

PLANNING AUTOMATION OF EAST ASIAN CATALOGUE RECORDS IN THE
NATIONAL LIBRARY

S W Wang, N White, A Gosling
National Library of Australia

I. INTRODUCTION

After adopting the Washington Library Network (WLN) and successfully establishing the Australian Bibliographic Network (ABN), the National Library of Australia is now seriously thinking of adopting a system which is capable of creating, copying, amending, displaying, retrieving and outputting bibliographic records of East Asian scripts for its East Asian collections. This paper aims to examine the present state of available systems and makes some suggestions.

II. DEVELOPMENTS IN THE AUTOMATION OF EAST ASIAN LANGUAGE
CATALOGUES IN JAPAN, TAIWAN, KOREA, CHINA, THE UNITED STATES
AND AUSTRALIA

In the last few years, dramatic progress in the automation of East Asian cataloguing has been made. The first Japanese MARC on tape was made available by the National Diet Library in Tokyo in April 1981. The first Chinese MARC tape produced in Taiwan was distributed by the National Central Library in Taipei in December 1982. The Korean MARC was tested on JEPCOM 4296 by the Central National Library in Seoul in August 1982. The National Library of China in Peking reported that attempts had been made in 1982 to establish a machine-readable record for 40,000 Chinese serials. In the United States, the Library of Congress announced that the first on-line cataloguing record in the Research Libraries Information Network (RLIN) database to contain Chinese script was created on 12 September 1983.

All these developments have proved that the automation of East Asian language catalogues is no longer a problem. The problem now is which automation system is suitable for East Asian libraries in Australia. Except for the RLIN/CJK system which covers the Chinese, Japanese and Korean languages, all other systems apply only to their respective languages.

Japanese MARC

- a. Record structure is based on ISO-2709 (Documentation-Format for Bibliographic Information Interchange on Magnetic Tape).
- b. Format structure follows UNIMARC.

- c. Character set: Japanese Graphic Code (NDC-70) devised by the National Diet Library in 1970 is used. It includes the following:

Symbols and marks	293
Numerals	90
European alphabets	423
Kana	341
Kanji	<u>4570</u>
TOTAL	<u>5717</u>

However, the Japanese Government officially adopted in 1978 the Code of the Japanese Graphic Character set for Information Interchange prepared by the Agency of the Japan Information Processing Development Centre as the national standard, JISC6222 - 1978. This code includes the following characters and is slightly different from the set used by the National Diet Library:

Diacritics	108
Numerals	10
Roman characters	52
Greek characters	48
Russian characters	66
Hiragana	83
Katakana	86
Kanji	<u>6349</u>
TOTAL	<u>6802</u>

- d. Cataloguing rules: The Nippon Cataloguing Rules. It conforms to ISBD but does not follow ISBD strictly in every detail in order to reflect Japanese special circumstances.
- e. Computer facilities: NEC Lingual Processing System Company's JEM-3100 II/D System, TS-3100 E/D (CRT terminal of JEM-3100 system), JEM-7000 line printer (used mainly for proof reading), and JEM-7500 soft-printer (used mainly for master sheet).
- f. Annual subscription rate by air parcel: ¥760,000 for Australia and New Zealand. Subscription orders should be sent to:

Maruzen Company
 PO Box 5050
 Tokyo International 100-31
 JAPAN

Chinese MARC produced in Taiwan

- a. Record structure on magnetic tape follows ISO-2709.
- b. Format structure was formulated with UNIMARC and LC MARC in mind.
- c. Character set: The Chinese Character Code for Information Interchange (CCCI) compiled and published by the Chinese Library Association of Taiwan in April 1980. This code is basically a 7-bit code of ISO-464 and contains the following characters:

Most frequently used Chinese characters	4807
Punctuation marks	35
Radicals	214
Numerals	41
Chinese phonetic symbols	37
Tone marks	<u>4</u>
TOTAL	<u>5138</u>

- d. Cataloguing rules: The Chinese Cataloguing rules revised edition is adopted for the description of bibliographic records. The rules were drafted in accordance with the principles of ISBD and the Anglo-American Cataloguing Rules, 2nd edition (AACR2). Some variations were made to reflect the peculiarities of traditional and current imprints.
- e. Romanization: The Wade-Giles system of romanization is used for transliterating title proper, statements of intellectual responsibilities, physical descriptions, series, subjects etc.
- f. Computer facilities: Wang Laboratory's VS 100 facilities are used. The Three Corner Coding Method, 2 Arabic numbers standing for each corner of a Chinese character, is adopted for inputting purposes.
- g. Chinese MARC tapes are available from the National Central Library in Taipei.

KOR (Korean) MARC

- a. Record structure is based on ISO 2709-1973.
- b. Format structure is based on UNIMARC, LC MARC, and UK MARC.

- c. Character set: Draft of standardization for computer processing announced by the Korean Ministry of Science and Technology in 1982. 1692 Chinese characters have been defined as the standard code. According to Hyeon Kyn Seob, Head, Office of Library Automation of the Central National Library in Seoul, about 5000 Chinese characters are actually used. The compositions of the Korean character set is not known.
- d. Cataloguing rules: The Korean Cataloguing Rules, 3rd edition which was revised in compliance with the principles set by ISBD and AACR2 is used.

Chinese MARC Format for Serials

- a. Record structure follows ISO-2709.
- b. Format structure was formulated as closely as possible to UNIMARC. LC MARC, Japanese MARC and Chinese MARC produced in Taiwan have also been consulted.
- c. Character set: The code of Chinese Graphic Character set for Information Interchange (CCGCSII) is used. It was drafted by the North China Institute of Computing Technology, adopted as the National Standard GB 2312-80 by the Chinese State Bureau of Standardization and published by the Technical Standard Press in May 1981. The standard totals 7445 Graphic characters,
- | | |
|--|-------------|
| including general characters | 202 |
| (including space,
punctuation marks, signs,
unit marks, and
tabulation marks) | |
| Ordinal numbers | 60 |
| Latin letters | 52 |
| Japanese Kana (83 hiragana
and 86 katakana) | 169 |
| Greek letters | 48 |
| Russian letters | 66 |
| Chinese phonetic symbols | 26 |
| Chinese phonetic-annotated
letters | 37 |
| Chinese characters | <u>6763</u> |
| TOTAL | <u>7445</u> |
- d. Serial cataloguing rules follow the principles of ISBD(S).

In Australia, so far no East Asian script record has been input onto any on-line database. However, East Asian records in romanized form are being used by several computer-based systems, notably the CAVAL system in Victoria established in 1980.

CAVAL (Cooperative Action by Victorian Academic Libraries) covers mainly academic libraries in Victoria. Six of them hold East Asian materials. Bibliographic records for East Asian materials are incorporated in the CAVAL system in completely romanized form. The romanized records are also produced in either COM fiches or cards for the union and individual library catalogues. The University of Melbourne Library maintains a parallel title card catalogue with title and author in Chinese and Japanese scripts added manually.

East Asian materials input onto CAVAL are coded in AUSMARC 3 Format, following AACR2 and LCSH (8th edition plus supplements).

The Wade-Giles system is adopted for Chinese materials and modified Hepburn for Japanese. CAVAL also follows LC's practice of providing cross references in Pinyin form for all headings.

Apart from the above, two Chinese coding systems were developed by Australian scholars: H.R. Hwa, C. Chung and Z. Z. Ding of the University of Sydney, and Zwi Barnea and Shui-yin Lo of the University of Melbourne.

Hwa and associates' system used English letters as code symbols and 8000 Chinese characters have been coded. The length of each coded word averages 4 letters. Barnea and Lo's method adopts 1, 2, 3, 4 and 5 to represent the following five fundamental strokes of Chinese characters respectively:

- . (dot)
- (horizontal)
- I (vertical)
- / (positive diagonal)
- and \ (negative diagonal).

About 4000 Chinese characters have been coded. The length of each coded word averages 3.7 digits (key strokes). A firm in Hong Kong, named Sinotronics Limited, using Barnea and Lo's method has produced the Sinotronic CS-4000 Chinese character processor.

Among many Library automation networks in the United States, only the Research Library Group (RLG) has developed the RLG/CJK system and the Online Computer Library Centre (OGLC) is working on a Chinese-Japanese-Korean library support package.

The RLG/CJK system

- a. Record structure: ISO-2709.
- b. Format structure: LC MARC with additional fields for East Asian scripts.

- c. Character set: RLIN East Asian character code (REACC). It consists of 14063 characters which were compiled from the Sinoterm set of 10934 Chinese characters, the CCCII set of 4807 Chinese characters, the CCGCSII set of 6763 Chinese characters, the JIS C622 - 1978 set of 6349 Chinese characters, and the Korean set of 2392 Chinese characters.
- d. Computer facilities: RLG uses the Sinoterm CJK system designed by the Transtech International Corporation in Massachusetts. The CJK terminal is actually a cluster of hardware and software including a micro computer, controller for one to four work stations and a specially modified General Electric 3000 series printer. The keyboard has 179 keys, 135 of which contain 245 Chinese character components, 96 Japanese katakana and hiragana characters, 33 Korean hangul letters, an extended roman alphabet character set. The remaining parts are function and control keys. The terminal uses component entry input system which allows the user to enter characters by keying a sequence of their constituent parts. It takes an average of 3.6 key strokes to create a Chinese character.

The OCLC/CJK system

OCLC announced recently that it and Asiagraphics of Mt Sinai, New York, had entered into negotiations leading toward the development of a Chinese-Japanese-Korean (CJK) library support package. The package and a CJK word processor, already developed by Asiagraphics, will be marketed and supported by OCLC. The system allows users to retrieve transliterated Chinese records from the OCLC database, and, using software developed by Asiagraphics, display on the terminal screen or print out these records in photographic characters using a dot-matrix printer attached to the OCLC M300 work station (a modified IBM-PC). The software will be able to support both the Pinyin and Wade-Giles transliteration schemes as well as full and simplified character sets of the Chinese language. Development of retrieval capabilities for Japanese and Korean records is underway. Production of catalogue cards in Chinese, Japanese, and Korean characters is also being developed. However, this system makes access to Chinese materials in the database based on phonetics rather than the components of the Chinese characters. Libraries adopting this system will not have to acquire a specially-designed keyboard, terminal and a special printer capable of printing Chinese characters.

Systems developed in Japan, Taiwan, Korea, and China use different Chinese character sets and apply only to their respective languages. No one system in these countries is

designed to cover all three East Asian languages. This makes it impossible for libraries in Western countries to adopt them. The OCLC system has not been fully developed yet and the suitability of it for us is also uncertain. The RLG system combines the features of Japanese, Taiwanese and Korean systems and follows closely LC MARC. It seems to us that the RLG system is the best system so far designed for the East Asian libraries in Western countries.

III. CHINESE, JAPANESE AND KOREAN (CJK) ENHANCEMENTS TO RLIN

RLIN developed by RLG in the United States, is a computerized library services system, with various programs relating to research library matters. Members share data in an on-line database, and are able to search bibliographic files, acquire, process and catalogue materials and update the on-line data as necessary.

In 1979, RLG applied for funds from several foundations for developing enhancements to RLIN to allow the inclusion of records in East Asian scripts. This application for funding was successful and the CJK project began in 1980. RLG also entered an agreement with LC involving cooperation between the two in the development of CJK. LC undertook to begin on-line cataloguing of its CJK material into RLIN once the project came to fruition.

In February 1981, RLG sent a request for proposals to more than 30 firms in the US, Japan and Taiwan. None of the proposals put forward were considered wholly suitable. However, after investigation, RLG entered into an agreement with Transtech International for the development of the CJK terminal. This CJK device is a modified version of Sinoterm, a Transtech product operational in Taiwan. Sinoterm is a component-entry device, using 245 "word-roots" consisting of complete Chinese characters, or basic strokes, or shapes, to compose, display and print Chinese characters. To create a particular character, a sequence of these key strokes is employed. For example, the character 司 is built up from the three elements 丨 - 口

A number of modifications to Sinoterm were required for CJK. Instead of one terminal there is a terminal cluster with a micro computer controller and from one to four CJK terminals. The bottom line of the display screen has been modified to include a "scratch pad" where Chinese characters and Korean syllables can be built up and then entered. The screen can display Chinese characters, Japanese Kana, Korean hangul, Roman letters and numerals. This display can then be transferred to paper. Initially, 14063 Chinese characters (including traditional, simplified and variant forms) and 1907 Korean characters

were contained in the CJK local dictionary. In addition, another file, the CJK Thesaurus, contains over 33000 characters which can be added to the local dictionary as needed.

CJK records can also be displayed on non-CJK terminals. In this case, only the romanized data, not the CJK script, is displayed. Cataloguing rules within RLG stipulate that there must be entry in both vernacular script and romanized form at least for the title, statement of responsibility, edition statement, imprint and series statements.

Following the successful development of RLG/CJK terminals, LC began using them to catalogue East Asian script materials on-line into RLIN on 12 September 1983. By October 1983, CJK terminals had been installed at six RLG member institutions, with another eight institutions joining this number by 1984. Sixteen institutions are now able to use RLIN for the creation, storage and retrieval of bibliographic records containing Chinese, Japanese and Korean characters.

RLG/CJK is at present restricted to its members. It is also incompatible with OCLC and with WLN, on which ABN is based. This is a serious problem for many East Asian libraries in Australia. The National Library of Australia finds it difficult to purchase the RLG system in its present form. However, RLG is currently aiming to develop software which would allow non-members of RLIN to use CJK. RLG hopes to make this software available for export within two years. To date there is no indication of the likely cost of such software.

Unless a better system than RLG/CJK becomes available, the National Library will most likely investigate the possibility of acquiring the RLG/CJK system.

IV. SHORT TERM MEASURES

In the meantime, several short-term measures are available as options. Since September 1983, the Library of Congress has been creating wholly romanized records for Chinese, Japanese and Korean monographs. Wholly romanized records for Chinese, Japanese and Korean serials have been available on MARC through the CONSER database since the late 1970s. These records and the CAVAL records mentioned earlier are now available on ABN. In addition, there are also the fully romanized records from WLN already on the database.

The Chinese, Japanese and Korean records have already proved useful in a number of ways in the Orientalia Section at the National Library, such as bibliographic verification in cataloguing and in reference enquiries. In addition, since the romanized record appears on the database quite some time before the LC card with the vernacular script is received in the Library, it can be used for interim cataloguing. All that need be added is the Library's call number and location symbol. And, of course, relevant subject and name authority cards made. The Library would not create original catalogue entries for Chinese, Japanese and Korean works for ABN as wholly romanized records are a stop-gap measure.

This will enable other libraries in the network around the country to know the work is held somewhere in Australia as, at present, a holdings search done on such entries will be negative. As these records are all for new works, none of them appears in UCEAM (Union Catalogue of East Asian Monographs) on microfilm. So this will make ABN a useful location tool for East Asian libraries as well as others.

Moreover, it will enable material that has to be catalogued as a matter of priority to be processed quickly and without the time-consuming business of creating an original record which will only be superseded when the LC card arrives. This will speed up the despatch of inter-library loans among other things.

For our own records, the call number assigned to the ABN record will also be added to the temporary card in the card catalogue. When the LC card arrives, the call number will be added to it and it will replace the temporary in the usual way.

This does mean the keeping of dual records but still the main emphasis will remain on the card catalogue because it contains the vernacular script. Also at present, there are still very few Chinese, Japanese and Korean monograph entries on ABN.

In addition, when books are catalogued using LC cards which bear the designation MARC: thus indicating the presence of a romanized record in the database, the call number will be added to ABN as well as an extra step in the cataloguing process.

For serials, the situation is slightly different in so far as the loss of characters does not impair comprehension as much as with monographs. Indeed, wholly romanized serial records have been acceptable for some time, as witness the CONSER database. The National Library considered

computerising the Chinese, Japanese and Korean serial catalogues in the late 1970s, using the SALSSAH programme but time and staff restrictions prevented this. Also there are more existing Chinese, Japanese and Korean records on the database for serials. The first stage would be to add the call number and holdings to those titles already held by the National Library. There are some problems with this as WLN records for serials are rather brief, containing the minimum of bibliographical information and no subject headings. The question would be whether to accept these or add more detail of our own from our existing serial catalogue. The LC records are more complete but only exist from the late 1970s. The retrieval rate on the most recent is not high for Japanese - out of 8 titles, only 1 was found. Nonetheless, it does seem much higher for earlier titles and for Chinese.

Then, time and staff permitting, would come the entering of original records where none exists, together with the call number and holdings. Again, this would be for the benefit of other libraries using the system since no union list of East Asian serials has yet been published. ULEAS remains in card form in the Orientalia reading room and the National Library's own published lists of Chinese, Japanese and Korean serials are now out of date. It will also complement the Chinese, Japanese and Korean romanized records originating from CAVAL.

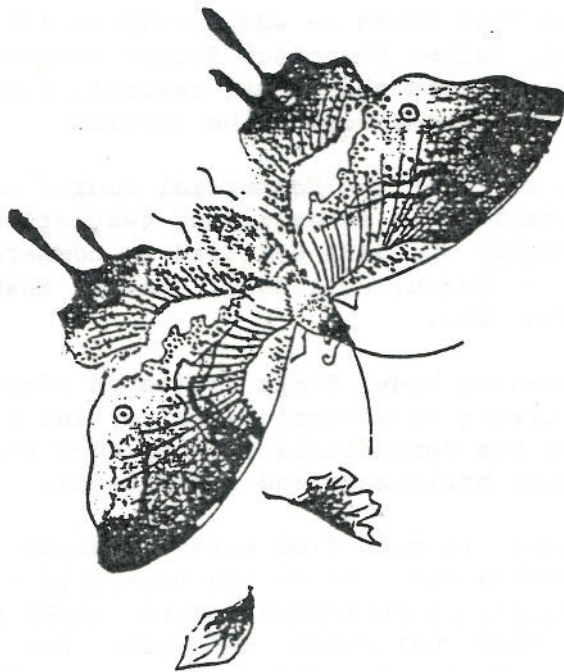
The serial card files would be retained because of the vernacular script. Another interesting possibility being considered by the Serials Section of the National Library is the computerisation of permanent serial records and visible indexes. This could also be done with the Orientalia serial records, though it might be better to wait for the adoption of a system capable of handling East Asian scripts since there is no real urgency in the matter.

V. LONG TERM PLANNING

Having examined various options for the automation of East Asian script records, the National Library is looking seriously at the possibility of adopting the RLG/CJK system. NLA is keeping in close contact with RLG, and during mid-April Mr Anthony Ketley, Director of Bibliographical Services, visited Stanford University for discussions with RLG, and for talks about their involvement with RLG/CJK. After publication of his report, we should have a clearer picture of when hardware and software will be available (RLG currently estimate the software will be ready for export by late 1985), costs involved and other details.

If RLG/CJK can be adopted, it will bring substantial benefits. It will enable more effective management of acquisitions, cataloguing, information retrieval and inter-library loans for East Asian script materials.

In the future, assuming the adoption of RLG/CJK or some other system, it should be possible to establish a nation-wide network for East Asian language materials. This would probably be through ABN, provided that ABN is capable of handling CJK scripts. If not, then other options would need to be investigated. At any rate, through new technology Australia should gain better collection management and development, resource sharing and vastly improved access to East Asian script library materials within the next few years. This is an exciting prospect for all those concerned with Chinese, Japanese and Korean collections in this country.



十竹齋珍藏