6-1-2000

Final Report on Pinyin Conversion by the CEAL Pinyin Liaison Group

Susie Cheng
Yu-lan Chou
Guo-qing Li
James Lin
Amy Tsiang

See next page for additional authors

Follow this and additional works at: https://scholarsarchive.byu.edu/jeal

BYU ScholarsArchive Citation
Available at: https://scholarsarchive.byu.edu/jeal/vol2000/iss121/8

This Report is brought to you for free and open access by the All Journals at BYU ScholarsArchive. It has been accepted for inclusion in Journal of East Asian Libraries by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Final Report on Pinyin Conversion by the CEAL Pinyin Liaison Group

Authors
Susie Cheng, Yu-lan Chou, Guo-qing Li, James Lin, Amy Tsiang, and Peter Zhou

This report is available in Journal of East Asian Libraries: https://scholarsarchive.byu.edu/jeal/vol2000/iss121/8
FINAL REPORT ON PINYIN CONVERSION

By

PINYIN LIAISON GROUP

COUNCIL ON EAST ASIAN LIBRARIES (CEAL)

March 2000

Susie Cheng, University of Hawaii
Yu-lan Chou, University of California at Berkeley
Guo-qing Li, Ohio State University
James Lin, Harvard University
Amy Tsiang, UCLA
Peter Zhou (Chair), University of Pittsburgh

Preamble

For nearly half a century, libraries in North America have been using Wade-Giles romanization for cataloging Chinese language materials. In 1997, the Library of Congress (LC) announced a decision to switch to Pinyin for the romanization of Chinese in cataloging and authority records. Pinyin is a romanization scheme widely used by governments, educational institutions, commercial publishers and news media in the Western world for transliterating Chinese scripts. Pinyin conversion, now scheduled to occur in 2000, will bring about systematic changes in millions of records, both those in Chinese and those in other languages that have headings, notes, or title added entries in Wade-Giles. This will be the single largest conversion of romanization systems in the history of American libraries to date. This report discusses issues related to the change to Pinyin in North American libraries and recommends the necessary steps libraries should take to adopt the use of Pinyin and to convert their cataloging records, following LC’s lead. This document is in the public domain, free of copyright restrictions. Therefore, we encourage CEAL members to adapt this report or change it into their own planning document in local deliberations on Pinyin conversion.
Table of Contents

I. History and current status of Pinyin conversion
   1. Use of romanization of Chinese scripts in Anglo-American cataloging
   2. Planning for Pinyin conversion
   3. A recommended timeline for CEAL libraries

II. Options for CEAL libraries and analyses of local conversion needs
   1. Conversion options using national utilities
   2. Name authority records
   3. Pinyin marker
   4. Other prerequisites for Day 1
   5. Split files
   6. Conversion of non-Chinese records and non-standard Chinese romanization forms
   7. Implementation of Pinyin conversion by CEAL libraries
   8. Implications for Public Services

Appendix: Major Pinyin conversion tools and documentation

I. History and current status of Pinyin conversion

1. Romanization of Chinese scripts in Anglo-American cataloging

   Before 1957, there were no set rules for cataloging Chinese language materials in Anglo-American cataloging. Different libraries varied widely with respect to the pattern and choice of language for bibliographic description and subject analysis. The use of Wade-Giles as the standard romanization scheme in American cataloging practice was initiated in February 1958 when the Library of Congress began to catalog Chinese materials in Wade-Giles, following the publication in 1957 of Preliminary Rules and Manual for Cataloging Chinese, Japanese and Korean Materials. While Wade-Giles became the standard romanization for Chinese scripts in North American cataloging, in 1958 China promulgated the Pinyin romanization scheme as its own standard for romanizing Chinese. That same year, soon after the new scheme was publicized, the British Library started to use Pinyin for the bibliographic control of Chinese language materials. Ever since its introduction, Pinyin has been gradually adopted by the international community as the standard Chinese romanization system.

   The Library of Congress (LC) first proposed conversion from the Wade-Giles system to Pinyin in 1979. The conversion was to take place in 1980, to coincide with LC’s introduction of computerized cataloging for Chinese materials. That plan failed to garner sufficient support in the East Asian library community and was given up. Again in 1990-91, LC publicly explored this issue and sought feedback from the library community. Though there was considerable support, strong concerns about varying standards for word division in Pinyin and the lack of computer programs for conversion of online records again defeated LC’s proposal. In 1996, the National Library of
Australia developed a conversion program that automatically converted 500,000 Chinese records from Wade-Giles to Pinyin. This influenced the Library of Congress and the East Asian library community to again explore the possibility of converting to the Chinese standard. In November 1997, LC announced that in the year 2000 it would begin using Pinyin as the new standard romanization scheme for cataloging Chinese materials and would at the same time convert all its existing Wade-Giles machine-readable cataloging and related authority records to Pinyin.

The East Asian library community endorsed LC’s decision. In May 1997, the Council on East Asian Libraries (CEAL) had already formed a taskforce to investigate the feasibility of adopting Pinyin as the standard Chinese romanization system for use in North American libraries. In its final report, the CEAL Taskforce supported the Library of Congress’ plan to convert to Pinyin romanization, recommending that such a program "be carried out only after a careful look at the impact of such a change on present national and local databases, on future additions to information about individual libraries, and on user access to the information." In May 1998, CEAL appointed a Pinyin Liaison Group to succeed the Taskforce and to represent CEAL in deliberations with LC, RLG and OCLC on matters related to the implementation of Pinyin conversion in North American libraries.

2. Planning for Pinyin Conversion

Since its November 1997 announcement, LC has issued Pinyin romanization guidelines and new classification schedules and has developed a conversion timeline in cooperation with RLG and OCLC. On June 29, 1999, RLG held a forum to discuss Pinyin conversion with representatives from LC, OCLC, CEAL and senior library administrators from eight research libraries with large Chinese collections. The panelists discussed RLG’s and OCLC’s plans for conversion, LC romanization guidelines, the sequence of conversion, and implications for local systems. On October 7, 1999, Harvard University organized a meeting at the Library of Congress with representatives from LC, OCLC, RLG, CEAL and selected libraries with large Chinese collections. During this meeting, representatives of these institutions reached an agreement on a conversion timeline for LC, RLG, OCLC, and individual libraries to follow. They also discussed various conversion options, name authority records (NARs), and issues related to local systems. On January 16, 2000, RLG held another forum on Pinyin conversion. At this forum, participants discussed RLG and OCLC conversion services, the conversion of non-Chinese records, and miscellaneous issues related to the implementation of Pinyin conversion. The latter included especially how to mark bibliographic and authority records that have been processed through Pinyin conversion.
The following key assumptions have emerged from these planning discussions:

- All libraries in North America will start to use Pinyin on Day 1 (October 1, 2000) in creating cataloging and acquisitions records for Chinese materials.

- The use of Pinyin in cataloging will be implemented in phases, with the creation of new bibliographic and authority records in Pinyin to begin on Day 1, followed by copy cataloging in Pinyin sometime within the following six months as more and more cataloging copy shows up in Pinyin. Each library will decide on its own starting date for copy cataloging in Pinyin (Day 2) which should be no later than April 1, 2001. This provides an interim period wherein a library can continue to do copy cataloging of Wade-Giles records in RLIN or OCLC while those databases are being converted.

- Conversion of Wade-Giles Chinese records in the RLG and OCLC union catalogs is expected to be completed by April 1, 2001. Conversion of Wade-Giles Chinese records in individual libraries’ OPACs is expected to be completed by October 1, 2001, one year after the initial switch to the Pinyin standard.

- Conversion will be across the board, including all Chinese records in the bibliographic utilities and in local files.

- Conversion will change current Wade-Giles headings to Pinyin in non-Chinese records.

- Most libraries will rely on national utilities to do the conversion. National utilities will decide on a pricing structure for their conversion services.

- The conversion of NARs by OCLC will precede, by a couple of months, the conversion of bibliographic records by RLG. All NARs that need to be changed will be converted before October 1, 2000.

- During the conversion period (Oct. 1, 2000-Oct. 1, 2001), a Pinyin marker (987 field in bibliographic records and 008/07 field in NARs) will be used in bibliographic records and name authority records that are created in Pinyin, converted to Pinyin, or marked for further review.

- Different libraries will convert their records at different times, with LC being the first to convert all its records. RLG will next convert libraries with large Chinese collections, while OCLC will convert newest records first and work backwards.

- During the beginning phase of conversion, RLG will focus on conversion of LC bibliographic records while OCLC will focus on NARs. OCLC will initiate the conversion of bibliographic records in WorldCat in October 2000 as well.
• The use of Pinyin and related conversion projects will follow LC’s new Chinese romanization guidelines as well as LC’s new cataloging policies and classification schedules.

• According to the coordinated timeline of the Pinyin Conversion Project (http://lcweb.loc.gov/catdir/pinyin/timeline.html), LC’s converted Pinyin records will begin to appear in the bibliographic utilities prior to Day 1 (October 1, 2000). While accepting Pinyin records created by the Library of Congress, libraries are encouraged not to adopt the Pinyin system in either original or copy cataloging of Chinese materials prior to October 1, 2000. It is also recommended that libraries not use the converted Pinyin NARs in bibliographic records prior to October 1, with the exception of the Chinese conventional place names that have already been established in Pinyin by a Library of Congress special project in 1999.

• The conversion specifications developed by the Library of Congress in consultation with RLG and OCLC will be made available to all libraries prior to October 1, 2000.

• OCLC and RLG will use the same specifications developed by LC so that the resulting conversion of the authority and bibliographic records will be the same. Due to differing database and record structures, however, the conversion programs and policies may differ in some respects. Libraries should consult the national utilities regarding any concerns related to the conversion specifications and their implementation.

3. A Recommended Timeline for CEAL libraries

January 1, 2000

- LC and national utilities decide on Pinyin marker
- OCLC and RLG define conversion options and services

April 1, 2000

- CEAL libraries deliberate on budget and conversion options
- CEAL libraries begin internal preparation for Day 1

October 1, 2000

- Day 1 occurs: official starting date for Pinyin
- LC uses Pinyin romanization for all cataloging
- CEAL libraries use Pinyin romanization for all original cataloging, new acquisitions and in-process records
- Converted LC bibliographic records have been distributed
- NARs have been converted by OCLC and distributed by LC
April 1, 2001

- Conversion of OCLC and RLG files has been completed
- Day 2 has been reached by all libraries: all cataloging is done in Pinyin. (All libraries have started doing all copy cataloging in Pinyin). Libraries can implement Day 2 earlier but not later than April 1, 2001, depending on their own situations
- OCLC and RLG are ready to release snapshots for individual libraries. (May happen sooner than this.)
- Individual libraries convert local machine catalogs to Pinyin (some libraries will start conversion earlier than this date)
- Post conversion clean-up work begins at national utilities and individual libraries whose OPACs are converted, though for some libraries, this may begin much earlier

October 1, 2001

- Day 3 occurs: all libraries cease using pinyin markers
- Conversion of individual libraries’ OPACs is completed

II. Options for CEAL Libraries and Analyses of Local Conversion Needs

1. Conversion options utilizing national bibliographic utilities

CEAL Libraries have the following options for conversion by the national utilities:

OCLC Services

Option 1: Conversion Based on the Library’s Local Database. Under this option, an individual library sends a file of MARC records from its local system to OCLC for conversion. Conversion may be limited to specific fields in the bibliographic records, but a final decision on this had not been made by OCLC as of this writing. Library then replaces its existing records with the converted local records returned by OCLC.

Option 2: Conversion Based on the Library’s Archive Records. Under this option, OCLC creates a file of a library's archive records and converts them to Pinyin. (Note that any editing done in a library's local system will not be reflected.) Library then replaces its current records with the converted archive records supplied by OCLC.

Option 3: Delivery of New Copies of Converted Master Records. Under this option, OCLC delivers copies of converted master records to which a library's holdings symbol is attached. (Editing done during previous uses of the record or in a library's local system would not be included.) A library then replaces its current records with OCLC’s Pinyin-converted master records.
Authority Records: At the time of conversion, OCLC can optionally provide a copy of converted National Authority File records associated with headings in these bibliographic records but will not convert authority records extracted from a library’s own database.

Batch loaded Records: Batch loading software will be modified so that incoming records are converted.

RLG Services

Conversion of Records in RLG’s Union Catalog (RLIN): All libraries with Chinese language records in the RLG union catalog will be converted in the October 2000 to April 2001 timeframe. RLG will first convert clusters that contain LC records followed by the conversion of clusters containing records of individual libraries, beginning with the largest Chinese collections. A library can order a snapshot of its converted records as soon as conversion of all its records is completed. RLG will add a "Current Pinyin Conversion Status" page to the RLG Web site indicating which libraries’ records have been completely converted and which ones are in process.

Conversion of Batch loaded Records: RLG will convert Chinese language records in Wade-Giles that are batch loaded after October 1, 2000, providing the source library identifies that a file contains Chinese language records requiring conversion. The library can request a copy of these converted records after they are loaded.

Libraries will handle post-conversion catalog maintenance and clean-up by themselves. Although the bibliographic utilities can provide some help in catalog maintenance, the responsibility for post-conversion maintenance will be wholly the responsibility of each individual library. Retrospective conversion projects already underway may continue to be in Wade-Giles, as both OCLC and RLG can convert these records into Pinyin after Day 2 as part of their batch loading programs. Therefore, libraries need not be concerned about not being able to complete retrospective conversion before Day 1. OCLC will modify batch load software to convert incoming bibliographic records and modify other services such as their "Authority Control Suite" and "Bibliographic Record Notification" to reflect the needs of Pinyin conversion. RLG will provide files of all changed headings sorted by the frequency of their appearance in bibliographic records to guide updates for authority records.

2. Name authority records (NARs)

OCLC will complete conversion of NARs before Day 1. As part of the Pinyin conversion programs, LC will compile a data dictionary of headings that should not convert. If there is sufficient interest, LC’s Cataloging Distribution Service will distribute a file of converted NARs. Converted NARs will also be included in the daily NACO distribution.

CEAL libraries should feel free to use the Chinese conventional place names that have already been established in Pinyin in the National Authority File, even in Wade-
Giles records, as these forms are specifically accounted for in conversion programs. Among Chinese conventional place names established by the Library of Congress in Pinyin, two name headings have been identified as being susceptible to double conversion. These are: Teng Xian (Shandong Sheng, China) and Pi Xian (Jiangsu Sheng, China) (Chinese: 腾县 and 邳县). These headings should be double-checked after bibliographic records and NARs are converted.

Due to the conversion timeline, LC’s bibliographic records and NARs will be converted before Day 1. Double conversion of NARs can best be avoided by taking great care in how one changes Wade-Giles records after LC’s converted headings begin to appear in the name authority file. It is advisable that CEAL libraries not use LC converted personal and corporate name headings in Wade-Giles records until such records are converted to Pinyin. This way, the risk of double conversion can be reduced. It will be perfectly safe to use LC’s converted headings in bibliographic records that include the Pinyin marker (field 987) coded to indicate that the record was either created in or converted to Pinyin.

Currently, LC is formulating plans for a moratorium on creating and changing authority records with Wade-Giles romanization. LC will issue guidelines for NACO/BIBCO libraries to help them minimize the risk of double conversion.

3. Pinyin marker

As CEAL libraries begin to catalog in Pinyin, they will implement a Pinyin marker in the 987 field of bibliographic records and in the 008/07 field of name authority records. Records converted by the bibliographic utilities will also include the Pinyin marker fields to indicate their conversion status. (See the instructions on the Pinyin marker in item 6, Appendix.)

4. Other prerequisites for Day 1

LC will complete changes in subject headings and classification schedules. LC and the utilities will conduct thorough tests of the conversion specifications for accuracy and will notify libraries of the final specifications.

5. Split files

While CEAL libraries will create all new records in Pinyin after Day 1, there will be a period (estimated to last no more than six months) during which the bibliographic utilities will contain a mixture of Wade-Giles and Pinyin records. During this period, to ease the work of copy cataloging, libraries may choose to accept a mix of records in Pinyin and Wade-Giles. In addition, libraries will have to keep their Wade-Giles records prior to total conversion. After Day 1, libraries should immediately begin to prepare for Day 2, the date when they will do all cataloging in Pinyin. In this period of split files, before conversion of a library's records is complete, libraries should make sure that when a Pinyin record is created or adopted that a Pinyin marker is properly inserted. It is
noteworthy that currently there are already records with Pinyin headings in the national and local databases without a Pinyin marker attached. Such occurrences should be minimized. OCLC and RLG should complete conversion of their entire databases of Chinese records by April 1, 2001.

6. Conversion of non-Chinese records and non-standard Chinese romanization forms

OCLC will scan its entire database for Wade-Giles headings in non-Chinese records. Wade-Giles name headings in non-Chinese language records will be converted to Pinyin by OCLC after Chinese bibliographic records are converted. The changed records will be distributed to member libraries if they have chosen conversion option 3 described above. Non-standard Chinese romanization forms will be ignored during this database scan. Non-standard romanization forms will be converted manually if necessary.

RLG may schedule the conversion of Wade-Giles strings in Japanese and Korean language records and in records with “Chinese” listed in the 041 field after completing the conversion of Chinese language records in the RLG union catalog, but this is not part of the 2000-2001 project currently underway.

In CEAL member libraries, the conversion of name headings from Wade-Giles to Pinyin in non-Chinese records will be mostly a local task. Libraries using OCLC services should make certain that changed non-Chinese records distributed by OCLC properly replace equivalent local records in their own databases. Libraries using RLG services are encouraged to contact RLG regarding how to convert their non-Chinese records, as this is not part of RLG’s announced 2000-2001 conversion project.

7. Implementation of Pinyin conversion by CEAL libraries

Libraries should start to plan the implementation of Pinyin conversion immediately. This should include deliberations on budget implications, conversion options, systems implications, the use of the Pinyin markers and Name Authority Records, cataloging workflow, and procedures such as the use of new cataloging schedules.

It is necessary that each CEAL library request its parent library to set up a taskforce with representatives from the Chinese collection and cataloging units, the library’s central cataloging department, information systems and other relevant personnel. It is also imperative that CEAL libraries begin to communicate with the national utilities regarding their conversion services.

Staff training and user education are also critical to the success of Pinyin conversion. Cataloging and acquisitions staff involved in the processing of Chinese language materials need to be trained in the new Pinyin romanization scheme and in LC’s new subject headings and classification schedules.
8. Implications for Public Services

Library users need to be informed of the conversion and to be provided with proper search guides. Special attention should be given to word division in Pinyin, as the standard being implemented differs somewhat from what may be familiar to users. User education is especially critical during the period when two romanization forms co-exist in catalogs and networks. Efforts should be made to direct users in how to search materials in the split files between October 1, 2000 and October 1, 2001. Libraries need to prepare proper users' guides and other relevant handouts to assist users in their search of the converted local OPAC. Pinyin conversion will also necessitate the re-labeling of current Chinese periodicals and their back files if such materials are shelved alphabetically by title. Such serials re-labeling projects should be coordinated with the conversion of the associated records to Pinyin.
Appendix: Major Pinyin conversion tools and documentation

1. LC’s Pinyin conversion timeline
   http://lcweb.loc.gov/catdir/pinyin/timeline.html

2. LC’s New Chinese Romanization guidelines
   http://lcweb.loc.gov/catdir/pinyin/romcover.html

3. Classification schedules: Chinese literary authors
   http://lcweb.loc.gov/catdir/pinyin/authors1949.html
   http://lcweb.loc.gov/catdir/pinyin/authors2001.html

4. Classification schedule: changes to Chinese conventional place names
   http://lcweb.loc.gov/catdir/pinyin/class6.html

5. Conventional Chinese place names
   http://lcweb.loc.gov/catdir/pinyin/placefaq.html

6. Pinyin markers
   http://lcweb.loc.gov/marc/pinyin.html
   http://lcweb.loc.gov/catdir/pinyin/authorities.html

7. LC’s announcement on Pinyin conversion project
   http://lcweb.loc.gov/catdir/pinyin/announce.html

Works Consulted

1 Cataloging service ; bulletin no. 42.


5 “Summary report of the CEAL task force to review a possible change from the Wade-Giles to the Pinyin romanization system” In Journal of East Asian Libraries, no. 115, June 1998, p.40-44.