

REPORT ON THE RETRIEVAL OF EAST ASIAN RECORDS FROM RLIN

INTRODUCTION

Trials on RLIN were carried out by staff of the Asian Studies Division of the ANU Library between mid-March and mid-May, 1986, in order to test the retrieval rate for Asian-language material. A dozen languages were tested; the East Asian results are discussed here.

The aims of the trials were to test the retrieval rate, to evaluate the usefulness of RLIN records for ANU cataloguing and to provide data for use in assessing the cost-effectiveness of a regular RLIN link-up.

BACKGROUND

ANU

The East Asian collection of the ANU Library comprises some 183,000 monograph volumes and 4450 serial titles. East Asian material is still ordered and processed manually, although East Asian serials are also entered into URICA in romanised form. East Asian material is classified according to the Harvard-Yenching classification system.

Chinese monographs are given interim cataloguing off-line; this consists of assigning a classification number and subject heading and checking the shelflist, author and subject authorities. Japanese monographs and serials are given full cataloguing off-line. (Full cataloguing includes the addition of vernacular scripts.) Serials are also entered on-line in brief format. Serials are organised alphabetically by title on the shelves and do not need a classification number.

RLIN

The Research Libraries Information Network (RLIN) is an on-line information system offering a bibliographic database with integrated search, acquisition, cataloguing and inter-library lending facilities and with supporting authority files.

Records are loaded from the US Library of Congress, National Library of Medicine, Government Printing Office and other member libraries, mainly, but not exclusively, in the USA. The British Library and Academia Sinica, Peking, are examples of overseas members.

Member libraries with major East Asian language collections include: Library of Congress, Stanford, Yale and Cornell Universities, the Universities of Princeton, Michigan, Columbia, Chicago, Harvard and the University of Toronto.

By March 1985, 26 member libraries had Chinese-Japanese-Korean (CJK) terminals for handling East Asian vernacular scripts. CJK vernacular scripts can be stored in the RLIN database but only displayed on a high resolution terminal with monochrome graphics adaptor (such as the IBM/PC AT or AX) accompanied by appropriate software. This equipment was not available at the time of testing.

METHOD

Access to RLIN was supplied through a Hewlett-Packard 2621P terminal and telecommunications link to RLG Computing Facilities in the USA via AUSTPAC.

One member of staff with typing skills and knowledge of Chinese and Japanese performed the search.

Samples of Chinese, Japanese and Korean monographs were taken from the new book collection. Reference material was not included. Chinese monographs published between 1974 and 1979 had already been unsuccessfully checked in the past against the holdings of 3 major US collections, namely, Library of Congress (LC), Stanford University and the University of Michigan, all of which were members of RLIN. In addition, since joining RLIN, member libraries have not input records retrospectively. Therefore a high retrieval rate for this category of material was not expected. A sample of Chinese non-export serials was taken from the new book collection of uncatalogued material. Since non-export material is difficult to obtain outside China, here again, a high retrieval rate was not expected.

The search was for known items which, for the most part, had no ISBNs or ISSNs, therefore searching was by title. Author searching presents more problems than title searching because of greater variations in romanisation and in forms of personal names.

LC records were obtained wherever possible.

RESULTS

Table 1

	Publication dates						Total records searched	Total records found	Time taken (min)	Average time per title (min)
	To 1979		1980-1983		1984-1985					
	sought	found	sought	found	sought	found				
CHINESE	64	22	128	81	78	11	270	114	310	1.14
Hit-rate Percentage		34%		63%		14%		42%		
CHINESE NON-EXPORT SERIALS *							184	74	151	.82
Hit-rate Percentage								40%		
JAPANESE	77	32	131	88	42	25	250	145	257	1.02
Hit-rate Percentage		41%		67%		59%		58%		
KOREAN	20	8	20	7	11	6	51	22	55	1.07
Hit-rate Percentage		40%		35%		54%		43%		

* Serials have no single date of publication

Retrieval

Best retrieval rates for Chinese and Japanese were for material published between 1980 and 1983. The poor figures for 1984 and 1985 reflect the delay between the publication of a work and its acquisition and processing by US libraries, particularly since reference works and other high priority cataloguing material were not included in the samples.

The percentage of Chinese serials retrieved was higher than expected. Restrictions on the export of serials have loosened to some extent. Whereas a library has or does not have a monograph, having a serial could mean holding anything from one issue to a complete run.

The low Korean retrieval rate may be due in part to problems of romanisation. However, the sample was too small to be conclusive.

Search time

The time taken searching for each title was fast and response time generally good. When a hit was found, it took about a minute to retrieve the MARC record. Browsing through multiple retrieval sets added considerably to search time and was avoided as far as possible by means of search strategy.

Search strategy

Some factors became apparent which affected retrieval.

The Library of Congress rules for Japanese word division do not cover every contingency and variations can occur. Searching on a truncated title phrase reduced misses caused by variations in romanisation. Alternatively, titles were sometimes searched more than once on the possible variations.

Chinese, being a tonal language, has a different problem. Firstly, the Wade-Giles romanisation system does not distinguish between homophones; secondly, it differentiates between aspirated and unaspirated consonants by means of a special character which is ignored in on-line searching. Thirdly, the 'words' are set out monosyllabically, resulting in the splitting of compound words. All three factors increased recall and lowered precision, while the third factor also increased the incidence of false drops. For Chinese, then, a title word (keyword) search produced very large retrieval sets. A relatively long title phrase was needed to ensure a small retrieval set.

Availability of RLIN

RLIN was available between 5am and 9pm US time, Monday to Friday, and 11am to 7pm on Sundays. In Australia, access was poor on Mondays and RLIN went down at noon, Australian Eastern Standard Time. Initially, searching was possible on other weekdays until 3pm but during the latter half of the trials RLIN closed down at 2pm. When lines were busy, it was difficult to log-on and response time was slow: the worst times encountered were between 8.30 and 9.15am daily. This effectively limited search time, assuming a staggered lunch hour, to just under 5 hours a day, Tuesdays to Fridays, which had to be shared between the divisions of the Library.

Vernacular script

A low percentage of records before 1980 contained CJK scripts (see Table 2). Material published between 1980 and 1983 contained the highest percentage of CJK. Library of Congress records rarely contained CJK; those that did were from the LC Special Projects section.

Library of Congress card numbers

Library of Congress card numbers (LCCN) were found in all LC records as well as in some of those from other libraries, even when no on-line LC record was present (see Table 2). LCCN was present in all displays except the multiple display and in field 10 of LC MARC display.

Cataloguing levels on RLIN

RLIN encodes cataloguing levels with numbers from 9110 to 9668; the lowest numbers contain the fullest cataloguing, i.e. 9110 to 9120. Highest numbers do not contain subject headings apart from 9665. 9668 is an acquisitions record containing fields 100, 245, 260 and 300 only.

Classification

A very small number of Harvard-Yenching classification numbers was found. It was not worth the extra time taken to look through holdings of libraries likely to use Harvard-Yenching classification, particularly after an LC record had already been found and printed out.

Cost

The cost of RLIN was US\$63 per hour for search-only access; this rate was reduced in April to \$45 per hour. ANU Library administration estimated costs, inclusive of telecommunication charges, at \$5 per minute or \$3 per record at the former rate.

USEFULNESS OF RLIN

Table 2

LANGUAGE	PUBLICATION DATES														
	To 1979					1980 - 1983					1984 - 1985				
	Hits	Usable MARC records	Usable LC records	LCCN present	CJK present	Hits	Usable MARC records	Usable LC records	LCCN present	CJK present	Hits	Usable MARC records	Usable LC records	LCCN present	CJK present
CHINESE MONO-GRAPHS	22	18	0	8	3	81	59	43	73	63	11	4	4	4	4
PERCENTAGES		81%		36%	13%		72%	53%	91%	77%		36%	36%	36%	36%
CHINESE * NON-EXPORT SERIALS											74	63	37	40	0
PERCENTAGES												85%	50%	54%	
JAPANESE MONO-GRAPHS	32	31	0	14	5	88	84	34	61	77	25	24	16	18	22
PERCENTAGES		96%		43%	15%		95%	38%	69%	87%		96%	64%	72%	88%
KOREAN MONO-GRAPHS	8	6	0	4	0	7	6	1	3	4	6	4	0	0	4
PERCENTAGES		75%		50%			85%	14%	42%	57%		66%			66%

* Serials have no single date of publication

The term 'usable' records, as referred to in Table 2, means those with content analysis.

East Asian copy cataloguing at ANU involves the checking of titles against hard copy catalogues or NUC microfiche, following which LC cards are ordered. Alternatively, the fiche or hard copy may be photocopied or photographed for printing. Classification numbers are added and authorities and shelflist checked, but the process avoids re-typing and re-checking the Roman and Asian scripts.

Although RLIN records were displayed in catalogue card as well as LC MARC format, the printout could not be used as copy or even as a draft card because diacritics and vernacular scripts were inaccessible on the existing equipment.

The clerical processes of re-making cards could be avoided by ordering LC cards from LCCN supplied on RLIN. By doing so, the average search time per record shown on Table 1 would be speeded up, as LCCN appears on the default display (unless it is a multiple display) and no further record need be extracted. One major drawback of ordering LC cards is the delay between ordering and their arrival. It should be noted also that LC cards cannot be ordered interactively through RLIN.

Since Library of Congress classification was not usable, Harvard-Yenching numbers still had to be assigned. Subject headings, therefore, were considered the most useful cataloguing aid present. The ready-made headings and punctuation in the catalogue card format mean fewer decisions for the cataloguer, and the usefulness of RLIN for copy cataloguing in general would increase where there was a lack of staff with relevant language skills.

Time spent on pre-catalogue checking, which took about 10 minutes per title manually against several sources, took about 20 seconds on-line.

CONCLUSION

RLIN brings together in one source records from the US libraries which ANU uses for East Asian copy cataloguing and provides LCCN even where no on-line LC record is present. This offsets the low percentage of LC records and the lack of vernacular scripts in them.

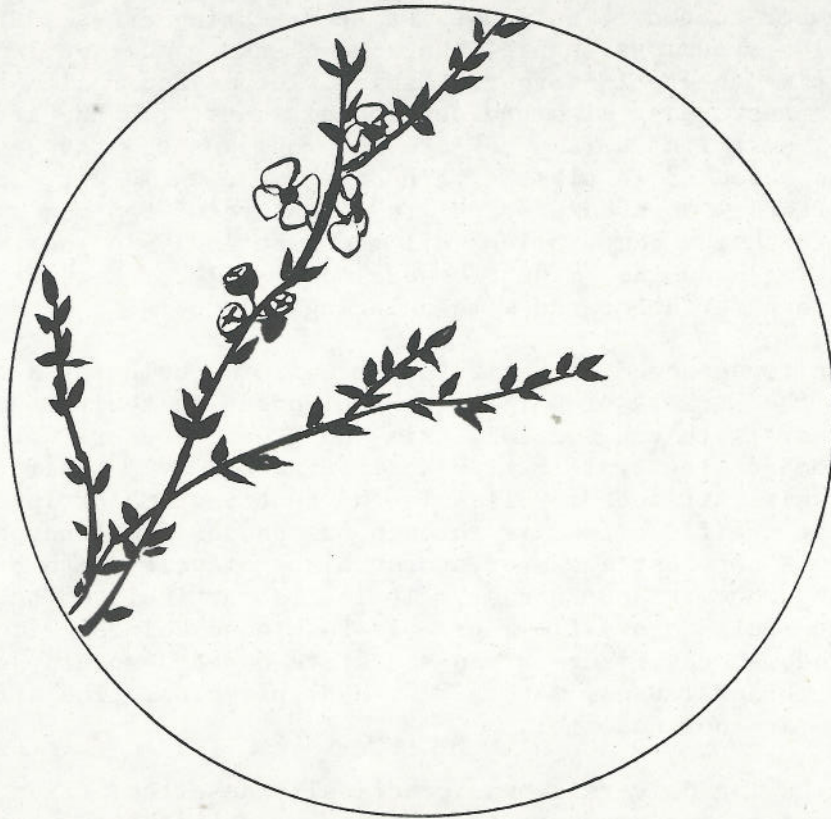
If draft cards are drawn up from RLIN printout, no clerical time and very little intellectual effort is saved except in assigning subject headings. On the other hand, by ordering LC cards using LCCN supplied on RLIN, savings could be achieved on the clerical processes of preparing catalogue cards, and great savings on pre-catalogue checking time. Searching could be done at L01 or DPO level from accession slips with titles already romanised.

As far as equipment is concerned, expensive hardware such as the Sinoterm terminal is unnecessary for search-only access, and no additional software would be needed to display vernacular scripts if LC cards were ordered.

At present the most efficient use of RLIN for East Asian cataloguing at ANU is as a source of LC card numbers. The situation may change and should be

reviewed at a later date. In the long term, the RLIN database may become available on compact disc complete with Asian scripts and would cut out the costs of telecommunications. Not only that, but it would be accessible throughout Australian working hours. In addition to providing cataloguing copy, RLIN can aid overseas inter-library loans and acquisitions, particularly in a specialised area such as East Asian librarianship. These factors should also be taken into account when weighing up the costs and benefits of a regular link-up with RLIN.

Susan MacDougall
The Australian National University Library



This is an abridged version of the article Report on the retrieval of Asian records from RLIN which appeared in the September 1986 issue of Cataloguing Australia, v. 12, no. 10, pp. 2-11. It appears here by kind permission of the editor.