BULIDING CHINA ACADEMIC DIGITAL LIBRARY

Feng Ying

Administrative Center, China Academic Libraries & Information System (CALIS),
Peking University, Beijing 100871, China
Email: fengy@calis.edu.cn

ABSTRACT

CALIS is a key-player in resource-sharing in China. This paper gives overview of CALIS and service network CALIS constructed. Then it gives the infrastructure of China Academic Digital Library Information System (CADLIS) in details.

INTRODUCTION

China Academic Library & Information System (CALIS) is funded primarily by the Chinese government, and operates under the leadership of Ministry of Education of China. CALIS is one of three public service system of national higher education, as China Academic Library consortium at the same time, intended to be a nonprofit organization in the future. The mission of CALIS is to promote, maintain and improve library resource sharing among Chinese universities, academic libraries, other libraries and information institutions. CALIS have made lots of efforts to build an open and distributed infrastructure. Full text databases on e-journals, thesis and dissertation (ETD) and reserved e-book, and abstract and Index databases of Union Catalogue, CALIS Current Contents of Western Journal, "Key discipline" and "network resource navigation" are the most important databases that CALIS imported and developed. Services that CALIS provided to member libraries is emphasis on cooperated database purchasing, union catalogue, interlibrary loan and document delivery, digital library model designing and training.

During first five year, started in 1996, CALIS established 3-tiered information service network that includes 4 national, 8 regional information centers and over 400 member libraries. At the same time, CALIS established information resource network that consists of imported and self-built resources as well. CALIS now is in its phase two, started in 2001 and will end in 2005. Goals of phase two is to establish the framework and infrastructure of an open distributed China Academic Digital Library (CADLIS) over thousand universities, to provide services for higher education, academic research and key disciplines, and to promote to make and use standards and specifications of digital libraries in China. So far, there're over 600 member libraries.

SCOPE OF CADLIS CONTENT AND FUNCTIONALITY

This section characterizes the CADLIS scope to describe that content and functionality it has. CADLIS is a distributed information environment, comprised by many digital resources and a extensible set of services.

Users
End-users are viewed primarily in their roles as educators and learners. The CADLIS information environment intends to provide not only information discovery, access and personalization service, but also virtual reference, research support, teaching reference and professional training services. End-users can get these services by CADLIS portals.

Digital Contents
Before the foundation of CALIS, there was little attention paid to the resources born digital. The efforts of CALIS makes digital resources become more prevalent in academic libraries in China. By the end of 2005, digital contents in CADLIS will be 30,000 Chinese and western journals, 20 million records of journal articles, 3 million bibliographic data and 6 million holdings in online union catalog, 1 million thesis and dissertation electronic full-texts, 20,000 reserved e-books, 300 electronic resources databases

and etc.

CADLIS uses OAI-PMH to harvest metadata from individual libraries, forming a centralized metadata database which is the one of basis of discovery service. Although metadata for different digital collections may not be stored in one physical location, for example, metadata of ETD is stored in Tsinghua University and metadata of special collections is stored in Wuhan University, it can be viewed as logically centralized. To speed up brief information display, some digital objects, such as thumbnails and first 12 pages of ETD, are harvested at the same time.

Digital objects, such as ETD full-text, are stored in and managed by dispersed institutions. End-user can access logical centralized metadata database to discover what they want and then request digital objects from a library.

Services
The basic services must include human interfaces that implemented by portal, plus support for discovery, access and personalization. Services for authorization and authentication, pay per view, and billing are included in CADLIS also. Some services, such as virtual reference, intended to be used or integrated in any application system of CADLIS.

SERVICE NETWORK OF CALIS

Since 1996, CALIS has established a 3-tiered service network, shown in Fig.1. National and regional centers are in charge of providing information and services at a national and regional level. To strengthen CALIS service capabilities and extend services to all universities in China, we built 15 provincial centers in phase two. Digital library test beds are responsible for research on application of key technologies in digital libraries, exploring new service models of digital library and etc.
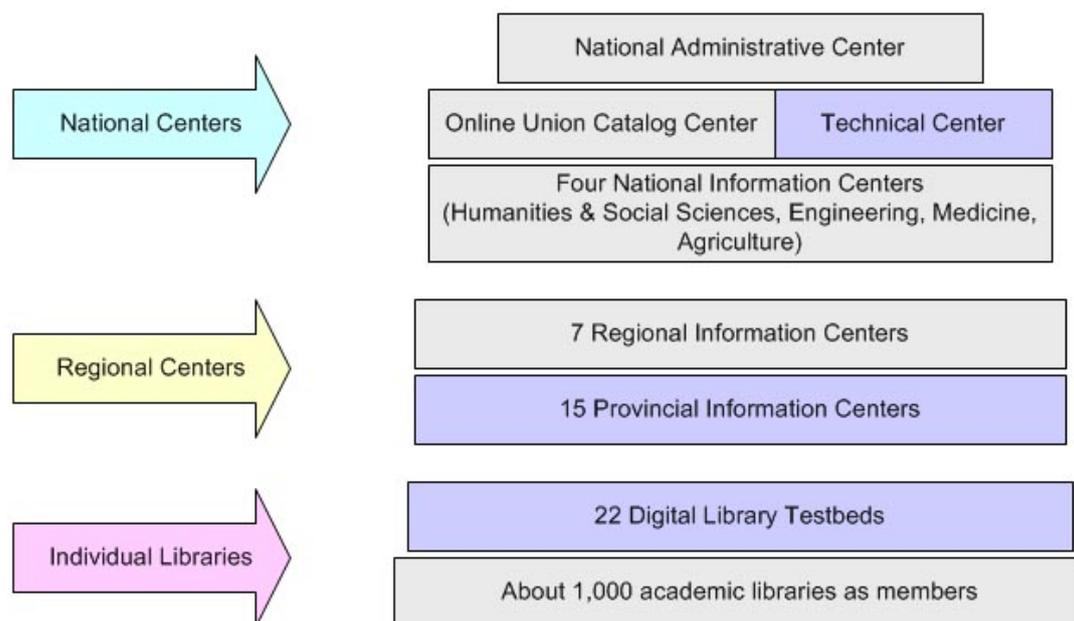
Fig.1 3-tiered organization of CALIS

STANDARDS AND SPECIFICATIONS

CADLIS is not a stand-alone digital library, but an open, distributed and cooperative digital libraries that include integration of many application systems. To give guidance to design, development and implementation of different application systems, to make sure that every piece of projects can be integrated seamlessly as a whole digital library by using the same technical standards and specifications, CALIS started research on international standards, specifications and key technologies on digital library in 2002. In February 2004, CALIS released the first draft version of standards and

specifications. Then the second version was released in November 2004. Now many of them are applied into CALIS projects.

These standards and specifications are guidelines for digitization and archival storage for resources, metadata specifications for different collections and their interoperability, specifications for digital object identifiers, interface for unified search, authorization, authentication, logs and etc, specifications for evaluation on resources and services, specific specifications and interface for application systems and so on. CALIS also takes OAI-PMH, OpenURL, METS as basic standards. Especially, metadata specifications are based upon the national project Research of Standards of China Digital Libraries funded by Ministry of Science & Technology.

TECHNICAL INFRASTRUCTURE

Goals of technical infrastructure are to build a distributed and cooperative information environment that support contents and services described above. The technical infrastructure must support flexibility, scalability and extensibility along all of these dimensions. And the infrastructure should enable CADLIS to be able to inter-connect with other national digital libraries.

Technical architecture of CADLIS
Fig. 2 shows the technical architecture of CADLIS. There're five levels that given different services.
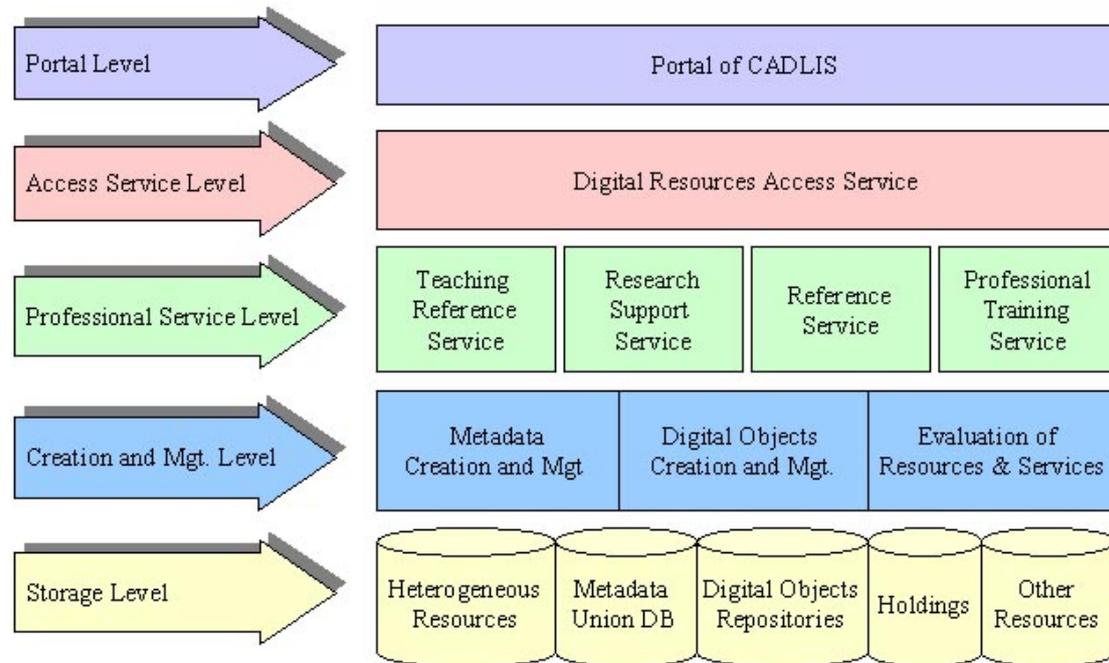
Fig.2 Technical Architecture of CADLIS

There's not only one national portal in CADLIS, but also many cooperative portals. For example, portals for engineering, humanities & social sciences, are all universities and discipline oriented. Portals for regional centers are regional area oriented and has End-users can interact with CADLIS via one portal and get more information and services from other portals transparently.

Here heterogeneous resources refer to those who are purchased and served by different systems. Metadata union DB refers to the centralized metadata database including self-built contents and resources that bought in raw data.

Deployment of CADLIS System
CADLIS system can be deployed at three levels according to the 3-tiered service network, shown in Fig.3. National portal are responsible for discovery and locate where resource are and how end-users can access. There're same applications of CADLIS both in national portal and regional portal, such as

union catalogue. What the difference is that only catalogue records from local area are stored in regional union catalogue, at the same time national portals holds catalogue records from all universities in China.

CALIS provide digital library system platform that can be tailored to its member libraries. The functionality that system has includes discovery and locating, management of digital contents, authorization and authentication, billing, logging and others.
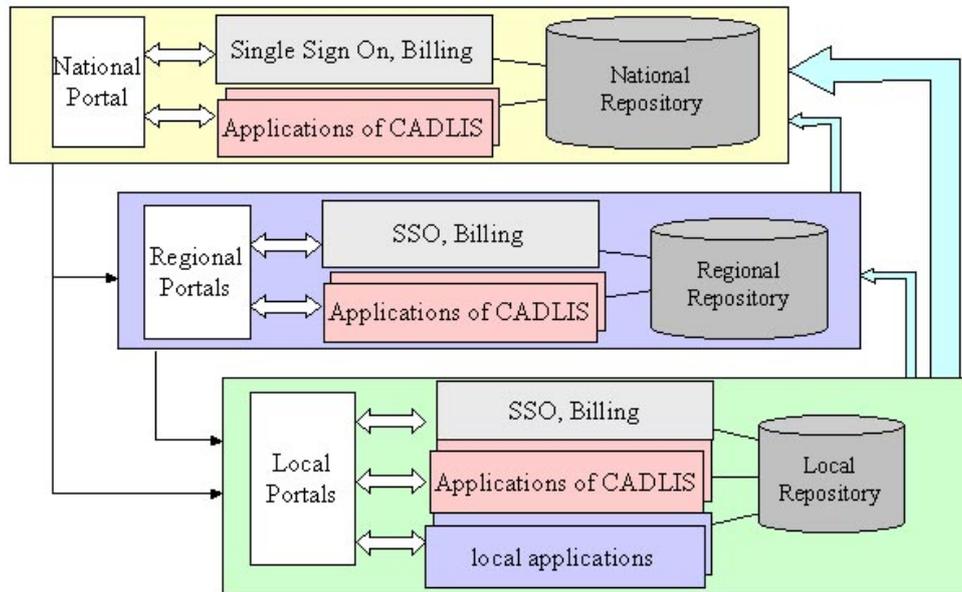
Fig. 3 Three Levels of Deployment of CADLIS

Digital Resource Access Service
Digital resource access service is the most basic and important service in CADLIS. It intents to be what you find is what you get. The following will give the architecture of digital resources access service in details.

Shown in Fig.4, the basic services for discovery, access and personalization are provided.
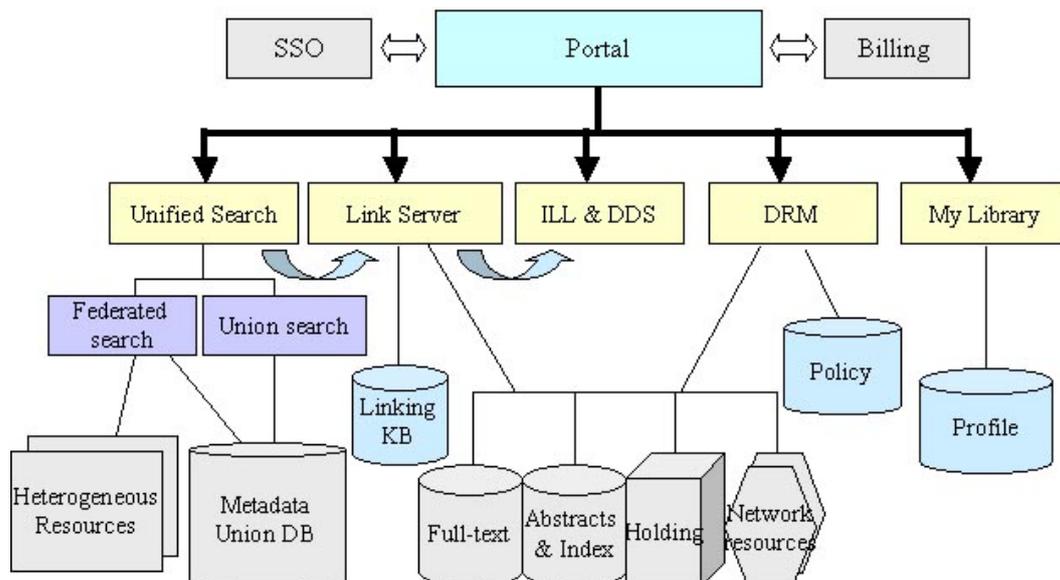
Fig.4 Architecture of Digital Resource Access Service in CADLIS

End-users can log on a portal to access to licensed and free resources. With single sign on function, users are able to access to different systems without multiple login. If user wants to access resources which his institution doesn't have, he can still get what he wants by pay per view.

User can use unified search or search function provided by specific applications. Results searched by unified search can be linked to other related resources, such as full-text, abstract and indexes, holdings and others, by CALIS link servers through OpenURL. Interlibrary loan and document delivery service is provided at the same time. For full-text like ETD, digital rights management (DRM) technology is applied to make sure the legal access to contents according to the access policies. User can also tailor his digital resources and services by my library function.

CALIS made its own unified search protocol, called CALIS ODL-MAP, based on Open Digital Library and extends to more functions, defining search, browse, index and other requests and responses in XML. Now CALIS unified search system supports Z39.50 sources, HTTP sources, CALIS ODL-MAP sources and those who can be accessed by internal interface. Fig.5 shows the interface of unified search system.
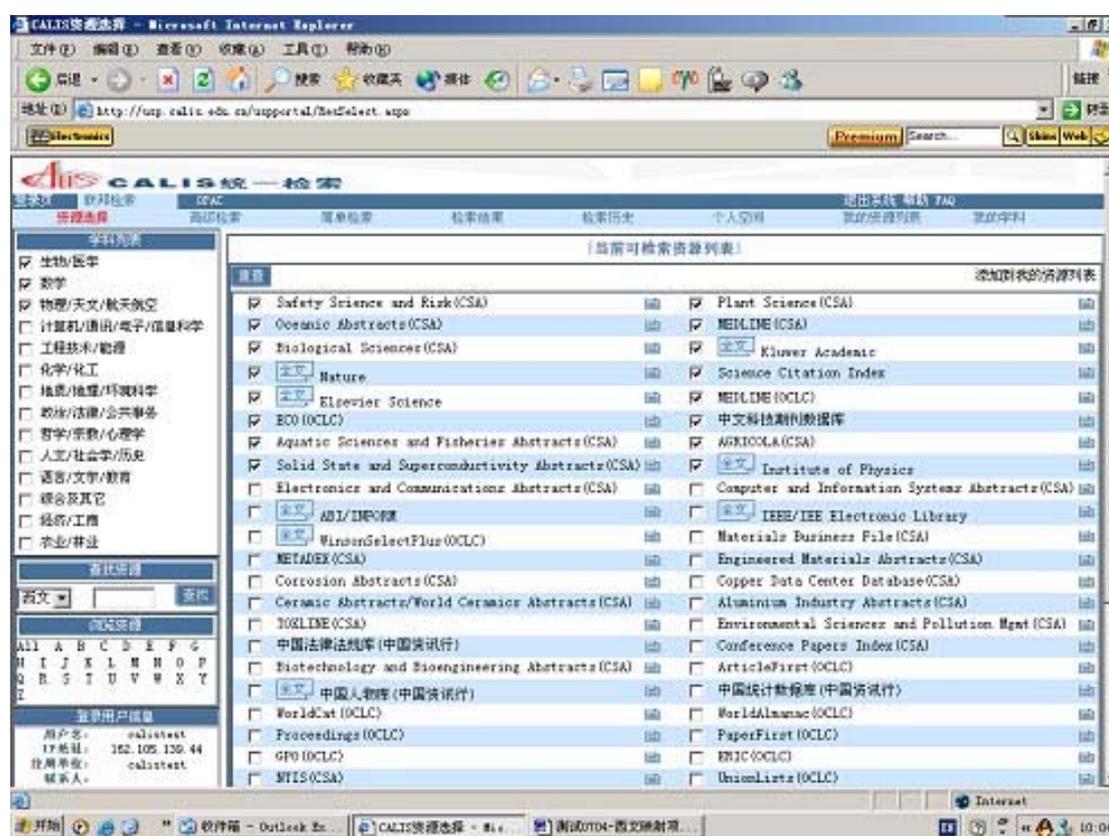
Fig.5 The Interface of CALIS Unified Search System

SOME OTHER SYSTEMS IN CADLIS

In this section, brief introduction of some important systems in CADLIS will be given.

Virtual Reference
CALIS pay lots attention to reference service and intends to make it applicable to any portals in CADLIS. CADLIS virtual reference is a two-level distributed service network including one national reference center and a group of local reference desks. This service network mainly relies on local librarians to provide services in cooperation with other librarians. Only recognized and qualified FAQ can be harvested or sent into central knowledge base. For libraries who are not members of CALIS virtual reference network, can rent a desk in national reference center for their local end-users.

Now there're 12 libraries joined. By the end of 2005, it supposes to be 20,000 FAQ records in central knowledge base.

Teaching Reference
To satisfy the needs for teaching materials and reserve books, CALIS built a distributed online teaching reference information and electronic reserved books databases. Teaching reference information includes courses, teachers, teaching materials, bibliographies and reserved book, allowing teachers and students to search and retrieve, covering major disciplines such as humanities and social sciences, engineering, medicine, agriculture, etc. Users can access to licensed electronic reserved books or read it online. Electronic reserved books are served in CEB format, developed by a local company, using DRM technology to protect authorized copies of materials. E-reserve is not considered in CADLIS phase two.

It will be 20,000 electronic reserved books in Chinese and 40,000 teaching reference information records by the end of 2005. So far, there's 53 universities joined this system.

Research Support
Research support system is one pilot projects of digital library test beds taken by Peking University. Its goal is to build an e-print and web publishing system for academic libraries, to build online communities for researchers for their cooperation and academic communication, and to facilitate sharing and commenting resources for researchers across disciplines as well as between universities and other institutions. Some contents are contributed by researchers themselves, some are based on digital resources provided by CALIS and academic libraries.

Online Union Catalogue
CALIS Union Catalogue database is the largest online bibliographic database in China, merged bibliographic records of 40 CALIS member libraries and holding records of 200 member libraries. It is also the first bibliographic database that can support data download and upload throughout the network. CALIS Union Catalogue database is not only one of CALIS services, but rather provides the foundation for CALIS interlibrary loan services. All records are stored in UTF-8.

By the end of July 2005, there's over 1,780,000 bibliographic records in CALIS Union Catalogue database, including books and journals in Chinese, western languages, Japanese and so on. So far, there's over 500 member libraries.

CHALLENGES OF CADLIS

CADLIS is such a large project that may meet many problems and challenges. Listed below are challenges that CALIS has faced.
- Can CADLIS services be integrated seamlessly? How to deal with legacy systems?
- How to make CADLIS resources and services sustainable to end-users?
- How to ensure the appropriate copy that link server linked to?
- How to ensure data integrity under CADLIS highly distributed environment?
- How to prevent access to inappropriate materials?
- How can CADLIS facilitate authentication through all over universities in China?
- How can systematic measures of usage of resources and services be employed to evaluate or improve the quality of CADLIS resources?
- Must special steps be taken to make sure long term preservation of CADLIS resources?
- What should be done to make the system reliable enough to provide 7x24 services?

Answers to some challenges might be found in technologies. CALIS is still working on this.

SUMMARY AND CONCLUSIONS

After 9 years development, CALIS has made significant progress in resource sharing among universities in China. Member libraries benefit a lot from CALIS. CADLIS enrich resources and extends service to more and more libraries and users. Through many efforts for resource sharing have been done, CALIS still faces with the challenges to provide services that are relevant to its members.

CALIS will continue to help its members to predict emerging needs and to promote the resource sharing action among member libraries to a higher level.

REFERENCES

[1] **Yao Xiaoxia, Chen Ling, Wang Xiaomei, "Current Status of Digital Library Construction in CALIS Member Libraries",** *Journal of Academic Libraries***, Volume 21, No.3, p. 21-25, 2003.**

[2] http://www.calis.edu.cn

[3] **David Fulker, Greg J., Hernon, P.** *Components of an NSDL Architecture: Technical Scope and Functional Model***. Portland, ACM/IEEE Joint Conference on Digital Libraries, JCDL 2002**