

RLG/CJK AND OCLC/CJK AND IS THERE ANYTHING IN THEM FOR US? : A LOOK  
AT TWO EAST ASIAN SCRIPT AUTOMATION LIBRARY SYSTEMS IN THE USA

---

I. INTRODUCTION

There are, or will be, two library automation systems capable of handling Chinese, Japanese and Korean vernacular scripts (hereinafter referred to as CJK) in use in the USA at present. These have been developed separately by two computerised networks of libraries, the Research Libraries Group (RLG) and the On-line Computer Library Center (OCLC). That of RLG is older and has been running since 1983 while that of OCLC is expected to be available mid-1986. I was able to visit Library of Congress and see RLG/LJK in use in August 1985. This paper looks at the present state of both systems, some future developments and aims to raise some questions as to what these mean for East Asian libraries in Australia.

II. BACKGROUND

Since the 1970s, computers have been used increasingly in libraries to control bibliographical records. Libraries have combined to form networks, centralising their records in a main computer to which all the libraries in the network have access. This cuts the costs of and saves time in cataloguing as it means one library in the network can retrieve the record of another which has already catalogued the item in hand and use it rather than creating a whole new record from scratch and thus duplicating effort. More importantly, from a scholar's point of view, the book gets on the shelves quicker. It also means a library can readily see where a book is held if it does not hold the item itself, and can arrange to borrow it through interlibrary loan.

Because of the complex nature of the script, East Asian libraries in the USA and elsewhere were left in the backwash of this revolution. Both RLG/CJK and OCLC/CJK were developed by their networks in response to the need to bring East Asian libraries into the mainstream of librarianship as it is practised in the late 20th century.

The Research Library Group, which owns and operates the Research Libraries Information Network (RLIN), includes a number of very large East Asian libraries such as Yale, Columbia, Princeton, Cornell and Stanford. They developed their system in conjunction with the Library of Congress and Transtech International Corporation of Massachusetts.

The Online Computer Library Center is the largest such network in the USA and also includes a number of East Asian libraries. Their system developed from a CJK word processor, as did RLG's. Theirs was packaged by Asiagraphics of Mt Sinai, New York.

### III. SPECIFICATIONS

RLG/CJK's specifications are as follows :

- 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 10,934 Chinese characters, the Chinese Character Code for Information Interchange set of 4807 Chinese characters, the Code of Chinese Graphic Character Sets for Information Interchange set of 6763 Chinese characters, the Japan Industrial Standard C622-1978 set of 6349 Chinese characters and the Korean set of 2392 Chinese characters.

In addition, there is also a CJK thesaurus of 33,000 Chinese characters stored in the main computer at Stanford which can be accessed.

- d. Computer facilities : RLG uses the Sinoterm CJK system developed by Transtech. This consists of one to four terminals, a printer and a cluster controller. The cluster controller houses a microprocessor which enables the terminal to process data and to communicate with the large computer at Stanford. It also provides the capability for the terminal to retrieve and display Chinese, Japanese and Korean characters.

OCLC/CJK's specifications :

Record structure, format structure and character set are the same as for RLG/CJK. In place of RLG/CJK's additional dictionary of 33,000 characters located in the central computer at Stanford, OCLC/CJK has a program module to generate characters not in the data base. With a 20 MB disk drive, it would be possible to accommodate over 30,000 characters. That would require local entry or a future character set enhancement by OCLC.

Computer facilities : The OCLC/CJK workstation is configured around an SC/350 (an IBM XT with a special PROM which provides the full ALA character set display and the OCLC telecommunication protocols); a 10 MB Winchester disk drive, an Epson LQ 1500 printer and three applications software packages : online cataloguing, card production and word processing. This workstation can be used as an OCLC M/3000 (a modified IBM-PC) or as a regular IBM XT.

The most obvious difference lies in the respective keyboards. The original Asiagraphics software was designed to adapt to different kinds and sizes of computer systems. Any roman keyboard input hardware can be used without modification as long as the display is capable of bit-mapped graphics. So therefore OCLC/CJK's workstation has a standard IBM PC keyboard.

Input is natural language. In the case of the Chinese character, 容 for example, the pronunciation 'rong' is typed in, then the tone '/' (second rising tone), then the radical 'B' (ie. 'bao', one of 52 radical families derived by Asiagraphics from the traditional 214)

and then the space-bar is pressed to display the character on the screen. If there is more than one match, ie. there happens to be several characters with the same pronunciation, tone and radical, the options are displayed and the operator is asked to make a selection. Many common characters and compounds have option abbreviations which speed up input. It takes 3 to 8 keystrokes per character. The speed seems, according to one observer,<sup>1</sup> to be only slightly slower than for English language. A speed of 60 characters per minute was quoted by OCLC.

RLG/CJK's terminal has 179 keys, 135 of which contain 245 Chinese character components, 96 Japanese kana characters, 33 Korean hangul letters and an extended roman alphabet character set. The remaining parts are function and control keys. There are shift keys to enable the operator to switch from Chinese characters to Japanese kana to Korean hangul and to the roman alphabet.

The basis of input is components of Chinese characters, represented by the 245 keys. The operator strikes those keys containing the components that make up a character, in the proper sequence, which is usually the order the character is written. As the character is being built up, the components appear one by one at the bottom of the screen in the area called the scratch pad. When the character is complete, the space bar is struck, the character appears in the text on the screen and the scratch pad is cleared. For example: 月 + 𠂆 + 复 → 月复

Should the components result in two or more options for characters eg. 亻 + 王 → 任 or 全, the operator would be asked to make a choice. Similarly when two different characters with different meanings which have the same simplified form, both will be displayed with the command 'pick'. Although the character will still appear the same in the text, it will be linked to the correct full form in the system. The system is very flexible. It accepts variants eg. 月 radical for 肉 radical. If a search is done on the simplified form, it will retrieve all variant forms, including the full form and vice-versa unless specified otherwise.

It takes an average of 3 to 6 key strokes to create a Chinese character. After two months experience a rate of 40 characters per minute can be achieved.

RLG/CJK is based on graphics, the shape of the character being broken down into elements, then built up. OCLC/CJK is a combination of graphics (their 52 radical families) and phonetics (pronunciation of the character plus tone). RLG/CJK would appear to be slower and requires special training for its peculiar keyboard. However, I found that it was not so difficult to use, given that I was totally unfamiliar with the layout of the keyboard, and I got a great sense of achievement when I built a character and input it. There is a course and sheet to familiarise users with the keyboard.

Staff in the cataloguing section of LC to whom I spoke said that it took about one hour every day for a month before operators felt competent. Initial apprehension as to its complexity soon gave way as they found it interesting or a challenge. If an operator has a reasonable idea of how to write characters, she should be able to type in correctly 70% of all characters. For the remainder, one has to try to second-guess the person who did the original graphics, I was told. They added that if one person does the whole cataloguing from description through to subject, as in Orientalia in the National Library, rather than dividing these tasks among different groups as is done at the Library of Congress, and that person does the inputting as well, she or he would become much more familiar with the keyboard much quicker.

Since OCLC/CJK is not yet up and running, there are no reports on what it is like to use. It will be possible to better assess OCLC/CJK hardware when the National Library has had a chance to use on trial an Asiagraphics word processor which it will be receiving in the near future. We actually have the hardware and are awaiting the software. However, it would superficially, at least, appear to be less daunting, being a standard keyboard and the literature claims that anyone who knows how to pronounce the character can easily learn how to use it. It certainly would be faster, especially if the operator can touch type as the layout would then already be familiar. The main problem I can see is with Chinese inputting with its need for tones. Tones are difficult enough for native speakers to identify let alone non-native speakers. Chinese staff in Orientalia in the National Library of Australia consider this something of a stumbling block. On the other hand, this might well prove less of a problem once staff actually start working with it and become familiar with the system, much as RLG/CJK's keyboard with its 'bits' of characters has proved to be.

Maintenance of OCLC/CJK's work station would be less of a problem as any service organisation familiar with IBM PCs could do it, while RLG/CJK's terminal needs specialists familiar with Transtech's Sinoterm and these are found only in America at present.

Both systems use MARC cataloguing, both receive LC MARC tapes and both are adjuncts of large networks with substantial East Asian input. OCLC is the larger with over 6000 libraries in the USA, Mexico, Canada and the United Kingdom relying on its 20 regional network affiliates and two service centres. In addition, it has signed a contract with the National Central Library in Taiwan to obtain 40,000 Chinese records for loading into OCLC. The National Central Library will convert them from Chinese MARC to LC MARC before shipment in 1986, and will supply up to 6000 new records annually. Agreement in principle has been reached with RLG to obtain CJK tapes from RLIN. 90% of these, however, have been estimated as being available through LC MARC distribution service. While Japanese MARC records are available, for some reason, there are no plans to load them.

RLG is not anywhere near as large as OCLC but does, as mentioned above, contain some of the country's most important East Asian collections. It has members in Canada (University of Toronto) and in Hawaii.

In both systems records can be searched either by romanised form or by vernacular, by author, by title and so forth. Both support pinyin and Wade Giles romanisation, as well as Hepburn for Japanese and McCune-Reischauer for Korean.

OCLC/CJK has a card production package which supports local catalogue card production and transfer of machine-readable records to a local library system if an interface has been installed. RLG/CJK does not have a card production facility. The Library of Congress, for example, still sends its manuscripts of East Asian cataloguing to Japan for photocomposition into cards. Apparently other libraries in the network make a printout of the record as displayed on the screen and photocopy it in reduced form on to cards. Cards produced in this manner that I saw seemed acceptable but I was told that many were rather blurred.

OCLC/CJK has a wordprocessing package which could be used for bibliographies, correspondence, notices, even telexes. RLG/CJK was not designed primarily as a word processor (despite being modified from one) but there are two screens, one formatted and the other not, which can be used for correspondence and so forth.

As RLG/CJK is, as I have indicated earlier, the only one fully operational I had a chance to examine it in August 1985 at the Library of Congress. Staff then claimed it took 2 to 3 months to get back up to the normal level of cataloguing output and after 4 months there was a slight increase. They felt that with a smaller library like the National Library, the period might be less as they had 60 who had to be trained.

One feature I felt was particularly useful was the 'save' function key which enabled data from the interim record, once verified, to be transferred directly to the full record screen without the need to type it in again.

At the National Library, East Asian script materials are input onto ABN (Australian Bibliographic Network) in romanised form. LC MARC records we receive appear likewise only in romanised form with the vernacular fields stripped off.

Various levels of display of records are possible as with ABN/WLN (Australia National Bibliography/Washington Library Network). One useful feature is that most levels include a holdings statement of libraries having the bibliographic item. With ABN/WLN it is necessary to do a separate holdings search after the bibliographic search to obtain such data. However, ABN has planned enhancements that will include holdings statements at least in printouts. Searches can be done on a range of things including author, title, series, ISSN, rid number, CODEN and publisher.

One disadvantage I noticed is that the bibliographic file is not linked to the authority file as it is with ABN/WLN. This means that after doing a 'term' search to determine the correct form of an author's name or to see if it exists, one cannot go directly to the bibliographic file by keying in the appropriate number. One has to sign out of the authority file and then sign on to the bibliographic file, a somewhat time consuming and tedious procedure. I was told that ABN/WLN is the only network which does have linked vocabulary and bibliographic files but that RLG was working on the problem.

Staff mentioned that they found the Sinoterm terminal a little uncomfortable at times because there was glare not only from the screen but also from the keys which seemed unnecessarily shiny. Also, the screens are fixed and cannot be adjusted up or down.

Both RLG/CJK and OCLC/CJK seem to be powerful searching tools flexible enough to pick up variants and simplifications of characters; both have input from LC MARC cataloguing; both are parts of networks with large CJK databases; both can be used as word-processors as well as for online cataloguing and bibliographic searching.

However, RLG/CJK's keyboard is large and fairly complex. OCLC/CJK's keyboard is as for a standard IBM PC but relies for Chinese, on tones which could be a problem. Being based on a standard IBM PC OCLC/CJK would appear to be more flexible as it does not have a special terminal and keyboard. This would make maintenance, at least for the hardware, less of a problem. OCLC/CJK has full word-processing capabilities including telex. The system will incorporate Chinese MARC records from Taiwan's national library in addition to LC and RLG bibliographic data, thus making it more comprehensive. It is a pity that Japanese MARC will not be included. Moreover, OCLC/CJK is claimed to be 50% cheaper than RLG/CJK. Superficially, at least, OCLC/CJK looks the more promising but is as yet untried. Meantime, staff at the Library of Congress using RLG/CJK have expressed a high level of satisfaction with it.

#### IV. THE VIEW FROM AUSTRALIA

For Australian libraries contemplating 'the great leap forward' of computerising their CJK records fully, there are other things to consider.

Both systems as stated earlier, are part of large networks, designed to use online the database of those networks. No stand alone version is envisaged in either case. RLG/CJK is restricted at present to its members. No decision has been reached regarding sales to non-OCLC participants. However, RLG plans to develop software which would allow non-members of RLIN to use its CJK. In March 1985 it announced it was ready to distribute software that would allow an IBM PC with specific standard hardware to work like a RLIN terminal, which would be the basis for development of additional non-Roman script capabilities, beginning with Cyrillic, Hebrew, Greek and Arabic.

There are several possibilities for Australian libraries. The first is to join either RLIN or OCLC as associate members. The advantages are direct connection to a very large database which would help with cataloguing, reference and interlibrary loans. The disadvantages are costs and lack of adequate backup, especially with RLG/CJK. Not only do the fees charged by the networks have to be considered, but also the satellite connect charges which also opens up the question of time. How far would either of these networks extend their hours of operation so we in Australia could get value for money?

In 1984, RLG/CJK charges were US\$34,000 for a basic cluster of one terminal, one cluster controller and a printer and \$50,000 for a cluster with 4 terminals. Communications were \$750 a month (within the USA, of course) and a maintenance contract was \$500 a month per cluster. OCLC is around \$28,272 for two units, the first year, less in subsequent years. Add to these figures 'satellite charges' and some idea of the costs involved can be gauged. How many library managements even of large libraries, in Australia would be prepared to pay sums like that on top of their other computing costs for roman script materials, administration and so forth. Especially when East Asian language material is usually seen as a small and specialist component of their whole operation and not necessarily a very important part of it?

At present, RLIN, OCLC and WLN are incompatible with one another. Libraries in this country on ABN and other systems would have to run two systems parallel, say ABN and RLIN or ABN and OCLC. Can this be justified in expense when even the National Library's CJK cataloguing is so much smaller than that of the Library of Congress which has had to adopt the two system approach. Added to this, the Library of Congress, which is somewhat closer geographically to Stanford, still finds RLG/CJK more expensive than their own system, especially with the time problems.

Another possibility is not to join the networks but to gain access to them through a service organisation such as CLASS (Co-operative Library Agency for Systems and Services, based in California). The University of NSW, for instance, teaches cataloguing on the RLIN system via CLASS. CLASS has been defined as a "self-supporting not-for-profit membership organisation that offers a range of products and services to some 2,800 libraries ... throughout the United States and more recently, overseas. One of those services is access to RLIN."<sup>2</sup> It is a broker of RLIN services to all libraries not eligible for membership in RLG and believes there is a real potential for RLIN internationally though for this year they are concentrating their efforts nationally.

A third possibility comes from RLG's plans for the future direction of the network : the distributive network wherein members will do all their technical processing locally instead of on a central shared mainframe. This change has been brought about by rapidly rising communication costs within the USA. RLG is to develop the 'RLIN network server' "a microcomputer-based device to provide the gate-way between a local system and the RLG computer ..... The network server ... will allow us to evolve from a network of terminals connected to a host computer to one that supports intersystem communication between local systems and the RLIN system, which contains the union data base."<sup>3</sup> Members need only go to the central computer when data for their cataloguing, searching etc. is not on their own systems.

Information from the RLIN computer could be transferred to the local system and vice-versa. This brings to mind the whole concept of CD-ROMs and microcomputers WLN is looking at. To be able to buy the data of a network, say on one or more CD-ROMs and use it locally is a future possibility. This would cut satellite connect costs as the mainframe central computer need only be accessed for data not available locally.

The University of Pennsylvania, a member of RLG, is thinking along these lines: "Penn, like other libraries, will still need to search the RLIN data base to see if another library has acquired a particular item, to see if someone else has catalogued it, and for interlibrary loan. But even these activities perhaps can be supported in the future using a copy of the RLIN data base copied onto an optical disk and updated once a month. We could then contribute completely new cataloguing records for addition to the RLIN data base via the network server...." <sup>4</sup>

Or non-members in Australia perhaps could simply rely on the updates to the CD-ROM and avoid satellite connect charges and direct membership altogether as they need not or could not contribute to the database.

There is also the question as to how likely either RLIN or OCLC is to become truly a worldwide network? Bearing in mind that CJK cataloguing will always be a specialised part of any library operation, it is not economic for libraries in this country to 'go it alone' to produce their own CJK systems, especially when two large networks already have their own. If enough overseas East Asian libraries showed interest, then perhaps the networks might see it as worth their while to be more open, flexible and less parochial.

One development which would help this 'internationalisation' of automated library systems is the Linked Systems Project (LSP). This five year project is now in its final testing phases. It involves RLG, WLN, OCLC and the Library of Congress to create the protocols that would allow communication between their respective data bases, which are presently incompatible with each other. This came about as the realisation dawned that there was never going to be a single, national data base for libraries in the USA. Initially the project will be applied to authority work : "eventually with LSP connections in place up-to-date authority records will be accessible on-line... If the record is there [the participant] can use it. If not, they can establish it. And they will be looking at the same file LC cataloguers are using minus the records added during the last 24 hours". <sup>5</sup>

However, the most exciting aspect is the long-term goal of using the link for the exchange of bibliographical records. "It is not practical for one organisation to build and maintain a data base that is a 'union' of all our databases. So it's really when you get into bibliographic searching and information retrieval across systems that the link begins to pay off".<sup>6</sup> "When the LSP link is in place... a user who can't find the record he or she needs in RLIN will be able to search on an alternate data base and transfer the record to RLIN for copy cataloguing... And the potential for referred searching does not need to stop at national boundaries; the British Library and the National Library of Canada are two organisations planning on implementing links."<sup>7</sup>

LSP could even handle non-MARC records and a whole range of databases such as Chemical Abstracts or Lockheed DIALOG. A possible future use of LSP is interlibrary loan and even the transfer of text.

The big 'if' is costs, presently an unknown factor. At present they are concentrating on solving the technical problems, leaving the policy and economic ones until later.

What LSP would mean to us in Australia, assuming its cost was not too prohibitive, is that it would not be necessary to run two computer systems parallel as ABN/WLN would be able to 'talk' to both OCLC and RLG. So we could gain access to their CJK records, to search them and to transfer data from them for copy-cataloguing purposes. There is still the question of hardware capable of displaying CJK characters. OCLC's program will run off an IBM PC with bit-mapped graphic display capabilities. Would RLG's do the same once arriving via LSP? And there still remains the vexed question of costs and satellite connect time.

One is left with several questions or options. Do libraries join one of these networks as a full member contributing data as well as receiving it, thus making this rather specialised area of librarianship truly global? Is this feasible not only in terms of costs and logistics to the institution involved but to the network itself? Do they have the capacity, even in a distributive network? Do they really want to, attractive as the idea might seem, especially to scholars and participants; the idea of all of the East Asian bibliographic records of the libraries of North America, the United Kingdom, Australia, New Zealand, perhaps Europe at one's fingertips? Or can one climb on the coat-tails of OCLC and RLG's CJK databases, via LSP and ABN/WLN, deriving cataloguing data from overseas but not adding to them, somewhat as ABN does already with data from LC, BNB and WLN? How cost-beneficial would it be to have satellite access to them via LSP? How soon or how likely will it be possible to buy their data bases in disk form? Will these be available to non-members? Or will we have to become associate members to receive the benefits not only of the database in disk form but of LSP? Will we have to buy the whole data base or will it be possible just to buy the East Asian component?

Where does this leave us as regards our own networks? Few of the 17 libraries contributing to the Union Catalogue of East Asian Monographs are on ABN, fewer still of them are cataloguing their East Asian material on to it. It is not just the lack of CJK facilities that seems to be the problem. Some library administrators do not see the need for computers at all in their libraries, let alone for the East Asian collections.

There are a number of potentially exciting possibilities such as LSP which will enable the searching of hitherto incompatible data bases; or the availability of databases on disk for searching locally. How and to what extent these will affect East Asian libraries still remains to be seen.

Nikki White  
National Library of Australia

FOOTNOTES

1. ISCI reporting on OCLC/CJK demonstration at ALA mid-winter meeting, Chicago, 19-20 January 1986.
2. Research Library Group news no. 7/8 (May/Sept. 1985) p.9
3. Research Library Group news no 9 (Jan. 1986) p.12
4. Research Library Group news no.7/8 (May/Sept. 1985) p.3
5. Research Library Group news no 7/8 (May/Sept. 1985) p.6
6. Research Library Group news no.7/8 (May/Sept. 1985) p.13
7. Ibid. " " "
8. Ibid.



Kimura, Eiichi, 1906- c  
 Ch\_ugoku tetsugaku no tanky\_u / Kimura Eiichi cho. Tokyo : S\_obunsha, 1981.  
 591, 18 p. (T\_oy\_ogaku s\_osho)

Series.

LCCN: 81808727

L.C. CALL NO: B126.K494

ID: ONTG1468999-B

CC: 9665

DCF:

ONTG (c-9665 Ca0TU) MNUO (c-9115 MnU [CJK])

:? ful

Cluster 1 of 1

ID:ONTG1468999-B	RTYP:c	ST:p	FRN:	NLR:	MS:n	EL:	AD:05-25-8
CC:9665	BLT:am	DCF:	CSC:	MOD:r	SNR:	ATC:	UD:01-01-0
CP:ja	L:jpn	INT:	GPC:	BIO:	FIC:?	CON:b	
PC:s	PD:1981/	REP:	CPI:?	FSI:?	ILC:	MEI:?	II:
MMD:	OR:	POL:	DM:	RR:	COL:	EML:	GEN: BSE:
010	81808727						
040	\$beng\$dCStRLIN						
050	B126\$bK494						
100 10	Kimura, Eiichi,\$d1906- c						
245 10	Ch_ugoku tetsugaku no tanky_u /\$cKimura Eiichi cho.						
260 0	Tokyo :\$bS_obunsha,\$c1981.						
300	591, 18 p.						
440 0	T_oy_ogaku s_osho						
500	Title on t. p.: Studies in Chinese philosophy						
504	Includes bibliographical references and index.						
740 00	Studies in Chinese philosophy						

:+? fin tp ronko no kenkyu#

FIN TP RONKO NO KENKYU# - 2 clusters in BKS

1) Sat\_o, Ky\_ogen. RONK\_0 NO KENKY\_U / (T\_oky\_o : S\_obunsha, 1978.)  
 MNUO (c-9114 MnU [CJK])

2) Sat\_o, Ky\_ogen. RONK\_0 NO KENKY\_U / (T\_oky\_o : S\_obunsha, 1981.)  
 NYCO (c-9114 NIC [CJK]) NYCP (c-9114 NNC-EA) ONTG (c-9665 Ca0TU)

:? ful 2

Cluster 2 of 2

ID:NYC084-B3244	RTYP:c	ST:p	FRN:	NLR:	MS:	EL:	AD:11-05-8
CC:9114	BLT:am	DCF:a	CSC:	MOD:	SNR:	ATC:	UD:11-20-8
CP:ja	L:jpn	INT:	GPC:	BIO:	FIC:0	CON:b	
PC:s	PD:1981/	REP:	CPI:0	FSI:0	ILC:	MEI:1	II:
MMD:	OR:	POL:	DM:	RR:	COL:	EML:	GEN: BSE:
010	81809323/AJ						
015	Ja81-24058						
020	\$cY4500						
040	\$cNNC\$dNIC						
050 0	B128.W25\$bL646						
066	\$c\$1						
100 10	Sat_o, Ky_ogen.						
245 10	Ronk_o no kenky_u /\$cSat_o Ky_ogen cho.						
260 0	T_oky_o :\$bS_obunsha,\$c1981.						
300	4, 384, 19 p. ;\$c22 cm.						
490 0	T_oy_ogaku s_osho ;\$vno. 23						
500	Title on added t.p.: A study of the Lun-heng.						
504	Includes bibliographical references and indexes.						
600 10	Wang, Ch_ung,\$d27-97?\$tLun heng.						
740 01	Study of the Lun-heng.						

:+? fin tp tonko hekiga e#

FIN TP TONKO HEKIGA E# - None in BKS