

DIGITIZING AND NETWORKING: THE SCHEME AND PRACTICE OF ESTABLISHING THE CASS ELECTRONIC LIBRARY*

By Lu Benfu, Zhou Junlan

Abstract: Digitization and networking are two crucial steps in a transition from traditional to electronic library. This paper presents the tentative ideas and practices of the library of the Chinese Academy of Social Sciences (CASS) in this line. As to the digitization, it first analyses the CASS information resources, focusing on the characteristics of disciplines, types of media and distribution of language varieties, and then presents an overall plan on digitization and the concrete steps to be implemented as well as the possible technical problems to be met in the course of digitization. Concerning the networking, this paper first identifies the model of information exchange of CASS, based on which it designs the overall objectives and schemes for establishing the CASS electronic network and also draws a logical structural diagram of the electronic network.

China Academy of Social Sciences (CASS) is located in Beijing. It is granted by the Chinese government. Its researches are mainly conducted on the national social sciences schedule in recent years. Some enterprises, organizations and individuals entrust it with research work. At present, some 5,000 faculties work in over 32 professional research institutes which including economics, law, sociology, trade, religion, philosophy, literature, nationality problem and international relation, etc.

Compared with natural science, the researches of social science are more dependent upon information and documents, so the information and intellectual density is high in CASS, and the home and abroad communication is frequent. Accordingly, the documents on social sciences are in a rich collection, even documents of some disciplines hold a lead in China. The research results are information products too, e.g. monograph, thesis, research report. CASS library is responsible for information receiving and literature collection, and information outputs are processed by publishers and distributors.

CASS decides to develop the traditional library into an electronic library, that is,

* Paper delivered at the Open Session of the section Social Science Libraries at the 62nd IFLA General Conference in Beijing, 27 August 1996.

information communication is carried out by electronic network. Some schedules were formulated, part of them have already been implemented. This plan embraces two aspects:

- Digitizing: transforming library's documents and information to computer readable form.
- Networking: receiving and delivering digital information through electronic network.

The following is an overview of the experiences made in implementing digitizing and networking in the CASS.

1. Characteristics of information resources.

As a research organization on national level, CASS processes relatively rich information resources, holding over 5,000 thousand volumes of books and 2,000 journals. To digitize such a gigantic library collection, its characteristics must be understood:

(1) Discipline distribution of information resources

From the discipline point of view, CASS' information resources basically cover all disciplines of philosophy, social science and human science, which is shown in the following table:

Discipline	Quantity (10000 vol.)	Percentage
Total	506	100
History	149	29
Literature & Linguistics	97	19
Economics	77	15
Sociology & Ethnology	48	9.5
Foreign Research	43	8.5
Politics & Law	41	8
Philosophy & Religion	36	7
Others	20	4

Resource: The Document and Information Center of CASS (1992)

(2) Media types of information resources.

On the whole, traditional text is the main medium type of CASS information resources, however, information collections on other types are tending to expand, like video-audio documents, microfiches, CD-ROMs, machine-readable databases, etc. Other specific types include maps, pictures, famous drawing, etc.

The composition of media types is demonstrated by the following table:

Medium Type	Amount (unit: 10,000 vol.)	Proportion
Total	506	100
Ancient Books	62.5	12.27
Books & Periodicals	425	73.6
Non-book Materials	22	9.81
Others		4.31

Source: The Document and Information Center of CASS(1992)

(3) Language distribution of information resources.

Because of the wide range of CASS's research work, there are many different languages used. Chinese and ancient Chinese has undoubtedly the largest proportion, in addition to it, there are more than ten kinds of languages of Chinese minorities, such as the Zang language, the Meng language, the Thai language etc. Concerning non-Chinese languages besides the large ones such as English, Japanese, Russian, German etc. there are Korean, Serb, Spanish, Italian, Brahma, Swedish - in total 10 different languages to be found.

2. Digitizing of information resources

According to the above statistical analyses, we can find that: wide academic ranging, variety of language, complicated medium are the characteristics of the information resources in CASS. Digitizing these information resources is undoubtedly a hard work, so it should be done in order of importance and urgency, in light of different media characteristics and information contents.

In practice, we have made the following rules of procedure: at first, we should think over our economics supporting ability. At present we cannot make more than one investment in a time, so the work must been done period by period. The second step is resource sharing. If external institutions already have relevant

digitized information products, then we do not need do the same work. The third is the „ripe technology rule“. We only can with one item for which the digitizing technologies are ripe, otherwise we can shall leave it for later.

(1) The overall plan of digitizing

Our final target is digitizing all kinds of information resources: texts, audio, videos, maps, pictures, tables and statistical data, to be able to finally build the electronical library, but at the moment the CASS library maintains unclear schedules. The overall plan is divided into three periods: The short-term work is developing and building the specialized databases which CASS and other departments are in great need of; the middle term work is digitizing the information resources which are only owned by our library, because the work can just be done by us; the last period work is purchasing directly digitized information products in China as well as abroad which will digitize the whole library holdings. The order of digitization work is:

- a) The information which our research mostly needs, such as the CASS library collection catalogue, library's periodicals holdings.
- b) The information which is used in international communications. The reform in China has been taken for more than ten years, the economy has developed at high speed, and great successes have been achieved, at the same time, works on economics, law and so on in these periods can be surely used for reference for the developing countries.
- c) The China ancient books and national minority's documents in CASS.

(2) Specific plan of digitizing

On the base of above rules, a whole set of specific plans has been made by our academy, and has been taken into practice. These are:

- a) Building machine readable catalogue databases of all the books hold in CASS. This is the focal point of our digitizing work, which will convert all catalogues into standard machine-readable MARC form. This concerns about 500 million volumes (not including ancient books). At present there are 20 thousand MARC data, and we are planning to finish converting the whole data of our academy in five years.
- b) Building machine-readable catalogue databases of the complete periodical's holdings in CASS. This is similar to the first one, the difference is that the

databases include not only periodicals catalogues, but also subject information on the periodicals contents. In order to save library storage space, all the obsolete periodicals will be made to microfilms, this will also be done in five years.

c) Building databases of specialized documents. This is mainly used for international communication. At present, our academy has built specialized economic document databases. The databases has collected the papers which were published in the chief economical periodicals in China in the last five years. The contents include resources, subjects, abstracts etc.. Another plan is to build a full-text database on law which was issued in recent years. The database is mainly about the law concerning foreign affairs or foreign nationalities, it therefore can serve for foreign businessmen, and foreign scholars.

d) Building multimedia databases of China ancient books. The contents include: the rare books in our academy, the original handwritings of historical famous persons, archaeological pictures and so on. At present the database of rare text books has been started, other multimedia databases such as original handwritings are being planned.

e) Building factual databases. The research in social science has the tendency of developing from qualitative analysis to quantitative analysis. Therefore, we are preparing to digitize the factual tool books in our academy, in order to meet varying needs.

(3) The relative technology problem of digitizing

In line with our practice, we are facing the following difficulties when we digitize information resources:

a) The problems which the MARC are already known. Machine-readable catalogues all come from LC-MARC and include CN-MARC made in China. Since the formatting is based on the computer technology of that time, now to some extent the MARC form is backward , the typical show is: cumbersome field-setting, duplicated contents. Although in theory different kinds of MARC records can be interchanged, it is difficult in practice.

b) The difficulty for choosing multi-language software platform. There hasn't been a good multi-language software platform which is compatible with Chinese, western languages, Japanese , Russian and so on. So it is easy to make mistakes. Though the international standard on large character groups has been officially issued, the relevant operating system has not been set out, so that there still

remains a problem to be solved. On the other hand, the Chinese minority languages have no relevant computer processing software.

c) The difficulty in building multimedia databases. At present the multimedia technology can not settle the digitizing problem of Chinese calligraphy and painting. Even if it had been settled, the retrieval would be difficult.

3. Information communication models

Information communication models state two ways of communication. Indirect communication is meant when information is spread in the form of documents. Direct information exchange means that research fellows obtain information directly in ways of engaging in advanced studies, attending academic discussions, and so on, and information spread by lectures, academic discussions, etc.

(1) The size and distribution of document flow

According to the latest statistics in 1994, more than 20,000 Chinese books, 4,000 books in foreign languages, 4,000 Chinese journals and 2,000 journals in foreign languages are being added to the collection of the Academy every year. The information of all these documents can occupy 500M computer storage space. There are two forms of documents input: one is purchasing, this is the main way to increase the information resource; the other is exchange or donation. The Academy keeps exchange relations with about 400 book- and document-agencies abroad, and exchanges about 1,000 books and journals every year, the total number of volumes is more than 10,000.

Abundant information resources present plentiful reference materials for the research fellows, create a good environment, and so the research findings of the Academy are rich. The research findings are mainly published in the following forms: treatises, academic dissertations, research reports, learned materials, translations, translation texts, reference books and software, etc. Accounting in word, the proportions of the findings in every forms can be found in the following table:

Type of Research Findings	Number of Word (10,000)	Percent Distribution
Total	55071.1	100
Learned Material	9464.5	26.79
Treatise	8497	24.05
Reference book	7351	20.08
Academic Dissertation	3708	10.21
Teaching Material and Popular Reading	2999	8.49
Translation	1211.6	3.43
Research Report	674	1.91
General Article	376	1.06

Source: Statistical year report of CASS(1992)

The information listed above can occupy 200M computer storage space.

(2) The forms of direct information communication

In CASS, direct information exchange still is active, and the forms are varied, mainly including scholars' reciprocal visiting, holding workshops, engaging in advanced studies, giving lectures, and so on. Much of lately information can be obtained by setting up relations with relative institutes home and abroad, and scholars' researching, discussing academic problems with general character.

The following table reflects the state of the scholars' international interchange:

<i>Forms</i>	<i>Number of Go Abroad</i>		<i>Number of Visitor</i>	
	1991	1992	1991	1992
Total	578	710	978	783
Academic Visit	235	241	365	297
Co-research	73	63	34	52
Lecturing	12	11	21	15
Advanced Studies	70	70	35	7
Bilateral Workshops	31	63	53	187
International Meeting	107	140	352	93
Working Visit	29	24	91	101
Other	21	98	20	31

Source: Statistical year report of CASS(1992)

The direct information exchange is effective, but it is greatly limited by many objective conditions.

(3) Table of academic network

CASS mainly keeps academic relations with the following institutes:

Inter-national	Research Institutes of UN	Embassy	Societies & Associations	The Press Offices	
National	Government Offices	Ministries & Commissions	Societies & Associations	Universities	The Press Editorial Offices
Local	Local Government Offices	Local Academies of Social Sciences	Local Universities		

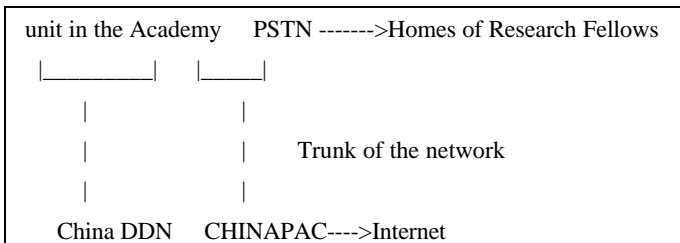
4. The objective and the plan of setting up the electronic network

From the information exchange models we have discussed above, we know that both information input and output are frequent in CASS. The electronic network is just the imitation of these models to satisfy the needs of information input and output.

(1) Objective & structural drawing

The objective of the electronic networks: transparent information sharing in CASS; domestic information acquisition and retrieval through electronic network; direct connection with international or domestic scholars of the same occupation; international; and domestic users' direct using of CASS's information resources and information products. In practical working, we have drawn up a plan for setting up the electronic network in stages.

Logical structure drawing:



(2) Plan & moves

First, setting up the CASS Scientific research information management network. There are three parts of business on the network:

a) Scientific research business

This is mainly for managing research projects, scientific research programs and research findings, the most important of all is the research findings management system, which will help gathering the findings of every year step by step;

b) Information business

Here is intended to store all the information resources (books, periodicals) in the CASS library and in every unit in computer network and to present retrieval service to internal users at any time.

c) Management business

It mainly indicates the office automation for storage of all kinds of administration information in the network in order to improve the management level of all the Academy.

Network technology design: according to information traffic, geographical information distribution and the frequency of use, the CASS DDN uses circle arboraceous topological structure. The trunk of the network uses optical fibers, and the information transmission ratio is 100M bps.

Working state: The CASS DDN has been carried out for one year, the computer network of the Academy library is established, office automation DDN is also completed, the trunk of the network is estimated to get through in 1996.

Second, extending the network into the research fellows' homes . Now many research fellows of CASS begin to write with another kind of pen-computer. It's estimated that the percentage of the families having computers can reach 20%--30%. Otherwise, most of families have installed telephone, so that the terminal users have the basic conditions to connect with the network. Using the public telephone network and modem, we plan to connect most families' computers with network in the next 3 or 5 years, and achieve three goals: terminal users' retrieving information of library at any time; direct connection with international and domestic scholars of the same occupation; adding their research findings to the network. Now, the library network has had only the ability to take in a few terminal users.

Third, to connect the network with the global Internet. At the end of 1994, Chinese Ministry of Posts & Telecommunications started CHINADDN & CHINAPAC, and reached an agreement with Sprint Comp. of America that CHINAPAC & CHINADDN can directly connect with the global Internet by using the route offered by the company. Within the next few years, we also plan to apply for a special DDN route from Beijing Telegraph Office - the institute managing the Internet in Beijing - and to start our own Scientific Research Information Network, to connect it with Internet, and at last to realize information sharing.

Lu Benfu and Zhou Junlan
Chinese Academy of Social Sciences
Beijing, 100732
China