



## **Open Access Resource Management: a New Role for the National Library of China**

**Wang Zhigeng**

E-mail: wangzhg@nlc.gov.cn

**Liang Huiwei**

E-mail: lianghw@nlc.gov.cn

**Zhang Wenjing**

E-mail: zhangwj@nlc.gov.cn

Digital Resource & Service Department  
the National Library of China  
Beijing, China

**Meeting:**

**132. National Libraries with CLM**

---

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY

10-15 August 2010, Gothenburg, Sweden

<http://www.ifla.org/en/ifla76>

---

### **Abstract:**

*With the development of OA movement in China, OA resources, mainly OA journals and OA Repositories, have been an unprecedented development. These resources not only facilitate the access to academic resources for researchers, but also have brought new issues to the management of those resources for Libraries. Libraries are facing challenges on bibliographic control and long-term preservation of OA resources. In this situation, the National Library of China (NLC) launched a project named Chinese OA Resources Portal (COARP) to collect and preserve the primary OA resources in China and provide discovery service for the user in an aggregated manner.*

*This article introduces the current status of Chinese OA resources and the issues the libraries facing. It discusses the topics of collection development, preservation and service delivery of the COARP project, with its focus on the principles of resource selection, content integration and method of long-term preservation.*

### **Keywords**

*Open Access, access policy, National Library of China, COARP*

---

## **1. Introduction**

OA is a mechanism of advocating fair principle to promote the free exchange of academic ideas and the resulting free access of literature information in the network environment. In recent years, China's OA is on the rise: on December 9, 2003, Lu Yongxiang, president of Chinese Academy of Science President signed the Berlin Declaration in Berlin on behalf of Chinese scientists; on May 24, 2004, Chen Yiyu signed the Berlin Declaration on behalf of National Natural Science Fund Committee, this shows that the scientific community and the technology funding agencies of China support OA. The emergence of a large number of OA resources such as The Miracle Library and Chinese Scientific Papers Online, not only provides a new way for researchers to publish, share and use academic resources, but also becomes an important part of library literature resources.

Chinese OA resources is increasing, but its organization and management is scattered and disorderly, and there are various problems such as scattered resources, range of species, short life, various application procedures, complex and varied data structures and organization styles, different search ways and display interface, so it lacks an effective mechanism of management, service and long-term preservation.

How to manage effectively and preserve this part of digital resources long and properly, its significance is not only able to adapt to the information needs of today's society, but meet the needs of future customers understanding the social situation and studying the historical experience of the current social development, and play a role in inheriting national and human culture. National Library of China serves as the repository of the nation's publications, should take the responsibility actively to carry out a unified plan for the management and preservation of Chinese OA resources from the national level, to establish appropriate management mechanism and preservation system, and to construct a technical framework model conforming to international standards with continuity and scalability to achieve efficient and reliable national management and preservation.

## 2. The Development Status of Chinese Access Resources

China has already embodied the spirit of open sharing almost from the day the Internet emerged, but the true OA movement began in 2003, China has undergone a process from awareness to practice.<sup>[1]</sup>

In the area of theoretical research, Chinese scholars began to pay attention to OA and publish papers from 2004. China Journal Full-text Database (CJFD) has published more than five hundred papers on OA research in the years of 2004-2010, and it shows an increasing trend (see Table 1). China has also undertaken a number of research projects on OA and published a number of research reports, and a number of foundation projects including the National Natural Science Foundation and the National Social Science Foundation have financed OA research topics. The study runs through the entire life cycle of OA, researchers focus on libraries and universities, research focus is mainly theoretical study based on OA, OA and libraries, OA publishing and copyright issues, practice progress and problems and so on.

**Table 1: Year 2004-2010 Tables of the Number of Papers on OA Research in China Journal Full-text Database (CJFD) Unit: Piece**

<b>Year</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>Total</b>
<b>Journal Articles</b>	3	24	75	116	112	150	21	501
<b>Thesis</b>	0	1	2	4	9	8	0	24
<b>Conference Papers</b>	0	1	4	3	2	4	0	14
<b>Total</b>	3	26	81	123	123	162	21	539

In practice, a group of young scholars created a Miracle Library in August 12, 2003. National Science and Technology Library established "Chinese Preprint Center", "Foreign Preprint Gateway" and "The Integrated Retrieval System of OA Journals". "China Scientific Papers Online" organized by the Science and Technology Development Center of the Ministry of Education(as of May 2010) has published

1 New Development of OA At Home and Abroad, Chu Jingli, Library Forum, June 2009

433005 papers online, the OA online resource integration platform has provided users with a variety of search services. Many agencies have established OA Resource Library one after another. National Library of China established "Open Library of Library and Information Science" to provide a free, real-time, open platform for the researchers and the staff of the general library community. In recent years, National Science Library of Chinese Academy of Sciences (CAS) has taken a series of measures on OA and established a number of knowledge bases to promote OA development actively.

Although the positive changes of China's OA theories and practice have taken place, but the problems can't be ignored. Throughout the actual development of international OA, frankly, OA in China is only in the initial stage, Chinese parties' attention on the OA isn't high, the participation in OA is not enough, there is still a big gap between China's OA journals and which of the world in the number, the number of OA journals, institutional repositories and subject knowledge bases is still limited, China's OA shoulders heavy responsibilities.

### **3. The Problems Faced by Chinese OA Resources**

#### **3.1 Resources Are Disordered and Lack of Effective Management**

Currently, the establishment of OA journals and OA Repositories is in the state of disorder, whose establishment and management depend entirely on institutions or individual will, none of the authoritative bodies or groups carries out effective management of the establishment and management at home and abroad, this not only leads to differences in the editing and publishing policies of OA journals, but also leads to differences in the software system functions used by OA Repositories. Moreover, this also causes that the resource development policies implemented by each library isn't uniform, metadata specification is inconsistent and the format of object data is not uniform, etc. It brings great difficulty to the management of OA resources, and also makes the management of OA Repositories an important problem

demanding prompt solution in the area of information management.

### **3.2 Metadata**

OA journals provide metadata associated with papers, OA Repositories also provide the cataloging and retrieval of metadata, but all metadata of repositories has not a well authority control currently, almost all the metadata records of journals and repositories do not contain classified information and quotation information. In order to improve the comprehensiveness and the accuracy of journals' and repositories' metadata retrieval services and facilitate the higher level to integrate and apply metadata, it's necessary to carry out authority control of OA journals' and repositories' metadata using the information indexing techniques and tools of libraries, including but not limited to name authority control, subject authority control, standardization of citations and bibliographies and so on.

### **3.3 Diverse Resource Forms, Insufficient Technologies**

The development of OA has experienced a variety of forms, the most important resource types include institutional repository, subject repository, self archiving, the open journals paid by authors, part of the open journals of OA, the open journals of delayed OA and blog, wiki, personal web sites, e-books, e-mail list services, forums, RSS seed, document sharing networks of P2P and so on, all aspects of OA resources involve technical issues from the processing, construction to the management, preservation and service of OA resources in the entire life cycle.

### **3.4 Resources Lack Long-term Protection Mechanism**

The common concern of users is obtaining OA resources long and effectively.

For the academic resources, users need a long-term and effective protection mechanism, so it not only considers long-term operation maintenance and system upgrade technically, but also takes into account the long-term effectiveness of the resources from the management.

For the OA resources in the non-traditional academic area, the problems they face are much more complex. The high dispersion of the resources stored on the network, the random of content changes and technical standards, especially the uncertainty of the legal property of intellectual property rights, bring many problems to the discovery, acquisition, management, use and long-term preservation of the resources.

In short, OA all require in-depth, systematic study from macro to micro, from management to services, from legal issues to technical issues.

#### **4. National Library's Role in the OA**

Libraries undertake the collection, sorting, organization, preservation and service functions of cultural information resources, the spirit of academic information equality and free dissemination advocated by OA is consistent with the spirit advocated by libraries, OA resources have become an important part of library literature resources, these decide that there is a close relationship between OA and the future development of libraries.

At present, none of the authoritative bodies or groups carries out effective management of OA resources. As the chief library in national libraries, National Library of China serves as the repository of the nation's publications, it should take responsibility initiatively to make full use of their advantages in resources and cooperation advantages in the field and technical force, to use the theories, technologies, methods and tools of information resource management to carry out scientific and effective management of OA resources on the production, dissemination, preservation and use throughout the life cycle.

##### **4.1 Establish a Registration Institution for OA Resources**

OA resources' registration system is an authority information system reflecting the creation, existence, development and demise of literature resources, is a central database system revealing the metadata attributes of literature information resources

on technology, content, right, management and service. It enables the registration, identification and analysis of resources. The priority of implementing OA resource management is to establish a perfect registration institution for OA resources.

National Library should take responsibility and be authorized by the State to build the National Registration Management Center for OA resources, to establish a registration institution for OA resources and a registration system for open resources, and they will be used to achieve the registration and the management of OA resources so as to guarantee the smooth progress of OA resource management.

#### **4.2 Establish a Long-term Protection Mechanism for OA Resources**

As a native network resource that can be accessible publicly, OA resources constitute an important part of human civilization and cultural heritage; the inherent properties of their own decide that OA resources are greatly rich while they perish continually. Whether from the side of long-time and effective demand of users for resources or from the side of Chinese civilization's heritage, OA resources require a long-term and reliable protection mechanism. National Library as the general stacks is struggling.

#### **4.3 Establish a Service Mechanism for OA Resources**

The value of libraries lies in providing access services of information resources for readers and users. The main purpose of National Library's carrying out OA resource management and long-term preservation is to provide one-stop and fully integrated appreciation information services and knowledge services of OA resources while ensuring long-term and effective obtainment of resources.

National Library achieves the discovery services of OA resource through the primary retrieval of metadata, achieves personalized services by adopting the custom management system of digital libraries, develops value-added services through the use of metadata and object data and through the statistics and the analysis of user behavior, provides users with the one-stop full-text search and retrieval services of

OA resources by indexing, processing and integrating the collected metadata and object data.

## **5. The Prospects of COARP**

In China, there is no thorough but simple way to realize the integrate search to the OA resources, nor is there an effective way to organize them. That is why a national OA resources repository should be built to provide the public the integration search of the OA resources and other value-added services to improve the reveal and effect of the OA resources and to improve the amount of the OA resources. The National Library of China is building the COARP to realize the comprehensive collection, arrangement, processing, integration, preservation, retrieval, publishing, providing of the OA resources, to provide the users a one-stop integrated search and other value-added services of the OA resource, and to ensure the long-time availability of the OA resources.

## 5.1 General Framework

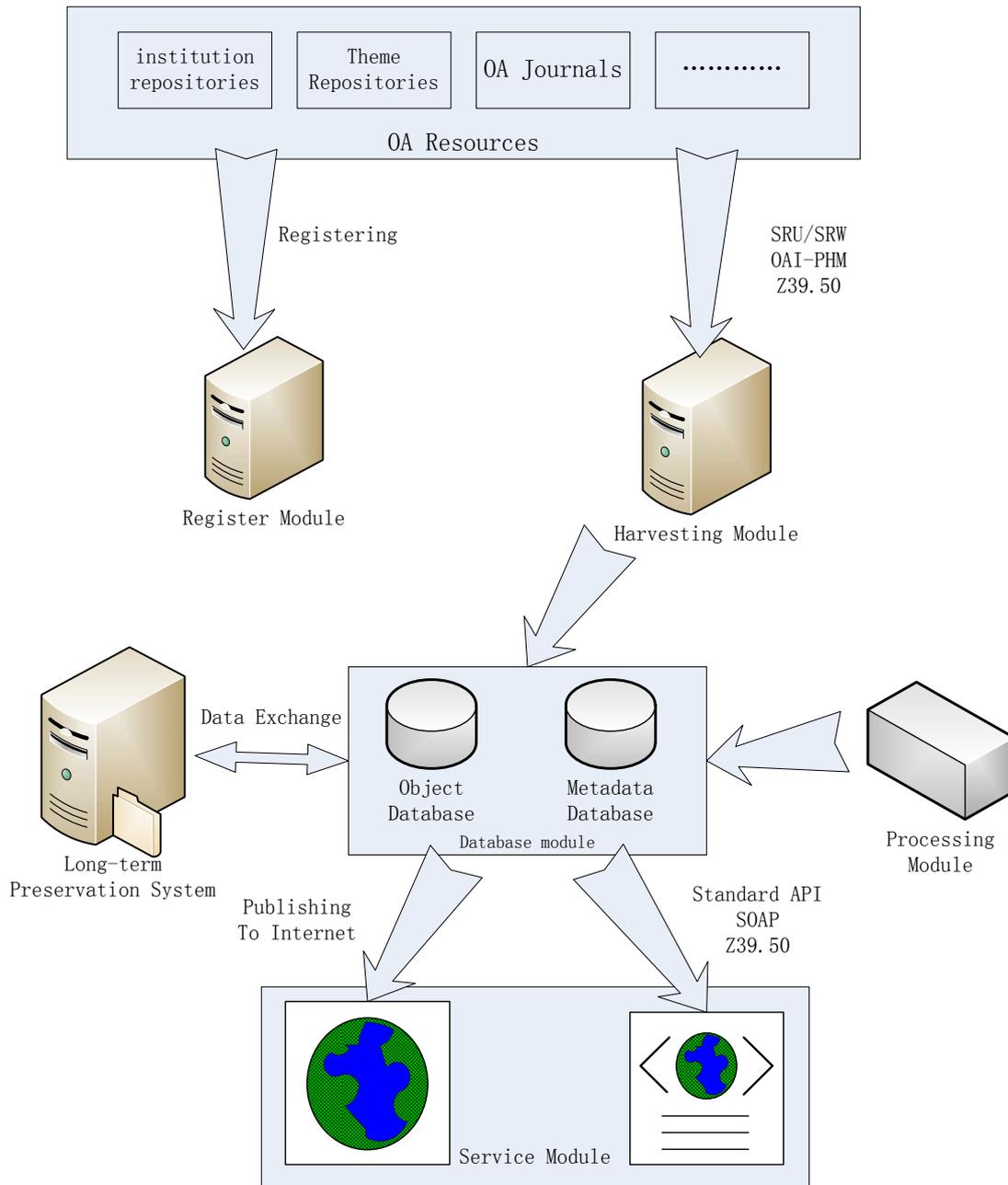


Chart 1: COARP General Framework

As shown above, COARP composes of 5 modules, including registering module, harvesting module, database module, processing module and servicing module.

OA sources can register in the registering module actively, or be added to the module manually by the administrators, adding the information such as the harvesting methods.

The harvesting module will harvest the OA resources by OAI-PMH or Z39.50 protocol periodically according to the information register module provides. The harvested information includes the digital objects and the metadata, which will be preserved separately into the object database and the metadata database.

The database module will pack the metadata and the digital objects according to the SIP regulations to put them into the long-term preservation system. And the module can also obtain the preserved resources from the system according to the DIP regulation.

The processing module can decide the criterion of the metadata, add and modify the metadata and put them into databases.

The service module can publish the preserved resources to the Internet according to certain requirements or rules, or provide standard APIs (SOAP, Z39.50) for other systems.

## **5.2 Registering of the OA resources**

As an existing form of the internet resources, OA resources should be registered in a specialized system. The properties, such as creation, existence, development and extinction, of the OA resources shall be recorded in the register system, so will the revealing of the metadata properties such as the technique, content, rights, management, service, and the realizing of the registering, identification and the analyzing of the OA resources, and the related services.

The OA registering system is mainly used to provide the register of the OA resource harvesting, the copyrights, and the long-term preservation of NLC. in the starting stage because of the policies and authorization problems, to control the information that which resource is harvested, which is not, or which is cataloged thoroughly, which is preserved already, and the copyrights. This system can be extended to be the entrance of COARP, supporting the importing, selecting, register of the OA resources to harvest, also the long-term preservation registering and management of the harvested OA resources.

The types of the OA resources include the OA journals, OA portals and others.

According to the content, the registering of the OA resources can be divided into harvesting register, cataloguing register, service register and long-term preservation register.

## **5.3 Resources Harvesting**

### **5.3.1 Harvesting Scope**

Usually OA journals and portals are mainly focused, actually the scope of OA resources are huge. So all the activities based on the thought “OPEN” can and should be included in our research, not only the general OA resources. The OA resources includes the serious web publications (such as institution libraries and OA journals), the free web academic information (such as personal web), the professional papers such as the research papers, the conference papers, the thesis, the monographs, the scientific data, the open source, and wiki, blogs, BBS, literatures, the audio and video data, which means all the resources that follows the principles of OA and available for users.

### **5.3.2 Harvesting Policies**

The harvesting activities follow the policies below:

**Legality:** the harvesting work will go according to the Press and Publication and intellectual property laws of our country, including but not limited in "Archives Act", "Intellectual Property Law" and "Press and Publications Ordinance", "electronic publication Ordinance", "Information Network Transmission Right Protection Ordinance". Legality is the protection to the harvesting planning, selecting, implements of the OA resources, and the prerequisite of the information services.

**Systematic:** as a whole, the OA resources libraries harvested should be a scientific knowledge system. It only can be formed after a long time accumulation. Harvesting in a mess cannot form a practical OA system. Systematic can be identified as the important insurance to the high quality information system.

**Standardization:** the harvesting work should follow certain criterions in the

library fields. The criteria includes but not limited in Z39.50, OAI-PMH, OAI-ORE. The standardization of the harvesting is the technical protection of the information services in libraries, and the prerequisite of the inter-operation between resources and the service systems in the OA resources discovery and service platform, and between the discovery and service platform and other resource service platforms in libraries.

Timeliness: information has its own life cycle, especially the “half life period” of the academic resources. The OA resources harvesting must ensure the timeliness to give full play to the academic value of OA resources, to promote scholarly communication and knowledge innovation, to improve the efficiency of turning the academic achievements into productive forces. Timeliness is the base of users in libraries enjoying the real-time notification.

### **5.3.3 Harvesting Methods**

The objects harvested in the COARP are the metadata and the digital resources.

Active harvesting is the main method for resource crawling, the publishers submitting is the secondary one. The harvesting system is divided into two parts. 1, publishers give active submission of the OA resources. This can be realized with the Legal deposit system of digital resources. The OA resources can be submitted to the deposit system, and then ingested into the harvesting system by some protocols.<sup>2</sup>, the harvesting system actively crawl and preserve the OA resources.

No matter which method is used, the timeliness is the core policy in OA resources harvesting, and the job shall go following all the related criteria.

## **5.4 Resource Integration**

In web environment, the Unified management system is often absent when the OA resources quantities, structures, distribution, outreach, content, control, transition are concerned. That is why the integration of the OA resources is important. In this project we harvest all the metadata and resources from the object sources locally, and then produce the metadata repository in a unified format by the scientific automatic transition and manual perfection to the metadata.

The integration of metadata is mainly realized by the automatic system. In the project we customize the metadata set of the OA library. This metadata set is a proper extension of DC. By the analyzing every website, the technicians make up the rules for automatic metadata extraction according to the metadata scheme of the project. And the rules can be continuous improved in practicing. So most of the metadata can be extracted automatically and the ones missed shall be manually made. The system supplies the full function of cataloging and audit to ensure the quality of the cataloging. Automatic classification and artificial audit are supported. The Basic elements of descriptive metadata of thesis are listed below:

<b>element</b>	<b>element refines</b>	<b>Comments</b>
<b>ID</b>		System ID , generated automatically by the system
<b>Title</b>		A name given to the paper. Typically, a Title will be a name by which the resource is formally known.
	<b>Alternative Title</b>	An alternative name for the paper. Index translated title, abbreviated title etc excluded formal title
<b>Creator</b>		An entity primarily responsible for making the content of the resource.
<b>Contributor</b>		An entity responsible for making contributions to the resource.
<b>subject</b>		The topic of the paper. Typically, the topic will be represented using keywords, key phrases, or classification codes.
<b>description</b>		An account of the resource.
	<b>Abstract</b>	A general description of the main content on paper
	<b>description relation Note</b>	Descriptions related to additional materials
	<b>Fund</b>	May includes project name and project code
<b>date</b>		
	<b>publication date/ Date of Publication/ Issued</b>	Formal publication date of paper
	<b>Harvesting Date</b>	Harvesting date of paper

<b>Resource Type</b>		Paper may include serials, proceedings and other thesis
<b>format</b>		
	<b>Format</b>	The file format, physical medium, or dimensions of the resource. e.g. PDF
	<b>Extent</b>	dimensions of the paper
	<b>page number</b>	Pages numbered
	<b>Technological environment</b>	E-book reading software. such as: Technological environment: Apabi reader
<b>Original Document Identifier</b>		An unambiguous reference to the resource within a given context.
<b>Host information</b>		Journal papers should include: journal title, ISSN, year, volume, page number (starting page, last page) and other sub-items of information Conference papers should include: the parent document title, parent document identifier, year, volume, page, session name, time, location and etc.
<b>source</b>		the web site information of resources on the network
<b>language</b>		A language of the intellectual content of the paper.
<b>full text</b>		Name of PDF file
<b>reference</b>		
<b>Location</b>		The location that storage the original object, express by URI
<b>related</b>		Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
<b>formally published</b>	the published status of the resources	
<b>peer review</b>	the peer review status of the resources	

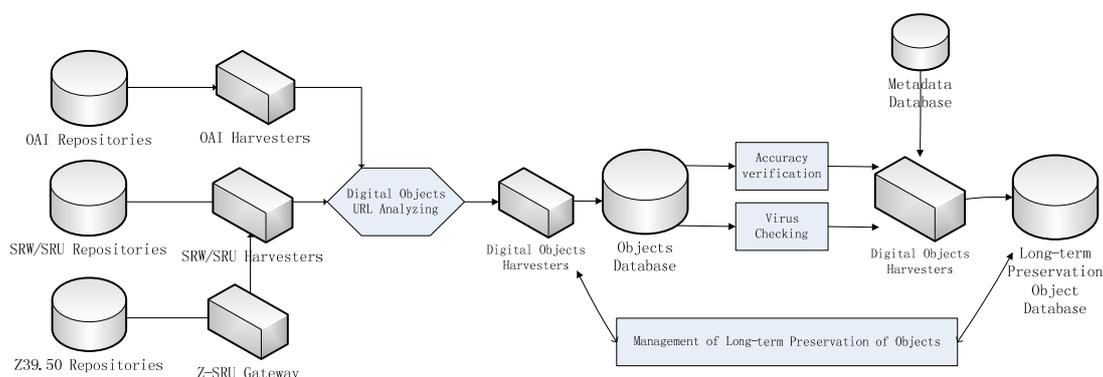
## 5.5 Long-term Preservation

In the long-term preservation of digital resources, the preservation abilities shall be considered to keep the original target of the producer-- Availability and reliability,

and the maintaining of the resources to keep the viability and the ability to be present and to be understood. The reliability means the byte files are complete to ensure the media is readable. The ability to be present means the byte files can be transformed to the file that can be understood by the human or machines. The ability to be understood means the present objects can be understood by the users. The aims of digital preservation are availability, capacity to be identified, comprehensibility, integrity, lasting viability, capacity to be present capacity and the authenticity of the digital objects.

Nowadays, the ISO standard OAIS are used as the technical structure of the digital preservation solution. The reference model stresses the whole procedures of the documents archiving, including processing, storage, data management, access and service. Also the transmission of the digital information to new media and new types, data model to express information, the role of the information preservation software and the digital exchange between archives are explained. Now OAIS is accepted widely. Many famous preservation system, such as the CEDARS The Distributed Digital Archiving Prototype, DSEP from NEDLIB, DIAS from e-Depot, CDS from Cornell, DAITSS from FCLA and the digital archive repository from CDL, are based on OAIS

We design the preservation flow according to the types and OAIS, which is shown below.



**Chart 2: preservation flow of OA resources**

Besides, National Library of China has done a lot of researches and experiments in the web information preservation. After the analysis to the results we find that we

need to preserve the web information in 2 ways. One is the original file format, such as HTML, word. Information can be preserved by URI to obtain the original certificate. The other is to transform the information into a unified format, such as TXT and PDF/A. The useless or duplicated information will be filtered. In this project, we use multiple ways to preserve the resources. First we preserve the original URI of the resources, by which we can call the original ones on the web. At the same time we transfer digital objects to PDF files and preserve. The reason is that PDF format is the ideal format for users to publish, preserve or spread the digital information in the world range, and is widely used in the digital documents publishing, long-term preservation, Digital Publishing and Printing and other fields. Besides PDF/A which is taken as format of the digital documents preservation has become the formal Chinese national standard after the approval of The Standardization Administration of China, And on September 1, 2009 comes into force

Policies, economy and management are involved in the long-preservation architecture of OA resources, so thorough analyze and coordination of organizations to the multiple elements involved in different levels and stages are required. The framework of the OA resources preservation architecture is brought up, in which all the elements that may be concerned are summarized. The related practices and researches are needed when the realization of the specific stages are concerned because of the initial stage of the long-preservation in our country.

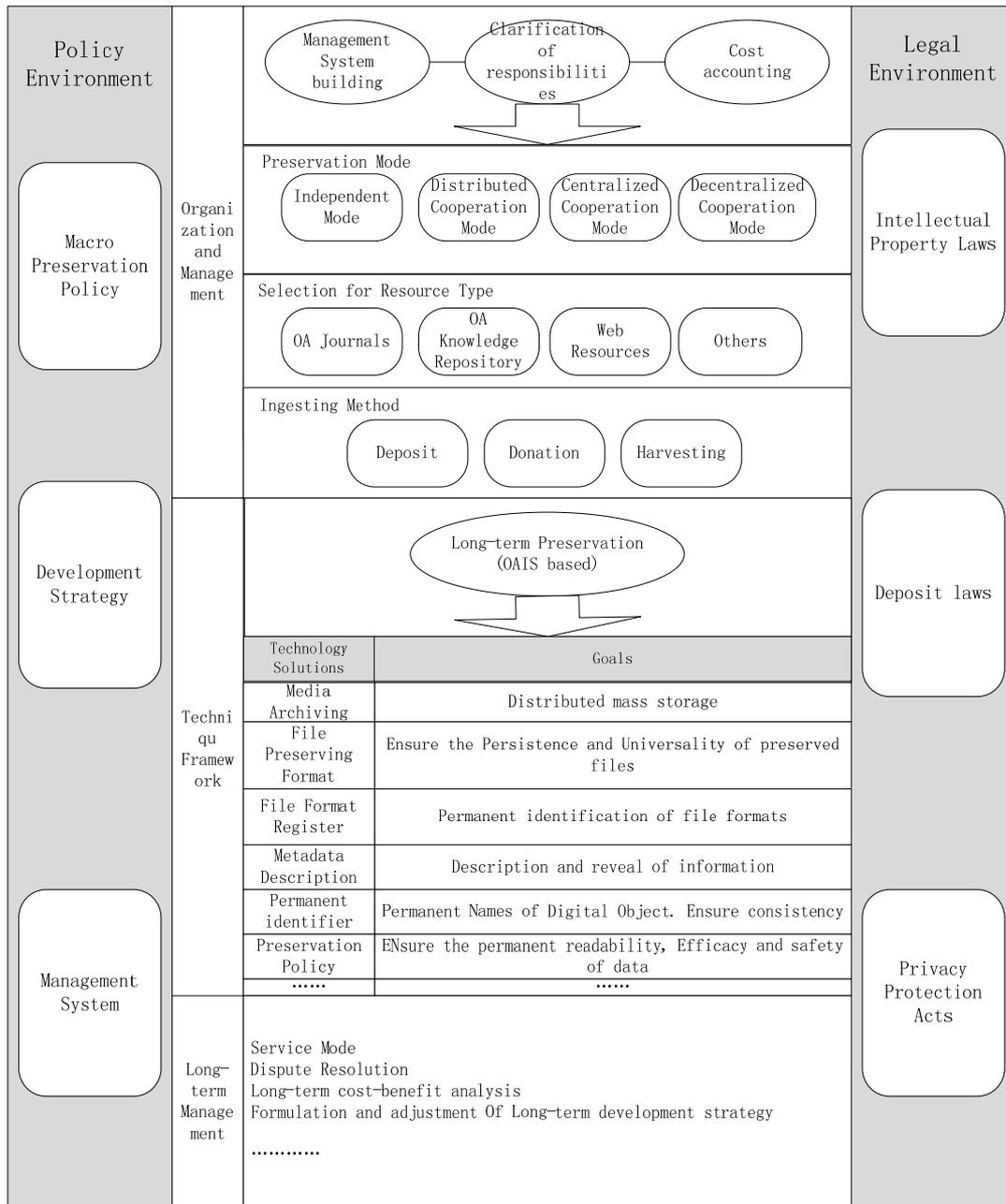


Chart 3: Long-preservation of OA framework

### 5.6 Resource Service

To provide comprehensive value-added information and knowledge services is the essential goal of the management and preservation of OA resources. OA services cannot exist without other services. That is why the libraries shall provide a whole information service environment. The lists of services we provide in this project are below:

Resource discovery Services: This is primarily realized by the retrieval of metadata. In this project we provide the retrieval service of the metadata of OA resources we collected. Users can find needed resources by simple or advanced search of the multiple metadata fields.

Full-text Acquisition Services: in this project, we do long-preservation to the full texts of the resources; besides we preserved the original obtain addresses by which users can access the original resources. We can provide the full text even if the resources on the internet disappear.

Collection Integration Services: we integrate the OA resources in the metalib system and provide services together with other resources. So users can search for all the collections and OA resources. That is easy for users and improves the utilization of the resources.

Value-added Service: the core of the OA resource management is the control and the added value. The value-added service means the functions which are developed based on the statistics and analysis of the utilization of metadata and digital object and the user behaviors. Theme service, citation analyzing service and utilization analyzing service are included.

Personalized service: in the future CRM of digital libraries will be used. It can provide personalized service to users according to the specialties, hobbies and other preset properties of the registered users, including the personalized UI, spaces, personal digital libraries, personal research communities and so on. The information-push services, such as the remind of new articles, new comments, citation tracking and retrievals, can be provided by email, SMS, RSS and so on.

## **6. Conclusion**

As a method to access information fair and free in modern society, today, OA not only has greatly changed peoples behaviour of learning and study, but greatly changed the original management and services mode of the library after a short stage of development, the library should not only consider how to resolve all kinds of the new

problems based on its own mission and profession, but to consider using their own professional advantage to study the related issues, to guide the OA movement development in health, science and sustainable manner more.

NLC is working to become the promoter and practitioners of open access movement in China. The COARP project aims to build the security mechanism of open access resources management and long-term preservation by the establishment of a standardized, sustainable and scalable technical and management framework. It will realize the systematic collection, organizing, preserving of Chinese open access resources. NLC is playing an important role in the harvesting, integration, preservation and service delivery for OA resources.

#### **[Reference]**

[1] Liu hongzhu, Qiu quan. Quantitative Analysis of Articles on OA in China during 2004 and 2007. [J]. New Century Library .2009/01

[2] CHU Jing-li. Recent Development of OA abroad and in China. Library Tribune. [J].2009/06

[3] Qin Ke. The pressure of OA on libraries and the appropriate countermeasures .[J]. Researches in Library Science.2006/8

[4] [EB/OL].[2010-05-12].<http://www.paper.edu.cn/>

[5] FENG Xiao wei. Influences of the Long-term Preservation of the OA (OA) Resources on the Development of Library. [J]. Sci-Tech Information Development & Economy.2008/25

[6] Digital Preservation and Asset Management.[2008-12-08].  
<http://www.jisc.ac.uk/whatwedo/programmes/preservation/assetmanagement.aspx>

[7]Budapest OA Initiative. [2008-12-08].

<http://www.soros.org/openaccess/read.shtml>