

Automation of Processes in the National Library of China: Historical Review and Future Perspective

BEN GU



Ben Gu has an MS in Mathematics from Fudan University, Shanghai, and a PhD in Management Science from Renmin University of China, Beijing. He began work in the National Library of China as a book selector in 1987, and became Chief of Book Selection in 1989. He was appointed Director of the Centre for Acquisitions & Cataloguing of Foreign Books in 1998 and Deputy Director of the Acquisitions & Cataloguing Department in 1999. He has more than 100 publications to his name, covering library science, the book industry, music and Chinese studies.

INTRODUCTION

The National Library of China (NLC) was established in 1909. As of December 2001, it had a collection of more than 21 million items, including 5.5 million volumes of Chinese monographs, 3 million volumes of monographs in foreign languages, 45,230 titles of Chinese periodicals, 41,436 titles of periodicals in foreign languages, and various special collections. About 1,400 full-time employees and several hundreds of part-time employees work in the library.

In the last two decades, the automation of processes in the National Library of China has progressed smoothly up to a certain point. However, the processing of publications in foreign languages is still to be computerized. This situation does not match the present state of library automation in countries from which these publications are imported. There are several reasons for this. As the country's national library, the NLC must give top priority to the processing of

Chinese publications; the prices of library systems for the processing of publications in foreign languages are comparatively high, and we can purchase an efficient system only when the budget is sufficient; and limited technologies have not previously allowed a library system to process all languages on a single platform. In this paper, I review the history of automation in the NLC and put forward some proposals for future development.

THE 1980s

In the 1980s, the NLC used LC MARC tapes to print bibliographic records of books in specific subjects, such as on China and Marxism, for acquisitions purposes. This was the earliest form of library automation; it simply provided more information for book selectors and did not change the manual work processes. During this period, the acquisitions librarians corresponded with

foreign publishers or booksellers by mail, at best by telex.

In 1987, when the present library building was opened to the public, the NLC formally began its library automation programme. It specified keyboards in Western, Japanese and Russian languages, and drafted system requirements. Because there were no software systems and software developers, the NEC computers were not used for the processing of publications in foreign languages. At the same time, staff tried to make more use of them. They used Japanese and Russian terminals for word processing and compiled some bibliographies. Because terminals did not have hard discs, data could be stored only on 360KB or 1.2MB floppies, which cannot process large amounts of information or accommodate databases. The bibliographies compiled at that time could not be converted into IBM-compatible formats, and they cannot be used now. It might be said that computers were used as typewriters.

In 1990, acquisitions librarians of Western monographs used dBase II to create an ISBN database on an NEC Chinese terminal with a 20MB hard disc. Since then, book selectors have been using it for duplicate-checking when they select new titles. The database was later updated and switched to IBM-compatible computers. Although quite a simple system, it has saved much manual work.

THE 1990s

Acquisitions

One of the main characteristics of the 1990s is the large-scale application of CD-ROMs. The first CD-ROM title we used was BIP (Books in Print with Reviews Plus). It has replaced the three volumes of *Books in Print*, four volumes of the *Subject Guide to Books in Print* and the biannual *Forthcoming Books*, saved much shelf space, improved search efficiency and timeliness. During this period of time, fax was becoming popular, and e-mail was becoming a main method of communication with foreign booksellers.

Cataloguing

Western languages

Automation of cataloguing began in the early 1990s with the introduction of Bibliofile

CD-ROM produced by TLC (The Library Corporation). Limited by the network environment and the computer expertise of staff, all MARC records downloaded were stored on floppies. Because of poor physical conditions and disc quality, many of the several hundreds of floppy diskettes were found unusable when bibliographic records were converted into ISO 2709 format in recent years. This has caused a waste of human resources in that some downloading operations had to be done again.

With consideration of prices and search interfaces, we began to replace Bibliofile with OCLC CatCD for the cataloguing of books in Western languages in 1998. The latter can be run on a Windows NT based network and shared by many users.

Chinese books

During this period, Chinese monograph cataloguers began to use the ACOS system based on NEC mainframe computers to catalogue new Chinese monographs and produce the *China National Bibliography*. After a few years, ACOS was replaced by a Windows-based system produced by Wenjin IT Centre, which is located in the NLC building and consists of former NLC computer technicians.

Japanese books

In the Section for Monographs in Oriental Languages, a Novell network was installed for the access to bibliographical CD-ROMs produced by the National Diet Library and Nippan. At first, staff used floppies to store downloaded MARC records for future conversion. Because the library did not plan to introduce a special system for processing Japanese books, the section stopped using the floppies in 1998. The CD-ROMs are now used solely for printing card catalogues of books searched in them.

Russian books

In 1998, some Chinese library software vendors were interested in testing the processing of Russian in their Chinese software. Theoretically, Russian can be processed only if necessary changes are made to the interface and a Russian operating system is used. Because of the limited market in China, the vendors did not pursue this

application. Even today very few Chinese libraries can process Chinese, Russian and Japanese in a single system.

International exchange

In early 1990s, the International Exchange Section used dBase II (later FoxBase 2.0) to make a simple database for the management of the addresses of exchange partners. The work in this section was thus automated to a certain degree and became more efficient.

Reader Services

OPAC

In 1999, Wenjin IT Centre produced a Web-based OPAC system for the use of public access and online reservation. As the first open access catalogue system in the library, it can be used to search bibliographic records of Chinese publications and publications in Western languages, although Latin characters with diacritics appear as strange Chinese characters. However, the system has several disadvantages:

- it is based on Windows Chinese version, and does not support Western languages, especially characters with diacritics, not to say other languages, such as Russian and Japanese;
- it does not have an English-language interface and is not easy to use for foreign users;
- it does not have many search options.

Closed-stack service

In late 1990s, Tsinghua University helped the library to make a system specially for the management of reader status and the delivery of call-slip information. This has replaced the former mechanical call-slip delivery system.

Open-stack service

A circulation system purchased from an American company in the late 1980s went out of service and was finally replaced by a new system produced by Wenjin IT Centre. With this system, the library can use smart cards jointly issued by the NLC and the Industrial and Commercial Bank of China to process user information (including ID numbers and colour photographs) and circulation records.

Infrastructure

In 1999, the NLC completed the construction of a gigabyte library-wide network, with every information node sharing 10 MB. It was the first Gigabit Ethernet library-wide network ever built in China. In 2000, the library implemented second-stage library-wide network engineering, upgrading network equipment, adding information nodes and computers and developing a VOD system.

The NLC began to be connected to CERNet (China Education and Research Network) based in Tsinghua University via microwave in 1995, and has since been connected to Peking University and the Chinese Academy of Sciences. Its connection with ChinaNet began in 1996, and with BCTV in 1997; it established a 100-MB connection with the State Council and a 1-GB connection with China CATV Net in 1999. The NLC now has Internet connections with almost all backbone networks in China.

Summary

We can summarize the following characteristics of library automation during the 1990s:

- Improvisation: automation in the acquisitions and cataloguing departments was not performed by computer professionals, and librarians had to spend quite a lot of time learning to write computer programs.
- Independence: various processes used different systems and were not linked, and data could not be shared.
- Incompleteness: only acquisitions and cataloguing of books in some languages were automated to a certain degree, leaving all other languages with a manual or semi-manual operation.
- A good network without an integrated system: a gigabyte library-wide network was constructed, but there was no integrated library management system for all the processes.
- No generally used standards: different sections, such as those of Chinese new books, Chinese rare books, audiovisual materials, etc. used different cataloguing formats for different systems, or different rules for the same formats.

At this time, library automation in the National Library of China lagged far behind not only national libraries of developed countries, but also other major libraries in China, such as Shanghai Library, Peking University Library and Tsinghua University Library. The reasons include the inadequacy of budgets, the large size of the collections, the pursuit of processing of all languages in a single system, and some management problems.

PRESENT POSITION

Present processes for foreign publications

Acquisitions

1. New title information is obtained from publishers, booksellers, CD-ROMs, the Internet, etc.
2. Book selectors make their selection on bookseller order forms or print interim order slips for acquisitions librarians.
3. Book selectors check the ISBN database and the author/title card catalogues for duplicates.
4. Acquisitions librarians prepare orders on interim order slips prepared by book selectors or order forms produced by booksellers.
5. Acquisitions librarians check library holdings for duplicates again and maintain author and title card catalogues.
6. Orders are sent to booksellers via e-mail and printed order forms.
7. Acquisitions librarians use typewriters to put short titles line-by-line on individual check-in forms.
8. Acquisitions librarians calculate total amounts of invoices for payment.
9. When a title has been processed by cataloguers, acquisitions librarians have to replace the card order form with a catalogue card.

Almost all the processes are manual. Acquisitions staff have to manually type bibliographic records two or three times. If the manual work of cataloguing is taken into account, there is even more duplication of work.

Cataloguing

Russian librarians use typewriters to type catalogue cards; librarians with knowledge of Korean, Arabic and Hindi librarians use PCs to print

catalogue cards; and librarians who know Japanese use CD-ROMs to print catalogue cards.

As for cataloguers of monographs in Western languages, they classify books according to the *Chinese Library Classification*, manually check staff catalogues to assign author numbers, use OCLC CatCD to download bibliographic records, and manually print spine labels. They then convert OCLC files into ISO 2709 format and forward them to computer technicians, which have to convert these records again for uploading to the OPAC system. Meanwhile, reader service librarians have to check the OPAC record by record when they receive newly catalogued books from cataloguing sections. The OPAC search has no authority control, but the cataloguing section maintains some reference cards for authority control according to LC authority file microfiches. Although the library has an OPAC for searching books in Western languages, catalogue cards are still used for all foreign languages, and cataloguing staff have to organize all the card catalogues in addition to processing MARC records. If any changes in bibliographic records are needed, all the cumbersome procedures have to be repeated.

Reader services

Systems for reader services are separated from the OPAC system, and we are unable to link bibliographic records with user data. For example, when users search the OPAC, they cannot view the status of books and so do not know if an item is available or not. Likewise, the check-out desk staff simply record the call number and user ID, and there is no means of knowing what titles a user has borrowed.

Looking for a good integrated library system

System selection

To solve the problem of not having a library system for the processing of books in foreign languages, the NLC began to select an integrated system for this part of work in late 1999. Among the principal requirements are that it must be technically up-to-date, able to do multilingual processing, extendible, capable of integration and consistent with the present processes.

Because there was no Chinese local library software vendor able to provide such a system, the system selection group visited major Chinese libraries using systems developed by foreign vendors and held discussions with representatives of major library software vendors in the world. After two years of consideration, the NLC decided to select Aleph 500 produced by Ex Libris, an Israel-based company, and signed the contract in September 2001. We find that Aleph 500 can meet most of our requirements; it supports Unicode, and is flexible, extendible and cost-effective. We also find that the Chinese and the Israelis are similar in their way of thinking, because they are both people with long traditions and with customs that differ from those of westerners.

It has been decided that the system will be used not only for publications in foreign languages but for Chinese publications.

Preparatory work

In November 2001, the NLC and Ex Libris drafted a timetable for the implementation of the new system. It is expected that the preparatory work will take about a year, and the STP (Switch to Production) will be completed by the end of 2002.

In January 2002, a working group was established to coordinate preparatory work on the new Aleph system. It consists of managers, librarians and computer professionals from different departments, who are given responsibility for organizing training, setting up parameters and representing NLC in contacts with Ex Libris.

Since Chinese publications have been catalogued in CNMARC format, which is based on UNIMARC, and publications in Western languages have been catalogued in MARC21 (USMARC), we have to create two bibliographic databases respectively in CNMARC and MARC21 formats. For simplicity, we plan to use MARC21 for all other foreign languages, including Russian, Japanese, Korean, Arabic, Vietnamese, Mongolian and Hindi, and are trying to find sources of bibliographic records for these languages.

FUTURE PERSPECTIVES: THE 21ST CENTURY

New processes

Acquisitions

1. New title information is acquired from publishers, booksellers, CD-ROMs, the Internet, etc., in MARC format, and then downloaded into temporary bibliographic databases; if the interfaces between Aleph 500 and booksellers are completed, direct downloading from bookseller databases can be used.
2. Book selectors make selection on booksellers' MARC data and check duplicates directly by searching data in related fields (author, title and ISBN).
3. Acquisitions librarians add order and vendor information.
4. Orders can be sent to booksellers via e-mail and EDI.
5. Check-in can be done by adding necessary information in bibliographic and administrative records.
6. Total amounts can be generated automatically for payment
7. Line items for general invoices can be generated from bibliographic records, and manual input is no longer necessary.

In addition, book selectors can utilize user statistics to analyse use rates of library collections so that necessary adjustments can be made to meet users' needs.

International exchange

Exchange librarians can use the acquisitions and serials modules by treating their exchange partners as if they were booksellers.

Cataloguing

Cataloguers can do original cataloguing, change bibliographic records created by acquisitions librarians, or use MARC records downloaded from OCLC CatCD and OCLC WorldCat or other services. Some booksellers can provide MARC records when they supply books. All the other work can be done in the system with little or no manual work. Card catalogues will be obsolete.

For Japanese monographs, we plan to use the Z39.50 service provided by the National Institute

of Informatics in Japan to download MARC21 records for cataloguing.

For Chinese monographs, we can use the authority files prepared by the NLC, which have never been used before. We have purchased LC authority files for the authority control of records in Western languages. For other languages, we can create authority records or purchase them from relevant countries if they are available.

Because the NLC does not have librarians for all languages, publications in some languages cannot be processed. For example, the Section of Oriental Books is responsible for Vietnamese books, but it has had no Vietnamese specialists for some time. With the application of the new system, Vietnamese books and other languages in Latin character sets can be processed by Western language librarians by downloading records from OCLC WorldCat. This is cost-effective for languages in which the library acquires of fewer than 1,000 new monographs annually.

Reader services

With the application of the new system, reader service librarians can manage all items and users in a single system; reference librarians can use the system to provide an SDI service; and readers can use the Chinese and English interfaces and the powerful search functions of the OPAC to find information more easily.

Problems and solutions

Hardware

In principle, most computers in use at present should be upgraded or replaced by new ones to accommodate Windows 2000 as required by Aleph 500. If we can find a way of using the present computers with Windows 98 in related languages, we can save much money and time.

Bibliographic records for cataloguing and acquisitions

In the new system, bibliographic records are shared in all the processes, including acquisitions, cataloguing and reader services. However, bibliographic information from different sources, such as booksellers and cataloguing tools, are different. For example, a multi-volume title may be described by acquisitions librarians as such but

by cataloguers as a series title consisting of various separate titles. Ways need to be found to deal with this sort of problem.

Authority control

Although Chinese authority databases have been maintained by the NLC for several years, they have never been used for authority control. The staff have to find means to establish a connection between authority and bibliographic databases.

After purchasing the authority files from LC, we should make necessary changes in the authority records, especially for those of Chinese names, and add Chinese characters so that users can find titles in all languages by searching a single name. The question is whether we should maintain one authority file, or one for our own use and another for the original LC data.

MARC formats

Theoretically, Aleph 500 can accommodate many MARC formats. If we wish, we can create many bibliographic databases with different MARC formats for different languages, e.g. CNMARC, MARC21, RUSMARC and JapanMARC. However, this would lead to difficulty in system management and low efficiency in OPAC searching. After investigations and discussions, we have decided to use only two MARC formats, CNMARC and MARC21.

Payment

Many new integrated library management systems have e-business functions to allow users to use electronic invoices and electronic payment. However, almost no Chinese booksellers provide such services. We should try to persuade them to add such functions and facilitate our acquisitions work.

Reader services

Aleph 500 does not fully meet all the requirements of the library, especially those of reader services, such as reading room statistics, the processing of readers' cards, and the display of photographs for control purposes. (The smart cards issued jointly by the NLC and the Industrial and Commercial Bank of China do not contain photographs. We therefore require the system to display photographs of card-holders when the use

the library. Ex Libris promises to develop this function, but has not yet done so). To solve these problems, we need to persuade the vendors to add new functions to meet our requirements on the one hand, and on the other we should adjust our own processes to the system.

CONCLUSION

'Therefore its name was called Babel, because there the LORD confused the language of all the earth; and from there the LORD scattered them

abroad over the face of all the earth.' (*Genesis XI*). It is impossible for us to unify all the languages in the world. However, it is our dream to process all of them, particularly in a single system. What we have been doing is to work towards realization of the dream. With the rapid Chinese economic development and the greater attention of the government, we think that we can do so in the near future. A brand new National Library of China will come into being.

ABSTRACT

As the largest library in China, the National Library of China (NLC) has spent more than 20 years in automating its processes, and has not yet had an integrated library management system, lagging behind libraries of developed countries and also other major libraries in China. In the 1980s and 1990s the NLC made slow progress; reasons for lack of success in the past include the inadequacy of budgets, the extent of the collections, the pursuit of processing of materials in all languages in a single system, and some management problems. In 2001, the NLC signed a contract with Ex Libris to implement its Aleph 500 system and aimed at comprehensive solutions. Aleph 500 meets most of the library's requirements, supports Unicode and is flexible, extendible and cost-effective. The new system is expected to help the library to fully automate all of its processes, provide services better than other Chinese libraries and process all languages in a single system.