CAS IR Grid: An Institutional Repositories Network in CAS

Zhongming Zhu, Ma Jianxia, Lu Linong, Wu DengLu, Wei Liu, Xiaojia Ma
Lanzhou Branch of National Science Libray, CAS

Dongrong Zhang, Xiwen Liu, Lin Li, Li Wang
National Science Libray, CAS

Introduction

The Chinese Academy of Sciences has more than 100 institutes and institutions located in 23 provinces and municipalities throughout China. When institu-tional repository service was introduce into CAS in the second half of 2007, a network of institutional repositories named CAS IR Grid was brought forth by National Science Library of CAS (NSL). It envisages that each institute establishes its own local repository which is treated as a node of the Grid, NSL constructs a centralized metadata repository via harvesting and aggregating metadata of academic resources stored in distributed institutes' local IRs and an integrated interface for the aggregated resources so as to provide value-added services. Currently, CAS IR Grid is progressing smoothly, there are more than half of CAS institutes having initiated IR service, and a pilot portal for CAS IR Grid has been launched as well in the June of this year.

Strategies and Framework

CAS IR Grid is to be developed to be: a knowledge management repository to facilitate capture, access, preservation, dissemination of CAS-wide knowledge attainments; an indispensible component of integrated information and knowledge environment across CAS; a sustainable knowledge capacity building mechanism for institutes across CAS; a culture of open access in CAS and China. Furthermore, it may facilitate national and international collaboration in development of high-level knowledge repositories network service.

The hub-and-spoke model was taken as basic framework to guide the developement of CAS IR Grid. In this model, different institutes or geographies are allowed to maintain separate content repositories, OAI-PMH standard based connectors are used to connect disparate institute repositories to a central data integration platform. Users can then access content through a single interface, namely federated repositories portal. Fig. 1 presents an illustrated diagram of the model.
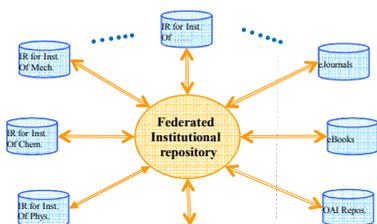

Fig. 1 Hub-and-spoke model for the development of CAS IR Grid

Functional Architecture

As shown in Fig. 2, CAS IR Grid adopts a three-layered architecture. From bottom to top, the content layer consists of IRs distributed in institutes and each has a built-in OAI-PMH data provider interface, which

exposes IR's metadata with default OAI DC or more applicable metadata format to be used by central repository in its upper layer. The aggregation layer has main components of the OAI-PMH harvester and the central repository, which harvests and aggregates metadata records and even content objects from institutes IRs. The interface layer provides various user services, such as browse, search, alerts, recommendation, knowledge inventories, statistics, etc. Furthermore, there are standardized interfaces provided in this layer, which support machine-to-machine or application-to-application data exchange or service interactions. For example, OAI-PMH, SRU, OpenSearch, SWORD, etc.
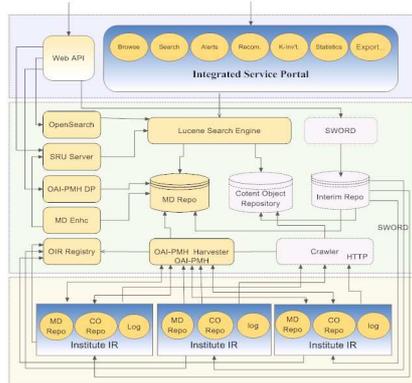

Fig. 2 Functional architecture of CAS IR Grid

Software for CAS IR Grid

In general, OSS solutions and patchwork prototyping method are adopted for the construction of CAS IR Grid platform. After some evaluations, DSpace is chosen as a prototype system to develop CAS IR Grid platform with extending and optimizing functions.

Firstly, CAS IR software is developed by carrying out following major customizations and extensions:
- Chinese compatibility localization
- Metadata schema extension
- Submission workflow modification
- Enhanced browsing and retrieval functions
- Advanced import function
- Added statistics module
- Improved OAI-PMH implementation
- Simplified installation
- User interface optimization, etc.

Secondly, based on the above entended version of CAS IR package, the platform of CAS IR Grid portal is developed by mainly adding modules to support metadata harvesting and aggregation. In practice, the OAI ORE patch for DSpace is adopted and integrated into CAS IR pakage. Beside its basic implementation of OAI harvesting fucntion, additional modules, namely IR registration management, metadata normalization, and duplication checker, are developed to extend and enrich functionalities of CAS IR Grid portal.

Development Status

To promote IR service across CAS, NSL establishes special team composed of subject libarians and technical staff members to provide help and services involved in building IRs in institutes. Currently, there are 63 institutes have installed CAS IR packages or launched IR service, and half of them are open to public access. Fig. 3 gives some deployment examples.

Fig. 3 Exampiar IR home pages of SIOM, SEMI and IMECH

CAS IR Grid portal is also progressing well, which now has a growing list of 34 IRs harvested and more than 85 thousand metadata records are collected. Fig. 4 depicts home page of CAS IR Grid portal.

Fig. 4 The home page of CAS IR Grid Portal(http://www.irgrid.cas.cn)

Future Considerations

Developing CAS IR Grid is a long-term task and is still in initial stages of the practice across CAS. Besides the continuous improvement of the software and platform, especially in support of management of compound content objects, semantic enhancement and relationship based analysis, much further work is required on increasing resonance and sustainability of IR service in institutes and among scientists community.

Acknowledgements