of responsibilities and sets up communication and collaboration mechanisms among working teams. However, in order to conduct IR promotion service methodically, we have established following official working procedures to guide the smooth running of the promotion.

- First, with the help of responsible subject librarian from NSL, the institute library on behalf of the institute submits an application request to NSL, namely consultation board of the project. This application should include clear-cut definition of strategic goals of building IR service, definite team organization, commitments to establish effective policies and supply supportive conditions and facilities, such as funds, basic hardware and software, etc.
- When consultation board receives an application, it performs a routine assessment of its feasibility, checks whether required conditions are met, and finally draws a conclusion if or not the application is passed through.
- If an application is approved, NSL assigns an agreement with the institute. As an official engagement between two sides, it prescribes bilateral responsibilities and obligations and is taken as a starter of building IR service in the institute.
- Finally, we enter procedure of bilateral practical co-operation. Those teams of NSL and the team of the institute will work together to push the implementation of IR service in the institute. During the course of this procedure, subject librarians will get much involved in consulting and helping institute work out a detailed implementation plan and finish constituting an adaptable policy framework; then the institute perform preparations such as hardware and software environment and other required pre-conditions to get ready for installation of IR package; technical support team help institute complete tasks such as remote or on-spot or self installation of IR package, post-configuration, training, trial running, official lunching, and other technical ones.

3.4 **Achieved an Expected Level of Deployment**

Until now, over 50 institutes have assigned building agreements with NSL, and nearly 40 institutes have launched IR service.

Among them, some have gradually established sustainable support mechanism for the development of IR service. For instance, the Institute of Semiconductors has incorporated the developing IR service into its informationization development, and established mechanism of combining content submission with assessment and reward of graduate students, researchers, and research units, to stipulate and encourage individual participation in IR building. The Institute of Software, the Dalian Institute of Chemical Physics and others have also set up similar mechanisms.

Concerning the content collection building, peer-reviewed journal articles, conference papers, theses and dissertations are the most concerned scholarly content. Many IRs also

have included presentations, books, book chapters or sections, patents, project achievements, and other materials. Moreover, research data are of most future concerned research outputs to be included in IRs.

In respect of content recruitment, it is still in an initial stage overall. Except some IRs have reached a several thousand-level, many of them have totally accumulated less than one-thousand items.

4. Development of CAS IR Grid Service Portal

4.1 Architecture of CAS IR Grid

CAS IR Grid service system itself can actually be regarded as a CAS-level IR system. Compared with institutes' IRs, it focuses on CAS-wide knowledge content aggregation, as well as knowledge dissemination, communication and discovery. Collection, deposition and preservation of knowledge artifacts will not be its essential services. A three-layered architecture of CAS IR Grid is presented as in Figure 4.

The content layer consists of IRs distributed in institutes. They are the nodes of the Grid, and each has built-in OAI-PMH data provider interface and can expose their metadata with default oai_dc format or more applicable qdc format to be used in the Grid system.

The aggregation layer provides metadata aggregation service. The harvester component collects metadata records from OAI-PMH data providers, in here are institutes IRs, and creates a central metadata repository, which serves as a base on which various tools and services can be developed.

The interface layer provides not only user-oriented services but also application-oriented services. User services will include basic and conventional services such as browse and search service. In addition, it will support services like dynamic creation of knowledge communities, institute profiling, generation of CAS-level knowledge inventories, and cross-subjects or cross institutes data linking, integration and discovery service.

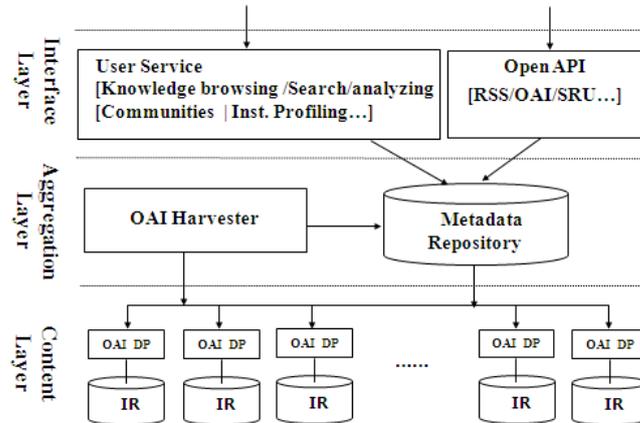


Figure 4: Architecture of CAS IR Grid

The application-oriented services are mainly standard-base interfaces, such as open interfaces of OAI-PMH, RSS, SRU/W, etc.

4.2 Development of Harvest and Aggregation Service System

We continue to follow our OSS strategy in development of harvest and aggregation components of CAS IR Grid. After a preliminary investigation and evaluation, two candidates were left, they are D-NET [6] and OAI ORE patch for DSpace [6].

Actually, we much prefer D-NET. It has powerful and attractive service oriented architecture, and is not just a harvester system; in fact, it is an open-sourced software package with much more complex functionalities to support running of a repositories network. Moreover, it is support platform of DRIVER Repository network. From the point view of ongoing collaboration between DRIVER and CAS IR Grid, it is a best selection for using it as basic platform for CAS IR Grid. Thus, as part of collaboration between two sides, DRIVER installed a test D-NET application in our local servers.

However, current D-NET 1.1 as an interim version is still under development, so it is not so stable to use it in a production environment. Moreover, from the point of view of using DNET in China, we need additional Chinese interfaces. But current version of D-NET has no built-in I10n/i18n support, so if we want to use it, we have to hack source codes to gain a localized version beforehand. Apparently, doing like this is time-consuming and not sustainable as well.

Therefore, we temporarily use alternative solution of building CAS IR Grid service portal, i.e. using OAI ORE patch for DSpace. This patch can function as a basic OAI-PMH harvester and an OAI ORE harvester, and we just use it as the former. It is relatively easy to be integrated into DSpace. Based on the integration, we have set up a test CAS IR Grid service portal (see Figure 5). However, the patch is just an implementation of OAI harvester; it is worth looking forward to new release of D-NET, because we are informed that the new release will have updates we need.



Figure 5: Test site for CAS IR Grid service portal

5. Future Considerations

Developing CAS IR Grid service is a long-term task, we will continue to try our best to push it forward. Our future concerns may exist in following aspects:

- Develop successive enhanced version of IR package. It is expected to support wide-expansion of knowledge content types, provide further search and discovery features, enhance capturing context and relationship between knowledge objects, and advance its composability and interoperability to integrate or to be integrated with other services easily.
- Spread IRs in most institutes across CAS. Currently, less than half of institutes have launched IR service; we need to help rest institutes join our IRs community in next years.
- Upgrade CAS IR Grid service continually. We will keep it harvesting IRs across CAS as many as possible, and develop advanced search and discovery functionalities.
- Make concrete progress in metadata sharing with DRIVER. In addition to current collaborations, which are particularly in the area of technology, we both sides need to promote bilateral collaboration to metadata exchange level.

6. Acknowledgments

We thank DRIVER, in particular, Dale Peters, Marek Imialek for their help in providing many consultation and support services involved in evaluating D-NET. We would also like to thank for hard work done by our colleagues from NSL.

7. References

- [1] Amaral M (2008) Institutional Repositories, Open Source Options, and Libraries. *Open and Libraries Class Journal*, 1(1)
<http://eprints.rclis.org/15355/1/Amaral.pdf>
- [2] Crow R (2002) The case for institutional repositories: a SPARC position paper
http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf
- [3] Day M (2004) Institutional repositories and research assessment: A supporting study for the ePrints UK Project
<http://eprints-uk.rdn.ac.uk/project/docs/studies/rae/rae-study.pdf>
- [4] D-NET
http://www.driver-repository.eu/D-NET_release
- [5] DSpace
<http://www.dspace.org>
- [6] OAI-PMH & OAI-ORE harvesting support
<http://jira.dspace.org/jira/browse/DS-289>
- [7] Zhang XL, Liu XW, Li L (2009) NSL OA Implementation: Promoting the Development of Sciences
<http://conference.las.ac.cn/Sino-German/2009/DOC/Session5/14.pdf>