Some Aspects of Using OCLC CJK for Public Service--A Japanese Librarian's View--

Kenneth K. Tanaka
SOME ASPECTS OF USING OCLC CJK FOR PUBLIC SERVICE
—A JAPANESE LIBRARIAN’S VIEW—*

Kenneth K. Tanaka
University of Maryland at College Park

I have used OCLC CJK for almost eight years, mainly for cataloging purposes. Despite this experience, I remain greatly challenged by the rapid changes and improvements in computers and in the OCLC CJK database.

Objective cataloging describes the bibliographic elements of individual works according to AACR 2 rules and assigns subject headings and classification numbers for content analysis of those works—that is, it creates bibliographic records. Public service searches and interprets the bibliographic records thus created in order to serve the needs of end users. For Chinese, Japanese, and Korean language materials, it is essential to view the original characters (vernacular) in addition to their romanized equivalents. Romanization does not convey information as precisely and as fully as the original characters themselves. The OCLC CJK database is extremely valuable to both catalogers and to public service staff because all the access points such as author, title, and so on are given in both romanized and original characters.


1. Bibliographic identification

If an item cannot be located in a library’s card catalog and in-house computer system, the OCLC database can be used to verify the author/editor, title, series, publisher, edition, and year of publication.

2. Holdings information

Once the citation is verified in the database, one can further determine if the item is available locally; and if not, for Interlibrary Loan purposes, which other libraries own the work.

3. Collection development

*This paper was presented in slightly different form at the panel discussion, “OCLC CJK for Public Service,” sponsored by OCLC on February 16, 1997, at the Mid-Winter ALA meeting in Washington, D.C.
The database can be used to identify materials for building a collection. However, if the item identified has been published very recently, there may not be a record yet in the OCLC database. There is a need for “Cataloging in Print for CJK” or “PromptCat CJK” programs, but overseas dealers in China, Japan, and Korea are reluctant to participate in these programs because it is not profitable and different cataloging rules, subject headings, and classification systems are used (e.g., Japan MARC).

There are other problems related to acquiring materials after being identified, problems not directly related to OCLC but which need to be kept in mind:

a. Academic publications older than two or three years are often not available because editions are small or press runs are small.

b. Several important periodicals are now only available in electronic format, at prices much greater than the print version. For example, an annual subscription for the CD-ROM version of Zasshi kiji sakai (Japanese periodicals index) costs nearly six times as much as the print subscription did. The online version of the same title is now accessible only when the Japanese online system is operating, or at night on the east coast of the United States.

4. Compiling bibliographies

Public service staff can assist a patron by conducting subject searches for materials to be used in a research project. This may include (1) checking related subject area records in OCLC by using “fin su” or “fin st” (subject, note, title) keyword search commands, and (2) verifying the bibliographic records.

5. Bibliographic instruction

Public service staff can conduct guided instruction sessions to using the OCLC database. Once users have mastered its use, they can perform their own searches at the OCLC terminals.

These instruction sessions should begin with searching for bibliographic records in WorldCat (the OCLC Online Union Catalog) using FirstSearch. FirstSearch in OCLC Reference Service provides keyword searching by author, title, and so on. This approach is easy and useful. However, WorldCat for FirstSearch users has certain weaknesses: (1) there is no vernacular, (2) there are no diacritical marks, and (3) a subject heading search with subject heading subdivisions does not function well unless the subject heading is a very specific one.

The following are examples of these weaknesses:

1. Japanese words with the same romanization may have different meanings. Without the vernacular, it is difficult to tell which meaning is meant.

For example, hashi may mean bridge 橋 , or edge 端 , or chopstick 箸 Engei may mean horticulture 園芸 or performing arts 演芸 Nichi-Bei Jikankyū Kyōgi may mean Japan-
United States Deputy Ministers Conference 日米次官級協議 or Japan-United States Conference on Wages 日米時間給協議

2. Diacritical marks often determine the meaning of words in romanized Japanese. Without them, it is difficult to know which meaning is intended. For example,

Oijsan is uncle おじさん but ojisan is grandfather おじいさん
Yosan is budget 予算 but yosan is sericulture 養蚕
Kinen is commemoration 記念 but kin'en is no smoking 禁煙

3. There were no results for a search of the subject heading "United States--Economic policy--1993-". A further difficulty for users is that they must be familiar with Library of Congress subject headings.

My experience is as a public service librarian and as the Japanese Bibliographer and Cataloger at the University of Maryland at College Park Libraries. UMCP Libraries still print CJK cards because the in-house system, CARL, does not have vernacular capability. When the CJK retrospective conversion project is completed and all materials are online (even though lacking vernacular), CJK cards will no longer be printed in-house because of the cost of maintenance and storage.

OCLC CJK

In the 1980s, I experienced frustration in romanized searches. Now, however, OCLC has the "scan" command, the "fin" keyword command, and the vernacular command, so that searching has become much easier. Besides, owing to the increase of OCLC member institutions' participation and the transferred bibliographical records from the RLIN system and the Waseda University Library's WINE (Waseda Information Network) MARC records converted into US MARC format, the availability of CJK bibliographic records has sharply increased. Formerly we experienced "log on" communication problems with OCLC CJK terminals three or four times a week when using dedicated line modem connections. Since our system was connected to the Internet late last year, this problem seldom occurs.

1. Romanization problem

It is very difficult to identify bibliographic records at a regular OCLC terminal without Chinese, Japanese, or Korean vernacular script. Chinese has two romanization systems, Wade-Giles and pinyin. The Korean romanization system, McCune-Reischauer, is not a well-constructed one. The Japanese-revised Hepburn system, known in Japan as the Hebonshiki へボン式 does not have many problems. However, in Japan the Kunreishiki 訓令式 (Japan Government system) is officially used in government agencies, including the National Diet Library, while commercial organizations tend to use the Hepburn system. For example, the 1979 edition of the Kokkai kokuritsu toshokan choshamei tenkyoroku 国立国会図書館著者名典拡録 (National Diet Library authority file for Japanese authors), edited by the National Diet Library, uses the Kunreishiki romanization system, while the 1989 edition of the Nihon choshamei jinmei
tenkyoroku 日本著者名人名典掲録  (Noted personalities in Japan since the Meiji Era), which is a compilation of Japanese authors’ name authorities edited by the editorial staff of the Nichigai Asoshiētsu, Inc., uses the Hepburn romanization system. The 1991 revised edition of the Kokuritsu kokkai toshokan choshamei tenkyoroku 国立国会図書館著者名典掲録  (National Diet Library authority file for Japanese authors) changed from the Kunreishiki to Japanese katakana readings without romanization. The following examples show the variant spellings of the two romanized systems for the same Japanese words:

**HEBON-SHIKI (Revised Hepburn system) KUNREISHIKI (Japan Government system)**

| Cha (Tea) | 茶 Tya |
| Chūjō (A Japanese surname) | 中條 Tyūzyō |
| Fujii Shin’ichi (Japanese full name) | 藤井真一 Huzii Sin’iti |

2. Word division problem

Word division is a difficult problem in Japanese romanization. When I was in Japan I was not aware of this problem because Japanese people use Japanese script, not romanization, to write Japanese. For instance, the romanization of 産業技術, which was published in 1995, will be Nihon ni totte Amerika towa nani ka, using the Library of Congress standardized word division system as given in the LC cataloging manual. Yet there are still problems in word division. Depending on whether a character phrase is romanized with hyphens or not, the romanization may vary. The following are examples:

<table>
<thead>
<tr>
<th>Hyphenated words</th>
<th>Without hyphen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shōgaku-dokuhon 小学読本</td>
<td>Shōgaku tokuhon (Primary school reader)</td>
</tr>
<tr>
<td>Otozi-zōshi 御伽草子</td>
<td>Otozi sōshi (Medieval short stories)</td>
</tr>
</tbody>
</table>

These minor things create searching problems in the OCLC database because the arrangement by title in the OCLC database is strictly word by word in alphabetical romanized order.

For example:

A. **One word** or **Two words**

Daihyakka 大百科 Dai hyakka (Encyclopedia)

B. **Variant methods of word division**

Hokkaidōritsu bijutsukan 北海道立美術館 (Hokkaido Provincial Art Museum)

Kanagawa kenritsu bijutsukan 神奈川県立美術館 (Kanagawa Prefectural Art Museum)

Furthermore, because this romanized title arrangement does not use the macron (a mark over a vowel indicating that it is long), searching is more difficult.
Advantages of OCLC CJK Plus

An OCLC CJK character search is advantageous in the following situations.

1. Identification of different editions
   Beijing edition vs. Taipei edition, because the same Chinese book is often published in mainland China and Taiwan. (In Japanese, han 版 means edition; however, some Japanese publishers use this term for printing.)

2. When the searcher is uncertain of how to read the author, publisher, or proper name in a bibliographic record, the correct romanization can be determined by means of a vernacular search.

A character combination in Japanese may have several different readings. For instance, the kanji 生 has 17 readings depending on how it is used in combination, such as ikiru 生きる (to live), seikatsu 生活 (living), shibafu 芝生 (lawn), issho 一生 (a lifetime). Readings of geographical names, personal names and pre-modern titles are often very difficult even for native Japanese. The following are good examples of this:

\[
\begin{align*}
\text{Oshamanbe-chō} & \quad \text{長万部町} & \text{(A town in Hokkaido)} \\
\text{Kunneppu-chō} & \quad \text{訓子府町} & \text{(A town in Hokkaido)} \\
\text{Datekurabe okuni kabuki} & \text{伊達競阿国戯場} & \text{(An 18th century kabuki kyōgen play)} \\
\text{Hikohohodemi no Mikoto ekotoba} & \text{彦火火出見尊絵詞} & \text{(LC subject heading; a medieval picture scroll of a Shinto deity in Japanese mythology who is known as Yamasachihioko, 山幸彦 the deity of hunting)}
\end{align*}
\]

A character search using “vt” (vernacular title), “vp” (vernacular personal name), “vc” (vernacular corporate name), and “va” (vernacular name/title) helps to solve the problem of how to read these combinations.

3. The same Chinese characters (kanji) will have different readings when read in Chinese, Japanese, and Korean.

Example: 華厳宗

<table>
<thead>
<tr>
<th>Language</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>Hua-yen tsung</td>
</tr>
<tr>
<td>Korean</td>
<td>Hwaŏn chong</td>
</tr>
<tr>
<td>Japanese</td>
<td>Kegonshū</td>
</tr>
</tbody>
</table>

Kegonshū, as the characters are read in Japanese, is a Mahayana Buddhist sect founded in China in the seventh century. By the middle of the seventh century it had spread to Korea, where Wonhyo founded the Korean Hwaŏn chong. The Japanese Kegonshū was founded in the eighth century. Although the Chinese and Korean sects are now extinct, the Japanese Kegonshū still exists. Its head temple is the Tōdaiji 東大寺 in Nara. This temple is famous for the huge
bronze Buddha enshrined there in 751. The subject heading for this sect has been established by the Library of Congress as follows:

<table>
<thead>
<tr>
<th>Hua-yen Buddhism</th>
<th>(Chinese sect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hwaon (Sect)</td>
<td>(Korean sect) Cross reference to Hua-yen Buddhism</td>
</tr>
<tr>
<td>Kegon (Sect)</td>
<td>(Japanese sect) Cross reference to Hua-yen Buddhism</td>
</tr>
</tbody>
</table>

CF:  Wonhyo (Sect)  (Korean Hua-yen Buddhism, its name taken from the founder 元曉, but no LC cross reference to Hua-yen Buddhism)

4. Tracing a Buddhist scripture’s uniform title from Chinese characters.

Hua-yen Buddhism (Japanese Kegonshū) was established based on the Kegongyō (Japanese reading), a very important Buddhist scripture. The Hua-yen ching is the name of the Chinese translation of the Sanskrit original. Note the following:

<table>
<thead>
<tr>
<th>華嚴経</th>
<th>(Chinese)</th>
<th>Hua-yen ching</th>
<th>(No cross reference to the uniform title)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Korean)</td>
<td>Hwaオン gyong</td>
<td>(Cross reference to the uniform title)</td>
<td></td>
</tr>
<tr>
<td>(Japanese)</td>
<td>Kegongyō</td>
<td>(No cross reference to the uniform title)</td>
<td></td>
</tr>
</tbody>
</table>

Library of Congress uniform title:
Tripitaka. #p Sutrapitaka. #p Buddhāvatamsakamahāvaipulyaśutra

(There are two Japanese readings for this scripture: Kegongyō and Kegonkyō.)

Searching problems

1. Availability of certain important characters (kanji)

Some kanji are available only in old kanji (kyūjitai 旧字体), but not in new simplified kanji (shinjitai 新字体 or jōyō kanji 常用漢字), because of character coverage limitation in the EACC (East Asian Character Code) set. (The original group of tōyō kanji 当用漢字 was expanded and became the jōyō kanji group, and Japanese simplified characters are considered now to be a part of the set of jōyō kanji.)

For example, sha 社 is available, but the simplified form sha 社, which is presently used, is not available.

2. Word division problems

If word division differs, the user may have difficulty locating the record, because OCLC records are arranged alphabetically word by word. For example, a well-known music publisher in Japan, 音楽の友社 is spelled out Ongaku no Tomosha, according to the LC name authority, while on the Waseda University’s WINE MARC record in OCLC, it is spelled as one word, Ongakunotomosha, because of the machine conversion of “kana reading” without word breaking
into a “romanized reading.” Also, according to AACR2, the first letter of each word for personal and corporate names must be a capital letter.

3. Searching difficult corporate name authorities.

If an end-user does not know how a particular corporate body is established, the record may be hard to find. The following are some examples:

<table>
<thead>
<tr>
<th>LC name authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyōto-shi 京都府   (Kyoto City)</td>
</tr>
<tr>
<td>Kyōto-fu 京都府   (Kyoto Prefecture)</td>
</tr>
<tr>
<td>農山漁村文化協会   (Rural Cultural Society)</td>
</tr>
</tbody>
</table>

4. Cumbersome approach to keyword subject searching in OCLC

When using the “fin su” command for subject heading, it there are several subject heading subdivisions, each word must be separated and connected with the “and su.....” command. The National Agricultural Library uses ISIS, an in-house system (Integrated System for Information Services). Doing subject heading searching in this system is very simple, and it is easy to get results quickly.

Example: s/Agriculture –Environmental aspects –Japan — Technological innovations.

5. Cross references to the Library of Congress subject headings and uniform titles relating to Buddhism

There is a need for more CJK cross references to Sanskrit forms related to Buddhist scriptures, sects, doctrines, and eminent Buddhists in ancient India.

Example: Moksa  (Subject heading)

(UF) Muki

This subject heading Moksa is gedatsu 解脫 in Japanese, which is a Buddhist doctrine meaning “spiritual deliverance from suffering”; however, there is no cross reference to Buddhism. I think this kind of subject heading in Sanskrit is too specialized for patrons to understand. Also, specific subject headings in Chinese, Japanese, or Korean need CJK vernacular. An example is the subject heading Myōkōnin 妙好人, which means devoted Pure Land Buddhist laymen (净士教信者) in Japan, especially Shinshū (Hoganji) Sect 真宗 (本願寺) laymen.

A practical solution to make existing Library of Congress Authority Files more useful (or user-friendly) to catalogers and other end-users of those files for CJK materials would be to attach vernacular characters below subject headings, name headings, and uniform title headings. These headings now appear only in the romanized version of the Chinese, Japanese, and Korean languages. However, there are some problems in cross references, if the vernacular form is involved in multi-language romanized forms. Take the example of 華厳経 (Kegongyō in Japanese) mentioned previously. This Buddhist scripture’s proper Chinese vernacular title is 大方広仏華厳経, though it is more commonly known in the shortened form 華嚴経. This important sutra (scripture) exists in China, Japan, and Korea. This means that 華厳経
can have its Chinese, Japanese, or Korean pronunciation. Furthermore, there were several translations into Chinese from the original Sanskrit text between the fifth and eighth centuries. The following are a few questions on this matter:

1. Which Chinese vernacular title should be used as the standard for the uniform title?
2. Should old Chinese characters or simplified Chinese characters be used? (Simplified Chinese characters were created after World War II in both China and Japan, but they are different forms. I believe that the old Chinese characters of the original translation should be used.)
3. Which tag number in the MARC format authority file should be used for the vernacular form in addition to the CJK romanized forms as cross references, because the uniform title of this Buddhist scripture is in Sanskrit?
4. Should all CJK romanized forms of the same vernacular title be added as cross references? Should all CJK romanized forms be indexed?

6. Name authority problems

Since the Library of Congress name authority records do not provide CJK characters, it is very difficult to identify authors. For example, Satō, Hideo is a common Japanese name. I found ten records for the name, including two born in the same year, “1937 -” and “1937 (March)” respectively.

7. Identification of tracing entries

In OCLC CJK Search Mode, all tracing entries are in romanized form only. Once end-users have learned to switch the display format from Card-view to MARC format by clicking on “Card view” menu under “Edit” main menu, CJK Search-only users can also display the record in MARC format and can find tracing entries with CJK vernacular data. But this procedure is rather complicated. Tracing entries with CJK vernacular is highly desirable for end-users in the OCLC CJK Search Mode.

8. Frustration occasioned by badly cataloged records and the lack of cataloged records

Since cataloging is done by people, human errors are unavoidable. Because OCLC operates through the cooperative efforts of member participants, the quality of records is uneven. Furthermore, the pressure to produce catalog records in any institution results in hasty products. Formerly the Library of Congress had a Tokyo Office for acquisitions under the Shared Cataloging and Acquisitions Program (May 1968-December 1985). Budget cuts forced the closing of several overseas LC offices, including the Tokyo Office, which resulted in the loss of many valuable acquisitions. LC catalogers cannot catalog books they never receive. In 1996, 63,054 new titles were published in Japan (not including uncountable numbers of non-commercial publications by federal and local governments, non-profit organizations, foundations, associations, companies, and private publications). The average cost of Japanese research books in hard cover is $50-$80 per volume. Because of limited allocations for acquisitions in East Asian libraries in North America, acquisitions of Japanese books are highly selective. And because most of the East Asian collections in North America support their institution’s East Asian studies programs,
the following areas of Japanese materials generally are not acquired:
(1) Science, technical aspects of agriculture, clinical medicine, and technology
(2) Subjects not related to Asian countries, especially China, Japan, or Korea (except books dealing with East Asian countries and other countries)
(3) Japanese translations from Western languages

In other words, acquisitions are focused on humanities and social sciences in East Asian countries and relations to and comparison with other countries. These factors result in a large gap in OCLC records. Many major East Asian libraries in the United States, including the Library of Congress, participate in the RLIN network, not in the OCLC system. However, following successful negotiations between these two major national online systems, RLIN CJK records are now available in OCLC.

Areas OCLC should address

1. The need for a local records file in OCLC CJK

The OCLC Online Union has been built on the principle of Shared Cataloging Work. Each member has a responsibility to build and maintain quality records in the Union Catalog. CJK users are provided with “Enhance” authorization. Though the actual workflow is ruled by the local library cataloging policy, CJK users are expected to enhance romanized records with CJK data when they have the book in hand. For their time and effort OCLC pays credits. Without this cooperative joint effort, OCLC could not exist. OCLC will enhance “national level” records, but each user must share “the lock/replace tasks” on non-national level records. However, in the case of national libraries’ records (such as the Library of Congress, the National Library of Medicine, and the National Library of Canada), even though we edit and add vernacular to the records, they revert to the original non-vernacular romanized form because the records of the national libraries may not be enhanced. Therefore, the record as it appears on OCLC may not be identical to the one on a local system. Also, if we do “edit” any OCLC record, the record in OCLC afterward reverts to the former original record. If the local file is available in OCLC CJK, it can substitute for printed cards in the public catalog file. (I have heard that 3.0 OCLC CJK Plus software for Windows 95 will have this feature and that it will be on the market at the beginning of 1998.)

2. The need for a Cataloging in Print Program for CJK materials

When we order newly published books, the bibliographic records are not usually available until the books are cataloged and input in OCLC. Like the approval plan with the PromptCat or the Cataloging in Print program for American publications, timely catalog records for CJK materials are crucial. Some Japanese publishers such as Nichigai Asoshibasu provide cataloging information on cards inserted in books; however, this cataloging information is based on Japanese cataloging rules and all entries including subject headings are in Japanese and so of minimal use for American users.

3. Union list of CJK journals with vernacular
The establishment of a separate file for a union list of CJK serials with summary holdings information in OCLC is highly desirable. OCLC provides the Serial Union List Group online local record creation facility, and union listing members must work cooperatively to keep up-to-date journals' holding information since the work of updating cataloging and holdings listing work is our member institutions' shared responsibility. I would like to ask the OCLC CJK User Group to coordinate and develop a standardized updating mechanism.

4. Access to the text of articles or tables of contents for CJK serials through Uncover

At present, OCLC's "First Search" databases, Article First and Contents First do not include CJK journals. A problem to be addressed is finding individuals to do indexing.

5. A brief summary or table of contents for non-subject literary works

6. Cooperative efforts toward retrospective conversion projects for CJK materials

OCLC CJK records of works published before 1980 are urgently needed. We encourage member institutions to convert their old records into OCLC as soon as possible. However, most retrospectively converted records are not full ("T") level due to costs and manpower shortages. Also, it is desirable for the Library of Congress to assign call number and subject headings to their minimal level "7" records.

7. OCLC supports for CJK user training and East Asian studies bibliographic instruction at local institutions

The forthcoming OCLC CJK Plus 3.0 new software will incorporate a new CJK Tour Guide in addition to CJK Help (Electronic user manual). CJK Help contains much information. Users may want to make printouts of search-related topics in Help for search-only end-users. Transferring data from Help and re-editing it into a booklet as a manual is an alternative possibility for end-users since OCLC does not plan to publish a book-format manual any more.