Addendum

Committee on East Asian Libraries

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

BYU ScholarsArchive Citation

Available at: https://scholarsarchive.byu.edu/jeal/vol1990/iss90/13

This Article is brought to you for free and open access by the 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.
Preliminary statements

1. The Library of Congress has been approached by many others who suggest that our romanization of Chinese should be shifted from Wade-Giles to pinyin, so that our records would be in harmony with the clearly prevalent use of pinyin. We know that all other U.S. Federal agencies use pinyin (including the Board on Geographic Names) and that on the international scene our records for Chinese publications are incompatible with those produced by European libraries (as is well known, the British Library has used pinyin since 1958). With, in addition, the use of pinyin by all news agencies, national and international, the Library of Congress and the North American library community are now isolated in their continued use of Wade-Giles. Spontaneously then, discussion has begun once again on the question of whether we also should switch to pinyin.

We realize that an essential first step is consultation with the Committee on East Asian Libraries, and the present meeting is our first opportunity to do so. It should be clear then that we have not made a decision to switch to pinyin. Instead, we have concentrated our attention on looking into the possibilities for dealing with the switch from Wade-Giles if a decision in favor of pinyin is made.

2. This document attempts to point to the problem we would have to face if indeed this decision is made, and to suggest the various approaches for dealing with it. The problem is unfortunately a large one due to its timing. In 1981, of course, we had a special opportunity to switch to pinyin since we had virtually no MARC records for Chinese at that point. All the MARC records created since then show Wade-Giles romanization in

- headings for personal names, corporate bodies, geographic names, uniform titles, and series;
- bibliographic description (titles, statements of responsibility, edition statements, places of publication, publisher' names, series statements, notes such as contents);
- some subject headings, aspects of classification and cutting, and the Dewey Decimal Classification, which includes headings in AACR 2 form.

There are now several hundred thousand MARC records in Chinese with tens of thousands of new records added annually, considering the files of others as well as our own.

3. Theoretically, there are various solutions to this problem, depending not only on resources in cataloging, but also on the means (1) to identify the headings that need to be switched,
and (2) to make the switch in both authority and bibliographic files. In the remainder of the paper we lay out for consideration the possibilities we see in trying first to identify the records and then to make the intended switch. In this first consideration we concentrate on automation. Then we turn to cataloging, in the gloom of limited resources that is the unfortunate reality.

AUTOMATION

The ideal approach to the problem of switching to pinyin would be to convert the retrospective files through the use of computers. The Library of Congress does not currently have a straightforward, cost-effective method for doing so. It is possible that one of the utilities may already have or will develop such a capability from which all of us could then benefit.

With respect to automation, some of the issues to consider are:

1. It would probably be impossible automatically and with sufficient accuracy to convert the bibliographic description to pinyin using computers.
   - It would be practically impossible to locate all the Wade-Giles romanizations since they may occur in any field in the record and may be intermixed with other languages.
   - Even if the romanizations could be located, there is currently no computerized global change facility that would allow us to change to pinyin. Either such a facility would have to be developed or the records would have to be changed manually.

2. It is somewhat more feasible to use the computer to convert headings in bibliographic records to pinyin, but even here a whole range of problems is evident and practical issues will probably preclude a global machine conversion.
   - Again there is the problem of identifying the Wade-Giles headings. Since no language or romanization code is carried at the field level in the records, the headings would have to be located through a text matching process.
   - For example, one option for identifying the headings would be to attempt to match the first three or four words of each heading against a list of Wade-Giles words. This would provide candidates for machine modification, but even this method is not precise: it is possible that a Wade-Giles heading could be entered subordinately under a conventional English language form; non-Chinese words can be present in Chinese headings; and some Wade-Giles words are also words in English and other languages, so false drops might occur.
   - If the Wade-Giles headings could be identified using the above method, then a computerized conversion to pinyin could take place. At present, however, there is no computerized global change capability to use for this, so one would have to be developed (or the records changed manually).

3. One approach would be to convert the authority file to pinyin and then to use that file for a machine conversion of the bibliographic files. Here, once again, we are presented with a list of problems which demonstrate the difficulty of implementing this option.
Although we have been supplying pinyin romanizations in cross references in our name authority records for over ten years now, the MARC authorities format does not provide a code that would enable us to locate easily the pinyin forms in order to "flip" them by computer with the Wade-Giles form.

In this case, then, a word matching algorithm would be used against the some two million cross references in our authority file in an attempt to locate the pinyin forms. The same problems and false drops in word matching mentioned above would probably apply here.

If the pinyin cross references could be located by computer, they could then be switched with the Wade-Giles forms in the authorized-heading fields. New software development would have to be undertaken for this or, failing that, they could be flipped manually.

After the authority file had been flipped, it would be compared against the bibliographic files in order to convert headings to pinyin. At present the Library of Congress has no computerized global change capability to use for this, so the work would have to be done by one of the utilities or else done manually, one record at a time. (LC's software development staff are concentrating on developing a new system for LC, so it is unlikely that they would be free to develop these new programs in the near future.)

After examining these issues and related ones in some detail, we have concluded that it is unlikely that automation would be able to play a major role in the switch to pinyin if that switch is undertaken in the immediate future. Automation would instead play a supporting role, assisting catalogers in their work wherever possible.

CATALOGING

The following scenarios outline possible approaches to cataloging for implementing pinyin romanization, in the atmosphere of a first consideration. It is hoped that they may be useful as a starting point in determining how cataloging might face the problem of shifting from Wade-Giles.

There are disadvantages common to all the approaches we have thought of, as well as advantages and disadvantages specific to each particular scenario. The common disadvantages seem to be:

1. It will be necessary for catalogers and searchers in library catalogs to know two romanization schemes. Arranging data in one alphabet which were created using the two different romanization schemes can cause confusion, and searchers would have to look under both Wade-Giles and pinyin forms to insure complete coverage. Even if structured headings are kept under one romanization, searchers would have to take into account the possibly confusing arrangements that would be produced when bibliographic titles romanized using Wade-Giles are arranged with bibliographic titles romanized using pinyin.

2. Expense; all of these scenarios will require extra work by catalogers, which, while diminishing over time, will reduce the number of new works that can be cataloged. Additionally, if automation is used to assist implementation, then there will also be costs in automation.
Scenario A. Superimposition

Retain existing Wade-Giles headings and LC classification and cutting for already established persons/bodies/uniform titles/numbers. Shift to pinyin only in the case of totally new entities.

Advantages
1. This is probably the cheapest solution in terms of immediate use of personpower and automation, since it would not require additional authority work or changes to existing bibliographic records.
2. This approach keeps the records under a particular heading or number according to one form in the files, i.e., the Wade-Giles form whenever the heading was established prior to the switch.
3. For entities already established, this approach assures uniformity in LCSH and LC Classification and Shelflisting—according to Wade-Giles romanization.

Disadvantages
1. It would slow down the progress of building the pinyin file.
2. New subheadings of already established entities will have to stay with Wade-Giles in order to match the "parent" heading.
3. Catalogers (and other searchers) would have to use continually and actively the two romanization systems.
4. There would be a mixture of romanizations in the same bibliographic record, with the description in pinyin and some or all headings in Wade-Giles.

Scenario B. Change headings to pinyin

As a particular heading is needed, all the occurrences of the heading would be changed to pinyin. This could be handled in the same way as the change to AACR 2.

Advantages
1. The pinyin file would be built up quickly.
2. This approach keeps the records having a particular heading together, under one form in the files, the pinyin form.
3. This approach lends itself to assistance through contractors, projects, and/or automation; once a cataloger has changed the authority records and made the reference from the former Wade-Giles heading (specially identified through use of the 4xx subfield $w/2 value "e" or a newly defined value especially for Wade-Giles), bibliographic records could then be "flipped" and redistributed at a later date.
4. Some frequently used headings could be identified by retriever runs and handled by special projects.
5. Most catalogers and reference staff grew accustomed to the AACR 2 change, so if similar procedures could be worked out, the routine could become familiar.

Disadvantages
1. The authority and bibliographic files might be out of sync for a long time if changes to the bibliographic files were made through a "flip" at a later date.
2. A single bibliographic record might have headings in different romanizations as headings for the various entries on the record are changed one-by-one from Wade-Giles to pinyin.
3. There would be a mixture of romanizations on many older bibliographic records, with the bibliographic description in Wade-Giles and some or all headings changed to pinyin.
4. If the AACR 2 procedures are followed, a bibliographic record might have to be handled and distributed many times, as each heading on it is changed from Wade-Giles to pinyin.
5. It will not be easily predictable for a user to know under which romanization to search.
6. Under this scenario there would no longer be a uniformity with the LCSH, LC Classification, and Shefflisting—as long as these retain Wade-Giles romanization.

Scenario C. Split files

As of a particular date, we would stop using Wade-Giles and begin to use pinyin for all newly cataloged items. Wade-Giles headings already present in the catalog would not be changed to their pinyin equivalents, but would be linked, if at all, by the authority files.

Advantages
1. There would be no mixture of romanizations on newly created records. (Of course some headings on older records would come up for changes due to some reason other than romanization, causing the question of changing the romanization at the same time to arise.)
2. There would be a predictable date for the change to pinyin, though that date might not correspond to the date of publication of the item searched.
3. This approach would not require much change to existing bibliographic records.
4. The procedures would be simpler for catalogers.

Disadvantages
1. This approach would require a lot of authority work. A new authority record would have to be created for the pinyin form (at the Library of Congress, the use of a proposed copy command might help), and the pinyin 4xx reference on the existing Wade-Giles authority record would have to be deleted or changed to a 5xx. The double reference structure (for the Wade-Giles and pinyin headings) would create confusion.
2. Records for one entity might be in two places in the file.
3. There would be general confusion for catalog users.
4. Techniques would have to be developed to provide for linkages and for relationships between editions or among various related works cataloged according to the two systems of romanization.
5. Squaring the two systems with the existing single system (Wade-Giles) found in the LCSH, LC Classification, and Shefflisting would be problematic.

Scenario D. Separate files for Wade-Giles and pinyin

Establish three separate new MUMS files for Wade-Giles records: Books, Serials, and Authorities. These Wade-Giles files could be searched separately or together with other files through the use of file qualifiers. Most bibliographic records to be moved to these files could be identified by retriever runs on the language code and the suffix/alphabetic identifier that follows the Library of Congress Control Number. The new Wade-Giles files would be largely frozen, though some records, e.g., for multipart items, could be recataloged using pinyin and then moved to the pinyin file, if needed because of the acquisition of new volumes. Newly cataloged records would be done in pinyin. As the cataloger used an existing authority record, he or she would use the proposed copy command to send an unaltered version to the Wade-Giles authority file, where it would retain the same control number, but with the prefix "wg" instead of "n," "no," or "nr." (The 4xx from the pinyin form could be deleted first, in the case of conflicts.) Then the cataloger would adjust the authority record in the Names file to pinyin.

Advantages
1. There would be no need to change headings on the Wade-Giles records.
2. There would be no mixture of romanizations on a single record.
3. There would be no mixture of romanizations of headings or bibliographic titles, unless the searcher intentionally searched the Wade-Giles and pinyin files together.
4. Both files would be under authority control, with useful references retained for Wade-Giles users.

**Disadvantages**
1. It would be necessary to search the Wade-Giles files as well as the regular files for complete coverage, though gradually the Wade-Giles files would become historical, rather like PREMARC.
2. It would not be possible to move all records containing Wade-Giles romanization to the new files: for example, publications in English with Wade-Giles headings for Chinese names would not be moved, nor would records in the Music, Maps, or Visual Materials files.
3. It is highly unlikely that automation, as practiced by LC (and probably by the utilities), could implement and support such an approach, either now or any time soon.
4. Accommodating the LCSH, LC Classification, and Shelflisting to the two systems of romanization and to the multiple files would be highly problematic.
5. It would add another level of complexity both to cooperative cataloging and to record distribution.

**Scenario E. Dual data in records**

Following this approach would mean including within each Chinese record both systems of romanization for some data, or perhaps for all data.

**Advantage**
Some version of this scenario would presumably result in a nearly ideal situation: we would not need to be concerned about the user's knowledge of a particular system of romanization.

**Disadvantages**
1. It might be necessary to change the MARC format to accommodate such dual data.
2. Costs would be considerable, both in cataloging and, initially, in automation.

**Scenario F. Dual records**

Following this approach would mean creating two records for each item in Chinese cataloged, one showing the romanized data in Wade-Giles form and the other showing pinyin.

**Advantages**
1. As with the preceding scenario, library users could approach the catalog with either romanization in mind.
2. "Duplicate" records are a tolerable consequence, since in large data bases users have been finding duplicates for a long time.

**Disadvantages**
1. The double records would have to be introduced to library users at the beginning, with a view to having them understand that (1) the records are not the usual duplication by catalogers that is heretofore not deliberate, and (2) the record they retrieve first may not be the record they really want.
2. While duplicate records are not necessarily a problem in large multi-library data bases, they would be a problem in the catalogs of individual libraries that derive from these data bases.
3. This approach would cost much in terms of cataloging resources; it would in fact greatly increase the cost of creating Chinese records, unless/until we could develop a machine translation between Wade-Giles and pinyin.
Scenario G. Interim proposal

1. The Library of Congress would work with the East Asian library community, OCLC, and RLIN in support of an eventual shift to pinyin.

2. Until we and our automated systems have come to the point of making the shift, we would provide, in addition to the pinyin cross references included since 1980 in authority records for most headings, an additional title added entry in bibliographic records giving the title proper also in pinyin romanization.

LC Suggestion

Given all the above, we at the Library of Congress tend to favor following SCENARIO G as an interim policy until a final decision is made.

Reactions

The present document is intended for those who wish to consider the possibilities for a shift to pinyin. After studying its contents, please send any comments you have BY JULY 31, 1990, to

Lucia J. Rather  
Director for Cataloging  
Collections Services  
Library of Congress  
Washington, D.C. 20540