

## You Say Pei-ching, I Say Beijing: Should We Call the Whole Thing Off?

A presentation  
by  
Carol Hixson  
Head, Catalog Department  
University of Oregon Libraries

as part of the

ALCTS/CCS/CC:AAM Program, Year of the Dragon, Mandate to Change: Challenges of the Pinyin Conversion in the Online Environment

ALA Annual Conference, July 9, 2000

When this program was first being planned and my participation in it was announced, a colleague on the CCS Executive Committee turned to me and asked, “What on earth do YOU know about Pinyin conversion?” And well he might have asked, since, at the time, I knew very little and was content to let my more knowledgeable colleagues with responsibility for large East-Asian collections do all the planning and worrying. But I am here precisely **because** I represent the great masses of catalog department managers in small and medium sized American libraries who know next to nothing and may think that they need know nothing about the Pinyin conversion. I was invited to address you because I am **not** an area-studies specialist and because I manage a catalog department in a library that does not have a large East Asian collection - which is the reality for all but a small minority of libraries in the United States. I’m here to present the outsider’s point of view and discuss what everyone needs to be aware of and thinking about – even libraries who don’t own **any** Chinese-language materials. I apologize up front to all speakers of Chinese for my total ignorance of the Chinese language and for what will certainly be horrendous pronunciation. Please understand that it is not a lack of respect but rather, instead, a lack of ability and instruction.

I have been asked to address catalog management issues for the generalist library. My definition of catalog management may perhaps be broader than yours, so let me explain my approach. Managing the catalog involves the maintenance and cleanup of all types of data in the online catalog to ensure consistency and usability. The catalog exists for our public, primarily, but also for our internal users, be they collection development staff, reference librarians, catalogers, or acquisitions staff. The catalog consists of bibliographic, authority, holdings, and item information. So, when I speak of catalog management in an integrated, online environment, I will be touching on technical issues of how we accomplish certain things, staffing issues, impact on other departments, and public service issues. My definition of a generalist library in this context is a library that has small or non-existent collections of Chinese-language materials.

As everyone knows by now, Pinyin is a system for writing the Chinese language in the Latin alphabet. Wade-Giles is the system currently in use in American libraries, having been adopted by the Library of Congress in 1958. That same year, China promulgated the use of pinyin and the British Library started to use pinyin for the bibliographic control of Chinese-language materials that same year. Pinyin has been widely adopted by the international community as the standard Chinese romanization system. The Library of Congress first proposed making the change to

pinyin in 1980 but the American library community was not ready to address the effort at that time. Finally, in 1997, serious planning for the conversion in American libraries began.

In the past few months, I've conducted a very informal, unscientific survey of some cataloging colleagues in libraries that do not have large (or any) Chinese language collections and have asked them what their plans for the conversion are. Among those who were even aware of it, almost none of the libraries had even begun to discuss the implications, although my query prompted the first in-house discussions in some cases. Making this change seems like a monumental waste of time to some people. Does it really matter to a native speaker of Chinese who would prefer to be reading the Chinese characters themselves whether we represent in our Latin-alphabet catalogs the city as Beijing or Pei-ching? Sure, maybe it's a big deal for the libraries that have large collections of Chinese-language materials. But what about the rest of us? Should we just call the whole thing off and not worry about any of it? Maybe, in this one case, we can give up our infamous catalogers' obsession with consistency and just let the old entries remain while the new forms come in on new records. I've been told that Chinese-speaking scholars tend to avoid the catalog anyway and prefer just to go to their subject area in the stacks and browse. And what about those libraries that do not collect anything in Chinese? Do they need to worry about it at all?

The answer to those rather obnoxious questions is that it matters a great deal to our Chinese-language patrons, whether they be scholars who are able to go directly to selected areas of the stacks to find relevant materials or whether they be native speakers of Chinese using a public library in Eugene, Oregon. It is a system that most Chinese-speaking students and scholars are already familiar with, since it is the system used in mainland China and Taiwan when they must communicate using the Latin alphabet. It also matters a great deal to library users with no knowledge of Chinese since it is also the system that has been used in U.S. federal agencies and the mass media for years. Whether they be scholars or general library users, most users of American libraries today are familiar with Pinyin (whether they realize it or not) and have been hampered in their research by having to learn to use Wade-Giles to find materials in American library catalogs. Consider the high school student who has read in a newspaper or heard on a television broadcast about the writings of Mao Zedong and yet will find nothing in an American library unless he looks under Mao Tse-tung. How will he even know that he needs to look under a different form of name? And, if serious scholars have tended to avoid our catalogs it is just as likely to be because we've been using the "wrong" system of romanization for years as it is the fact that we've been romanizing at all. The thought that a lack of consistency is somehow more tolerable in this case because a library may be dealing with materials of interest only to a minority of their users should be abhorrent to the American library community. A mixed file representing headings in both forms of romanization will severely hamper anyone doing any level of research in any library.

To those of you who are from libraries that do not collect **any** Chinese-language materials who think that this does not affect you at all, think again. Even if you do not collect materials in Chinese, you most likely collect some materials about China or by Chinese authors in translation. What does the general library need to consider and plan for? I'm going to begin by discussing the issues that affect every library and then move on to some issues that have more of an impact on libraries that collect some Chinese-language materials.

**Analyze the scope of the problem in your library**

The first step for a any library is to analyze the scope of the problem. Even if you don't have many (or any) Chinese-language materials, the problem may be larger than you realize. Take a look at some of the wealth of information that is available on the Library of Congress' Pinyin Web site (<http://lcweb.loc.gov/catdir/pinyin/>) and familiarize yourself with the national timeline. Take a look at some of the planning documents that have been put together by RLG, OCLC, LC, and CEAL and see what the big players consider to be significant issues. Though the scale is smaller, many of the same issues still need to be addressed

What are the issues relating to pinyin conversion that touch every library? Chinese place names are being changed in subject and name headings; author cutters in LC classification are being changed; series and uniform titles are being changed; major changes are being planned in the DS, G, and PL schedules of LC classification.. If you accept records from the Library of Congress or other libraries with no or minimal review, you will be feeling the effects of these changes.

### **Geographic name headings**

In August 1998, the Library of Congress began the process of revising the Chinese place names currently established in a conventional English-language form to the form used by the United States Board on Geographic Names. The Library then began working on other names, primarily the provinces and the provincial capitals and other major cities. In general, they have made the change to the authority records for these headings, as well as to authority records whose entry element or qualifier uses the geographic name. Once the change to these headings is made, it is expected that all new cataloging will contain the new forms. These changes to authority records are being made before the Library converts from the Wade-Giles system of romanization for Chinese to pinyin because many of the forms used for these geographic headings cannot be converted easily by machine manipulation since they are not systematically Romanized from Wade-Giles. Others represent a policy decision that precludes automatic conversion. The CPSO has a list of the affected geographic names on its Web site at:

(<http://lcweb.loc.gov/catdir/cpsso/china.html>) detailing the date on which the change was made, as well as some information about related bibliographic records in which the change has been made in LC's file.

How many general libraries were aware of these changes and have been keeping up with them? I've searched academic and public library catalogs over the last few months and have found a mixture - some libraries seem to have been keeping up with the announced changes and updating records in their own databases and others have many occurrences of the former Wade-Giles geographic names still in their catalogs. At the University of Oregon, we have changed many of these headings, but not all. It requires checking each heading, cleaning it up if it's still in Wade-Giles, and bringing in the new authority records. If your library has not been keeping track of these changes and making corrections as you go along, chances are you have split files in your catalog. You may even have outdated authority records that are misdirecting your patrons away from the new, correct form to the older, now incorrect form. Is this something you can live with? Will you attempt to clean these headings up systematically or will you wait until a conflict enters your catalog from a new record entering your database? Is your online system set up to identify conflicting headings. In many cases, there will not be cross references even in outdated authority records that will enable you to identify a conflict, even if your system generates reports like this based on authority records. Though your holdings in this area may be small, the labor it

will take to find and correct those headings that you have used may be substantial. Will you wait, instead, until an entire file of changed authority records is made available and have that run against your entire database to find and correct old headings that way? Will this even be possible? LC will not be making a file of changed authorities available.

## **Classification**

Major changes are anticipated in the DS, G, and PL schedules of LC. For Chinese literary authors in the PL schedule, LC plans to end the most recent time period in 1999 (1949-1999). Beginning with the year 2000, cuttering will be based upon the new Pinyin system. The changes to Chinese conventional place names are reflected in the current schedules, although the cutters have not been changed if they had already been established in Wade-Giles. There is, instead, a cross reference from the new form of name to the previously-established cutter. How will your library handle the changes to cutters? Do you plan to revise your catalog records so that all cutters are current? If you don't revise the cutters, will the split in your online shelf list cause your catalogers and your users confusion? Will the use of two different cutters for the same place cause your users problems? Have you alerted your public services staff? If you plan to revise cutters and re-label pieces, does your administration understand the impact on your current cataloging productivity?

LC has posted changes already-made changes to the DS schedule on its Web site at (<http://lcweb.loc.gov/catdir/pinyin/class6.html>). Proposed changes to the PL schedules are also posted at (<http://lcweb.loc.gov/catdir/pinyin/authors1949.html>) and at (<http://lcweb.loc.gov/catdir/pinyin/authors2001.html>)

Recommendations for changes to the DS schedule have been sent to CPSO. After the changes have been approved, they will be posted on the pinyin home page; this should occur in advance of the ALA conference. Let me briefly explain the changes, and the rationale for making them.

There are currently two classes for history of Chinese localities: DS793 (Provinces, dependencies, regions, counties, etc.), and DS796 (Cities, towns, etc.). If a county is changed to a city (as often happens), new works on that subject will be classed in DS796, not DS793. Except for cartographic material, we follow Cutters that are listed in the schedule; otherwise, we Cutter localities according to current established form.

The Library sees several disadvantages to continuing the current practice after pinyin conversion. Mixing Wade-Giles and pinyin Cutters in the same class number would be confusing. The printed schedule for DS796 would become very crowded if we provided x-refs. for all of the converted headings. Librarians and users would still have to distinguish between different kinds of jurisdictions. The frequently changing jurisdictional categories are time consuming and

inconvenient for catalogers to deal with, and make it hard for reference librarians to find material on the shelves.

For this reason, CPSO has proposed beginning anew with a new class number for Chinese history at the time of pinyin conversion. Class DS793 will still be used for Provinces, regions, etc., retaining current Cutters for provinces and the references from pinyin forms that were added to the schedule in 1999. Class DS796 for Cities, towns, etc., will be abandoned, except for the major municipalities which are already classed there -- Hong Kong, Macau, Tianjin, Shanghai and Chongqing. A new number, DS797, will be established to include cities, prefectures, counties, and towns. In DS797, Cuttering will be carried out by province. Beijing will continue to be classed in DS795.

Inaugurating a new class number for Chinese history would bring with it several advantages. Old Cutters would pose shelf listing problems after conversion; there is a great advantage in simply being able to Cutter what one sees. There would not be a need for many x-refs. in the printed schedule for the new number.

Not having to distinguish between different kinds of local history will speed processing, and make it easier to find material on the shelf. Many counties and prefectures have changed to become cities, and that pattern is likely to continue. Having all jurisdictions below the provincial level included in one class number would mean that even if there is a jurisdictional change, most material about one place would be Cattered in the same location.

LC would not reclassify already cataloged materials. Added volumes to existing multipart items and analytics in monographic series classified as collections will continue to use existing call numbers even if the numbers have been made obsolete. The list of Geographic Area Codes (GACs) will be changed to incorporate pinyin changes to the names of provinces.

The new numbers for cities will not be used until Descriptive Day 1 for pinyin, October 1. CPSO plans to issue the documentation supporting implementation before that date.

### **Authority records**

LC's Cataloging Distribution Service determined that, based on constituent response, it will not provide a special pinyin authorities distribution. RLG and OCLC are considering making a separate file of converted authorities available to their members. They are also offering to provide a customized file based on the use of headings in a library's converted bibliographic records.

What can we expect from these converted authorities? LC, RLG, OCLC and other major players have worked out specification for the machine conversion of authority records very carefully. For many records, we can expect that machine conversion will have been done on the 1xx field to change it from Wade-Giles to pinyin. In those cases, the former 1xx based on Wade-Giles will be moved to the 4xx and marked as the previous form of entry (\$w nne). We can also expect that the records will be coded in the fixed field (008 byte 07) to indicate that it has been converted. (For details about the authority record marker see: <http://lcweb.loc.gov/catdir/pinyin/authorities.html>). In some cases, a person will have had to review the authority record because the conversion software will be unable to determine that it should be converted by machine. In the instances where a person has had to review the record, we can expect that the 1xx will be converted and that there will be a 4xx representing the W-G form of name.

Originally, the plan had been only to convert the 1xx and the Wade-Giles 4xx. This would have created substantial problems in public catalogs, particularly for author/title authorities. However, LC and OCLC have agreed that 4xxs and 5xxs will also be converted, if they meet the conversion specifications. This is very good news for all of us.

Additionally, 670s will not be converted. After the bib files are converted, this will mean that a person trying to search for the bib record on which the authority was based will not be able to find it using the romanization that appears in the 670. Since many of us are non-Chinese speakers trying to resolve conflicts in our catalogs, this will make the job of the generalist dealing with these headings that much more difficult.

Other things you'll need to consider are: How does global update work in your local system? Are global changes even possible? What about place names that are embedded in subfield \$z of 650 fields or that have been used as qualifiers. Does your system support the finding and changing of character strings from one form to another in particular fields, regardless of placement in the string? If it doesn't, how will you deal with embedded character strings like this? Do you have the staff available to track down the needed changes and make them manually? And, if you plan on doing global updating, test it out on some of these records first to see how it will work.

Can you overlay authorities in your local system or will you need to delete the existing authorities and load the new ones in? Not all of the local systems currently under development are able to overlay existing authority records.

Discuss these issues with your catalog management and authority control staff.

### **Original, copy, and retrospective cataloging**

Because the changes to authority records and bibliographic records will **not** happen simultaneously, libraries that do any Chinese-language cataloging face a dilemma: either they must live with a mixture of romanization schemes in their local catalogs, or they must stop all cataloging activity in the area until all bibliographic and authority records have been converted, both locally and nationally. The abbreviated timeline of what has been happening or is expected to happen is this:

*1999 – LC converts authority records for Chinese conventional place names and also changes those headings on its bibliographic records. All new cataloging done at LC or in the major utilities should use the new forms of these headings in original cataloging*

*July 2000 - LC declares that all newly-established subject headings will use pinyin and begins to convert subject headings manually*

*August 2000 - RLG begins to convert LC's bibliographic records; OCLC begins to convert national name authority records (NARs)*

*August/September 2000 - LC's bibliographic records converted by RLG and redistributed*

*September 2000 - LC loads converted NARs and bib records into its local database; NARs are redistributed*

*October 1, 2000 – The official Day 1: new bibs and authorities are to be entered in the utilities using pinyin (including acquisitions records); copy cataloging will continue to reflect a mixture of romanization until all bibs have been converted*

*October 2000, OCLC's Conversion of WorldCat begins ( Chinese records first (Lang = 'chi'); Newest records converted first; Then, non-Chinese records; Especially, Japanese and Korean; Batchloading software modified so that incoming records are converted as they load)*

*April 2001 - OCLC and RLIN complete conversion of their bibliographic files*

*Sometime in 2001 – RLG and OCLC offer services to member libraries to assist them with the conversion of their local data*

Some things are mandated in this process: before October 1, 2000, libraries are **not** supposed to use pinyin description in new records that they contribute to the utilities (although my catalogers have found evidence that some libraries in OCLC already are using pinyin). After October 1, 2000, all new records entered into OCLC or RLIN are supposed to use pinyin. But there will be a lag time of approximately six months (if all goes as quickly as planned) before the existing bibliographic records in the utilities' databases are all converted.

### **Copy cataloging for new materials or retrospective conversion**

Libraries trying to do Chinese-language copy cataloging in this time period have three options: defer the copy cataloging until the conversion of the bibliographic records in those databases are complete, convert the records manually, one-by-one, as they are brought into the local catalog, or accept whatever form of romanization exists on the record and wait to collect only the local records needing conversion and have them converted by the utilities or manually, as a group. All bibliographic records that have been converted (or created originally using pinyin) in the national utilities are to use a special 987 field marking them so that both catalogers and local systems can recognize that they have been converted or reviewed. The absence of a 987 field or the status code of "r" (indicating a record requires manual review) will allow some systems to identify only the groups of records that need conversion or review.

So, what will you do locally with your Chinese copy cataloging? Can you afford to postpone all copy cataloging until everything is completed in the national files? If you're a library with a small East Asian collection but with some permanent staff devoted to cataloging these materials, the answer is probably "no". If you're a library with the occasional Chinese-language title that you outsource to a vendor, the answer is probably "yes". At the University of Oregon, waiting to do any copy-based cataloging (either of new materials or retrospective conversion of card catalog records) until everything was completed nationally was not an option. We discussed the possibility of starting to convert, manually, the descriptive elements of our new and retrospective conversion COPY cataloging immediately and planning to do global updates of authority-controlled fields after the NARs became available. We would have had to keep track of the headings on those records, load the authority records separately, and run global updates. Or, we could wait until Day 1 (Oct. 1), begin our new and retrospective conversion cataloging on that day using pinyin (which would mean some manual conversion of descriptive elements in existing OCLC records since they won't be finishing the bib file conversion until 4/01) in both descriptive elements and authority-controlled fields. Or we could accept records with descriptive elements in either Wade-Giles or pinyin and deal with the conversion of records in our local database as a group, after the national work was completed. We have tentatively decided to start changing the descriptive elements on our copy cataloging to pinyin on Day 1, October 1, at the same time that we start to supply new original records nationally using pinyin. This seemed like as clean a break as we could manage.

What about retrospective conversion? We have decided to postpone retrospective conversion of titles requiring original input until Day 1 when the utilities have indicated that libraries are to enter new records into their databases using pinyin. However, we will treat retrospective conversion based on existing bibliographic records in OCLC exactly as we do all other copy cataloging: we will start using pinyin in the descriptive elements beginning Day 1.

### **Acquisitions records**

What about acquisitions records? Most of our Chinese materials are ordered from Hong Kong and the People's Republic. Both places use pinyin. Taiwan is itself shifting to pinyin. There might be someone somewhere who will not deal with Wade-Giles romanization, but since pinyin is ubiquitous they need to know pinyin. Since we include the character text along with the romanization on the order slips, vendors should have no problem at all (romanization by itself often won't work since Chinese has so many homophones -- the word "shi" with the fourth tone can represent 49 different characters, for example)

At what point do we start to key temporary bibliographic records into our local catalogs using the new pinyin system? Are the vendors that we deal with prepared to deal with the new romanization system? Do we convert the romanization of our acquisitions or in-process bibliographic records into pinyin? After discussions between the bibliographer for East Asian materials, Acquisitions, and Catalog Department librarians, the University of Oregon has decided to begin the input of new order or in-process records using pinyin sometime in the summer. The vendors are prepared to accept the new scheme (and will, in many cases, find it easier than having to deal with Wade-Giles romanization) and it will actually be easier to train our student assistants, since they all arrive knowing pinyin romanization and will now be able to begin work immediately. Since the records are not being shared nationally, we are able to begin using the

pinyin before the national Day 1. Since records input in the summer will likely not be cataloged for several months, we should be fully into cataloging in pinyin by the time we are ready to process these materials.

### **Conversion of existing local records**

The two primary utilities, RLIN and OCLC, are offering a variety of services to assist libraries in converting their records. RLG will first convert clusters of LC records followed by the clusters containing the records of individual libraries. A library can order a snapshot of its converted records as soon as conversion of all of its records is completed. OCLC will offer three conversion options for libraries: 1) Conversion based on the library's local database; 2) Conversion based on the library's archive records; 3) Delivery of new copies of converted master records. The utilities have also offered to provide matching authority records based on the bib file converted. As an OCLC member library, the University of Oregon plans to send OCLC a set of records from its own database and have those records converted by OCLC and acquire the matching authority records from them at the same time. If we were to try to convert all of our existing records by in-house review and manual conversion, we would be unable to do any new cataloging or retrospective conversion of manual catalog records for a considerable period of time. While my catalogers would feel more comfortable being able to do all of this work personally, it is not a model that we can afford. We have made an administrative decision to trust the work that has been undertaken by LC, OCLC, and RLG collectively and use the vendor's machine conversion to assist us. We have also decided that we will not be undertaking any systematic review of the machine-converted records once they are returned to us.

But what do you need to consider if you choose such an option? First, you need to estimate the number of bibliographic records that would need to be converted and how you would identify them (perhaps have your Automation or Systems Dept. run a program singling out records based on the language code in the fixed field). Then you need to prepare a budget for what it would cost to take advantage of any of the options your utility is offering and present that to your Library Administration. You also must work with your Systems staff to make sure they are prepared to help you through whatever technical issues arise. If you wish to send a local file to your utility for processing, you must make sure that you are able to extract the records and send them in the appropriate form to your utility. Your library must be able to extract appropriate MARC records from the local system. You or your Systems staff will need to work on this with your utility. You must know your local system's capabilities. Is it set up to accept 987 fields? Do you want to have those fields loaded into your local catalog or stripped out on import? Work with your systems staff, create a test file, run a small test to look for unexpected bugs. Be sure you understand how records are loaded into your local system

At the University of Oregon, we estimate that we have about 11-12,000 bibs that need to be converted from Wade-Giles to pinyin. Our estimated budget, based on the still sketchy data available from OCLC (Price range expected to be about 20 to 25 cents per record), is that the conversion and supply of matching authority records would run between \$3000 and \$4000.

There are some words in Japanese and Korean language records that are Chinese and will need to be converted.

Incorrect language coding may keep a record from being converted. Incorrect tagging may keep other fields and subfields from being converted.

Some words may be converted to wrong word if diacritics are not correct.

### **Public service issues**

“Gap” period will be a time of challenge for OCLC, CJK 3.01 users, library administration, and library end-users. LC, OCLC and RLG are working to keep the “gap” as short as possible. But, even so, there will be a period of time ranging from a few months to perhaps a year when there will be a mixture of romanization schemes in bibliographic and authority records. What do you, as catalog managers, need to do to prepare your colleagues and your users?

Get the word out. UO put up a Web page some time ago outlining the issue and attempting to reassure users. We have met with PS heads to discuss coming changes and to let them know that the changes will be implemented gradually. Help your public services colleagues prepare bibliographic instruction sessions that take this conversion into account, if the change is considered significant enough to warrant it.

If you plan to revise the cutters for place names and authors, make sure that you have discussed the impact of this on the staff who will be doing re-labeling of spine labels.

If you have Romanized records, have the new romanization guidelines available for your staff. Include Acquisitions staff if they have been involved in creating brief bib records in your local system or in finding records in the utility. For a while, they will have to know both systems of transliteration while the records in your utility and your local catalog are being converted.

Are you part of a consortium database? Have you had any discussions with your other member libraries? What are your sister libraries doing?

### **Administrative issues**

Make sure your administration is informed. Estimate how many records will be affected. Have a budget for conversion in mind. Decide if you will do it all in-house or will rely on the help of one of the utilities or another vendor. Decide if you want to convert only CJK records or all records that contain Chinese data, such as names, uniform titles, subject headings.

Involve your systems people in planning for the conversion. Be aware of efforts taking place in institutions that use the same online system as you.

### **Conclusion**

Anyone who categorizes catalogers as rule-bound, navel gazers hasn't been paying attention for the last 25 years. I've been working in catalog departments, first as a student in my college library then as a cataloger and now as a manager, since 1975. Those have been 25 tumultuous years. During that time, catalogers have become accustomed to crises. Perhaps we even thrive on them. First came the widespread adoption of the MARC format and computerization. The move to online systems, with varying degrees of integration between different library functions, is

another challenge that most of us have faced. Next came AACR2 with dire predictions of widespread chaos. Retrospective conversion is somewhere in there (for some of us, that is still an active issue). The increasing use of copy cataloging, paraprofessionals doing increasingly more complex work, and outsourcing are trends affecting all catalog departments dramatically. Many of us are now facing the move to the next generation of online system with sweeping promises of increased flexibility and customization that often initially, at least, seem to bring a reduction in system functionality. The need to provide content analysis and bibliographic control for electronic resources is a challenge in which many people see the end of cataloging and library catalogs. Redefinitions of seriality loom ever closer on the horizon. For larger libraries with specialized East-Asian collections, conversion from Wade-Giles to Pinyin ranks right up there with the other crises. Yet, even for those larger libraries will time prove this to have been less of a crisis than we all feared? Certainly, the planning effort for this conversion has been tremendous and the coordination is greater than any other change I've witnessed in 25 years.