Summary Report on Pinyin Conversion

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/vol1999/iss119/11

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.
Summary Report on Pinyin Conversion

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/vol1999/iss119/11
SUMMARY REPORT ON PINYIN CONVERSION

CEAL Pinyin Liaison Group
July, 1999

Table of Contents

I. Work of CEAL Pinyin Liaison Group
II. RLG Forum on Pinyin Conversion in New Orleans
III. LC Pinyin Conversion Plans
IV. OCLC Pinyin Conversion Plans
V. RLG Pinyin Conversion Plans
VI. Issues and Impact on CEAL Libraries

CEAL Pinyin Liaison Group members:

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

I. Work of CEAL Pinyin Liaison Group

In May 1998, the then CEAL President Tai-loi Ma appointed a Pinyin Liaison Group to represent CEAL in deliberations with LC, RLG and OCLC in matters related to the forthcoming conversion of Chinese language records from Wade-Giles to Pinyin romanization. In April 1999, CEAL Acting President William B. McClory appointed additional members to this Group. The Group also serves as a conduit for communication among CEAL libraries on Pinyin conversion.

The Group decided to issue brief summary reports when warranted to inform CEAL libraries of developments in Pinyin conversion. With Day One of conversion only several months away, it is imperative that CEAL libraries start preparing for this event. In the meanwhile, the Group is studying various Pinyin conversion issues closely and is preparing a final report to CEAL Executive Committee later this year.

Enclosed in this summary report are brief discussions of recent developments in Pinyin conversion planning and of issues that need further study. We chose not to go into detailed analysis of those matters here. Such analyses will be presented in our final report.

II. RLG Forum on Pinyin Conversion in New Orleans

On June 27, 1999 during the ALA annual convention in New Orleans, RLG held a forum to discuss Pinyin conversion. Representatives from RLG, LC, OCLC, and senior library
administrators from Columbia, Cornell, Harvard, Princeton, UC-Berkeley, University of Chicago, University of Michigan and Yale University attended the meeting. Five members of the CEAL Pinyin Liaison Group also attended the meeting. These people conducted a panel discussion on various issues and ramifications of Pinyin conversion before an audience of more than one hundred people.

During this meeting, LC, RLG and OCLC individually introduced their plans for Pinyin conversion. The CEAL Pinyin Liaison Group presented a checklist of issues and ramifications. Panelists and the audience then engaged in a lively discussion on this change and its impact on library databases and cataloging work. The complete meeting minutes and handouts of this RLG Forum will be posted on the RLG website later this year.

III. LC Pinyin Conversion Plans

LC’s Chinese monographic records (fixed field language=chi,USMARC 008/35-37) will be converted by RLIN. LC’s Chinese serial records will be converted on OCLC. Wade-Giles headings on non-Chinese monographic and serial records will be converted to Pinyin as well. LC anticipates that certain subfields may have to be put aside for manual conversion. LC will review records not converted and make changes if necessary. In addition, 67,000 pre-MARC records will be loaded into the DLC local file.

October 1, 1999 is Day One for conversion of subject headings. LC will start using Pinyin romanization in new subject headings from then on. Manual conversion will start on October 1 for subject authority records with headings in Wade-Giles forms. LC hopes that it will be possible to convert some NARs by machine. It will probably be necessary to convert others, such as non-unique names, manually before and during the bib conversion. Insofar as possible, NARs with headings in Wade-Giles forms that are not represented by access points on Chinese bib records will be identified and converted.

LC began to make changes in Chinese conventional place names in August 1998. More than 260 NARs for geographical locations have been changed to current Pinyin forms. By mid-July, the NARs related to those geographic headings (such as subordinate bodies) will also have been changed. The total number of changed NARs will amount to more than 4,500. The corresponding headings on bib records have not yet been changed.

LC will start to convert its bib records in Spring 2000.

IV. OCLC Pinyin Conversion Plans

OCLC is actively working with LC and RLG to review draft specifications of the conversion and to coordinate the implementation timing with LC and RLG. OCLC plans to convert all 756,000 Chinese records in Worldcat, including CONSER Chinese serial records and will explore ways to identify non-Chinese records for conversion. For authority records, OCLC has a proposal to identify related authority records and modify them (by using WLN’s authority management software). The identification of related authority records may assist in finding Wade-Giles segments in non-Chinese bibliographic records for conversion.
OCLC will mark all records after conversion and modify batch-load software to convert incoming bib records. OCLC will modify other services to offer different conversion options such as Authority Control Suite, Bibliographic Record Notification, and the Local Database Creation.

V. RLG Pinyin Conversion Plans

RLG will convert all Chinese language bibliographic records in the RLG union catalog (estimated 2.5 million records) including those with or without CJK scripts according to LC’s functional requirements. RLG will provide files of all changed headings (estimated 1.0+ million headings with 25,000 of ten or more appearances in RLG union catalog), sorted by frequency of their appearance in bibliographic records, to guide updates needed for authority records. It will also provide snapshots of all converted bibliographic records to the records’ owning libraries with an option to export the Chinese character aggregator as a defined character.

RLG will provide snapshots for each library’s own Chinese-language RLIN records after RLG has converted them to Pinyin romanization. RLG members will receive discounts; non-members will pay the standard rate. RLG will incorporate conversion into data loading programs for incoming files still in Wade-Giles after the RLG union catalog conversion is completed.

RLG has based its project on a number of planning assumptions. First, Day One would represent a clean break: that is, all bibliographic records created before Day One are in Wade-Giles, all new records after it are in Pinyin. Only Wade-Giles strings would be converted, and there will be only one conversion program to be used on all records. Conversion will be done cluster by cluster, rather than library by library, with the exception of LC’s records. All variable fields in the bibliographic records will be converted unless otherwise specified. Updated authority records will be distributed and incorporated into libraries’ local catalogs (through linked authorities or authority control); it was assumed that it would be undesirable to have a program automatically change authority records for headings that can affect thousands of bibliographic records. Non-Chinese bibliographic records with changed headings would be contributed back to RLG union catalog as part of ongoing record.

There will be a period of "mixed" Wade-Giles and Pinyin romanization in the RLG union catalog for some period of time. RLG has completed a review of LC’s initial functional requirements. It is implementing new programming tools for conversion, and expects to run the first test run against LC sample Chinese records in September 1999. By the end of 1999, it will complete a series of test runs incorporating LC revisions until LC signs off on the conversion program. In the beginning of 2000, RLG will first run the conversion program against test records from several RLG member libraries. Then, in Spring 2000, full conversion of Chinese-language bibliographic records in RLG union catalog will start, beginning with LC records. It is expected that by September 2000, RLG’s bibliographic record conversion will be completed. Then, RLG will provide snapshots to owning libraries as requested. RLG will not know exactly how long the bibliographic conversion will take until it conducts the first production runs.
VI. Issues and impact on CEAL libraries

The following are some of the major issues identified in Pinyin conversion, among many issues that have been brought up in the RLG Forum and in other discussions during the past several months. We also included this Group’s positions on some of those issues in the following discussions.

a) LC Pinyin romanization guidelines, Chinese conventional place names and retained Wade Giles personal headings: LC needs to expand its Pinyin romanization guidelines on geographic name headings not specified by BGN and on retained Wade-Giles forms after Day One. A mixture of both Pinyin and Wade-Giles headings will exist in the databases after conversion. A policy on how to romanize geographic names of Taiwan and Macao should be added to the guidelines.

b) Both bib and authority records need to be marked on the field and record levels so that they will not be converted more than once (to avoid such erroneous multiple conversions as tu->tu->du). Bibliographic records that have been converted should be flagged by an internal field to assure a one-time conversion only. This might pose a problem for authority records that may need to have multiple machine upgrading.

c) Individual libraries should have input on what fields in the record to convert or not to convert. We hope that LC and RLG could share conversion specs with CEAL libraries and involve them in such discussion and testing. We agree that the final decision on conversion specifications remains with LC.

d) Each CEAL library will set its own Day One for Pinyin conversion. The question is how long it will take for RLG to complete the snapshot by cluster. How do the national utilities deal with the clean-up job of Wade-Giles records created after Day One? RLG has urged each library to commit to the same Day One for its own records the day that both description and access points for all new Chinese-language records created are in Pinyin romanization. We learned that RLG is willing to include in the MARC records it exports an internal field indicating that the records have been converted to Pinyin romanization.

e) We believe authority records should be converted as early as possible before Day One or at least at the same time with the conversion of bib records. There will be a gap in the conversion of these two groups of records. Such a gap in conversion will impact local library’s copy cataloging policy and practice. Libraries will have to deal with mixed subject and name headings in both Pinyin and Wade-Giles.

f) What should be done before Day One? What is the "spread" of "Day One" dates? Will NARs be converted through coordination among NACO libraries? The conversion specs should be the same for every library. Day One should not occur before the marker field has been decided.

g) Conversion has to deal with Wade-Giles forms in non-Chinese records and non-standard Chinese romanization in name authority files. Conversion also has to take into consideration that the ultimate goal is to achieve a complete retrospective conversion of all Wade-Giles records.
including those in card catalogs of some libraries. This would mean that before total conversion, some libraries, if not all, will have to deal with split files for a long time.

h) Conversion plans should include Music/AMC records where no language code is present, but Wade-Giles forms exist. This is also true of finding aids, metadata and other source files that are still in Wade-Giles. How will this conversion impact western language cataloging such as dissertations?

i) Conversion should ideally be synchronized among RLG, OCLC and LC. Conversion of authority records and bib records also should be synchronized. As indicated by participants at the RLG Forum, local support for Pinyin conversion will likely be minimal. Most libraries perhaps would like national utilities to do as much as possible and offer an affordable price for their conversion service.

j) Regarding consortium level planning, we noted that currently both RLG and OCLC libraries are engaged in communication within each group through EAMEMNET and OCLC CJK Users Group. University of California System has formed an advisory group on Pinyin conversion. A survey done by the CEAL Pinyin Liaison Group in February 1999 indicated that most libraries are not well-prepared for Pinyin conversion. Therefore, more consortium level planning and cooperation would be needed to cope with this change.

There are a number of unanswered questions: How long will the conversion of name authority files take? How do Chinese cataloging staff deal with the ever-changing geographic names and subject headings in their daily copy cataloging? Should individual libraries wait until after RLG, OCLC and LC have converted all their bib files before they start using Pinyin in new records they create? Or should they start to use Pinyin while conversion by LC and national utilities is in process? For LC, there will not be a Day One, but many Day Ones in different phases of the conversion with different files. How should individual libraries decide on their own timeline under such a fluid situation? While a one time conversion is ideal to everyone, it would seem that some bibliographic records will already have some access points in Pinyin while the rest of the record will be in Wade-Giles. How will libraries handle this? The current LC and RLG conversion plans could create a situation under which certain access points of a record could be changed and updated multiple times, although conversion of bib records will be done only once. How can this conflict be resolved?

The CEAL Pinyin Liaison Group will continue to monitor and engage in discussion, planning and implementation of Pinyin conversion. We will monitor cataloging policy changes related to Pinyin conversion and be engaged with LC, RLG and OCLC on Day One decision, system specs, and NAR conversion. We will also pursue answers to questions raised in this report. We are very pleased to see LC’s recent modification in its planning to start converting NARs before bib records, which addresses some of the concerns of the CEAL community. We hope LC will come up with a more definitive timeline for the conversion of name authority files. We urge CEAL libraries with staff resources to conduct NACO collaboration projects with LC in order to have as many NARs converted as possible before bib conversion. We also plan to assess the impact of Pinyin conversion on public services at East Asian Libraries and at general reference service departments in our future report.