Technical Bulletin 240
Pinyin Conversion Project

About Technical Bulletin 240

Contents

Technical Bulletin 240 covers the following topics:

Route to

Cataloging and local system staff.

Why read this?

To learn about the OCLC® conversion of transliterated Chinese characters in OCLC-MARC records from the Wade-Giles to the pinyin romanization system. This technical bulletin also defines the conversion markers for authority records (field 008/07, Romanization Scheme) and bibliographic records (field 987, Local Romanization/Conversion History).

Record processing information

Important: Sections 4 and 5 describe changes that may affect local system processing of OCLC-MARC records.

Conversion start date

In phases between August 1, 2000 and April 2001. OCLC will notify users of project progess via logon Messages of the Day.

Questions

Contact your OCLC-affiliated regional network, service center, or OCLC distributor with questions or for support. Contact the Training and Implementation Specialist at OCLC if you are an independent institution.

User Support

User and Network Support hours at OCLC: 7:00 am-9:00 pm US Eastern time, Monday through Friday; 8:00 am-8:00 pm US Eastern time, Saturday. Saturday support is for telecommunications and hardware only.

USA and Canada 1-800-848-5800; Mexico 95-800-848-5800; International and local (Central Ohio) 1-614-793-8682.

User and Network Support via e-mail: *support@oclc.org*

Documentation affected

OCLC-MARC Tape and Export Record Formats

OCLC Cataloging Service User Guide < http://purl.org/oclc/cug >

Bibliographic Formats and Standards < http://purl.oclc.org/oclc/bib-formats >

Authorities User Guide< http://www.oclc.org/oclc/man/5411aug/frames_man.htm >

Concise Input Standards < http://www.oclc.org/oclc/man/7366cis/toc.htm >

OCLC Batchload User Guide < http://purl.org/oclc/batchloading-guide >

Guidelines for National Level Enhance Participants
< http://www.oclc.org/oclc/cataloging/enhance/guidelines.htm >

Enhance Training Outline
< http://www.oclc.org/oclc/cataloging/enhance/outline.htm >

1 Overview of Pinyin Conversion

Background

OCLC, in close cooperation with the Library of Congress and the Research Libraries Group, has developed plans for the conversion of authority and bibliographic records that use the Wade-Giles Chinese romanization scheme to reflect pinyin romanization.

For the romanization of Chinese, most North American libraries have been using the Wade-Giles system, especially since its adoption by the Library of Congress in 1957. The pinyin romanization system, adopted by the People's Republic of China in 1958 and accepted by the government of Taiwan in 1999, in the meantime has become the standard for the rest of the world. It is recognized by the International Standards Organization (ISO) and the United Nations.

Because pinyin is now the world standard, existing authority and bibliographic files will be converted to reflect pinyin. LC, RLG, and OCLC have been consulting regularly, especially since June 1999, in the planning, coding, testing, and implementation of the conversion project.

Project schedule

The table below outlines the schedule of the Pinyin Conversion Project components discussed in this Technical Bulletin.

Who	Task	Begin Date	End Date
OCLC	Extraction and conversion of Chinese authority records in the national authority file	August 2000	Sept. 30, 2000
OCLC, RLG, LC	Moratorium on creating and changing authority records with romanized Chinese language data in 1xx, 4xx, 5xx, 64x, and 663 fields	August 1, 2000	Sept. 30, 2000
OCLC, LC	Convert and distribute authority records	August 2000	Mid-Sept., 2000
OCLC	Convert CONSER Serials Records in Chinese	Sept. 2000	October 2000
RLG	Convert and distribute Chinese bibliographic records to OCLC, LC	August 2000	October 1, 2000
OCLC	Modify OCLC batch processing	Sept. 2000	
OCLC	Update validation tables for online users and download them to OCLC CatME for Windows and OCLC CJK software users	Sept. 2000	
OCLC, RLG, LC, NACO	Use pinyin romanization in any new or changed authority records. Use 008/07 in authority records to indicate conversion to pinyin	October 1, 2000 (Day One)	
OCLC, RLG, LC, all users	Begin cataloging of bibliographic records using pinyin romanization. Use 987 in new and changed records containing romanized Chinese language data	October 1, 2000 (Day One)	
OCLC	Convert WorldCat bibliographic records	October 1, 2000	April 2001

1 Overview of Pinyin Conversion (continued)

Impact on OCLC CJK software

Aside from updates to the validation tables to accommodate the pinyin authority and bibliographic record markers, there will be no major changes to OCLC CJK software. The Wade-Giles to Pinyin search key conversion program in OCLC CJK software will remain unchanged until most institutions have had a chance to convert their local catalogs. At that time, the direction of the conversion program will be reversed, in keeping with the request of the OCLC CJK Users Group. As always, this program should be used with descretion depending on individual institutional needs.

Pinyin Day One

October 1, 2000 is the mutually agreed upon date for the implementation of pinyin in United States bibliographic systems. Users should refrain from using pinyin for the formulation of systematically romanized access points and other data in bibliographic and authority records until that date. On October 1, 2000, all users should cease using Wade-Giles romanization and begin using pinyin romanization exclusively for all cataloging. See LC's "New Chinese Romanization Guidelines" for the Pinyin Conversion Project on the Web at < http://lcweb.loc.gov/catdir/pinyin/romcover.html >

More information

For further details on the OCLC Pinyin/Wade-Giles Conversion Project see the OCLC Web site at < http://www.oclc.org/oclc/pinyin/index.htm >.

2 Authority Records

Authority record conversion

OCLC, LC and RLG have identified authority records that represent headings used in Chinese bibliographic records. OCLC will convert authority records for use in the national authority file. This conversion began on August 1, 2000 and will be completed by October 1, 2000 ("Day One"), at which time all LC Chinese current cataloging will begin to reflect pinyin romanization. Converted authority records will be marked with appropriate codes in the 008/07 (Romanization Scheme) fixed field.

Moratorium on record creation and editing

From August 1 to September 30, 2000 there will be a moratorium on the creation of, or change to any authority record that contains (or will contain) systematically romanized Chinese language data in 1xx, 4xx, 5xx, 64x, or 663 fields. LC will not delete authority records with systematically romanized Chinese language data in 1xx, 4xx, 5xx, 64x, or 663 fields during this period. This moratorium will ensure that

- All appropriate authority records are converted by OCLC during the identification and conversion process.

- No new Chinese language elements are introduced into the authority file that would be missed in the conversion process, or worse, be converted twice, thereby introducing errors into the file.

End of moratorium on record creation and editing

OCLC and LC expect to complete the conversion and distribution of the estimated 180,000 authority records by mid-September 2000. Since October 1, 2000 is the mutually agreed upon start date for the implementation of pinyin, LC and PCC participants will not use pinyin for the formulation of systematically romanized access points in bibliographic and authority records until October 1. The moratorium on the creation, change, and deletion of authority records will continue through September 30, 2000.

NACO participants

NACO participants should use the appropriate Pinyin Conversion Marker in new authority records created starting October 1, 2000. If updates to an existing authority record involve pinyin romanization, the record should include the appropriate Pinyin Conversion Marker starting October 1, 2000.

3 Bibliographic Records

Record conversion

Conversion of bibliographic records is based on the specifications developed in cooperation with LC and RLG. Converted bibliographic records will be marked with a locally defined MARC 987 (Local Romanization/Conversion History) field so that risk of double conversion is minimized.

CONSER serial record conversion

OCLC will convert all the Chinese language CONSER serial records (estimated to be about 7,000 records) in one group, early in the conversion process. These will be reloaded and distributed to LC separately. Additional Wade-Giles text in other CONSER records will be converted to pinyin as encountered and will be distributed as part of the regular CONSER distribution.

Library of Congress record conversion schedule

RLG will convert LC's Chinese language bibliographic records and distribute them to LC and OCLC. This conversion is planned to be completed by October 1, 2000 ("Day One"), at which time all LC Chinese current cataloging will begin to reflect pinyin romanization.

OCLC record conversion schedule

OCLC will begin to convert the WorldCat bibliographic file, working backwards from the most recent records, soon after October 1, 2000. The expected completion date is April 2001.

New record creation

All OCLC Cataloging users should include an appropriate Pinyin Conversion Marker 987 in new bibliographic records created beginning October 1, 2000 that contain romanized Chinese characters. They should also add an appropriate field 987 to any existing bibliographic record containing pinyin romanized Chinese characters that is locked and replaced beginning October 1, 2000. Records that include field 987 will not be converted subsequently by OCLC's pinyin conversion programs, eliminating the potential for erroneous conversion.

Batchload processing

OCLC will modify batchload processing to ensure that incoming bibliographic data is appropriately evaluated and converted to pinyin if necessary. This conversion processing ensures that existing matching software works correctly following conversion of Wade-Giles data in WorldCat. It also prevents the re-introduction of Wade-Giles data into WorldCat starting October 1, 2000.

Local data conversion options

OCLC will offer a variety of local data conversion options to both members and non-members. The options are variations on the three scenarios below, each of which provides the option of including corrected authority records.

- Full conversion services based on a library's local database

- Full conversion services for a file created by OCLC Local Database Creation, using a library's OCLC archival records

- Delivery of new copies of converted OCLC master records, which include vernacular data, when present

Additional conversion options and variations are available. For details on the local conversion offerings, see the OCLC Web site at < http://www.oclc.org/oclc/pinyin/index.htm >.

4 Pinyin Conversion Marker for Authority Records: Field 008/07

Field 008/07

The 008/07 field ("Romanization scheme") will indicate whether an authority record has been converted to pinyin. Name, series and subject authority records will be given a pinyin marker during the conversion process. When the Library of Congress declares that all authority records in the name and subject authority files have been converted, the marker will no longer be used. If authority record conversion has not been completed by October 1, 2000 ("Day One"), catalogers may be asked to code new authority records 'c' or 'n' until conversion has been completed.

Value 'c'

Use value 'c' to indicate headings which have been converted to the pinyin form according to the romanization guidelines prescribed by the Library of Congress.

Note: No category of headings that are "already in pinyin form" exists, since technically no headings have been romanized according to the new guidelines. Current headings which appear to be in pinyin form either also represent valid Wade-Giles forms, or have not been romanized but established according to usage.

Value 'n'

Use value 'n' to indicate headings which were considered for conversion but were not converted because they were not romanized according to Wade-Giles guidelines.

Note: Use value 'n' for headings which appear to be in pinyin form but in fact have not been romanized. For example, the heading:

Liu, Chang, 1954-

appears to be in Wade-Giles form. In fact, the heading was established for the author of an English-language monograph. The form of heading was based on how the author's name appeared on the title page. After evaluation, code this heading 'n' because it has not been romanized according to Wade-Giles guidelines. It would not be converted to pinyin.

More information

For the most current text of the Pinyin Conversion Marker for Authority Records, see the LC Web site at < http://lcweb.loc.gov/catdir/pinyin/authorities.html >.

5   Pinyin Conversion Marker for Bibliographic Records: Field 987

Field 987

Use local MARC field 987 to record temporary information about the conversion status of MARC records that contain romanized Chinese data. Participants in the Pinyin Conversion Project, an effort to convert romanized Chinese data from the Wade-Giles to pinyin romanization schemes, will use field 987. Beginning October 1, 2000, add this field to all records that include romanized Chinese characters.

Production of field 987

Field 987 may originate from manual input or system generation.

Effect on record processing

If a record includes field 987, that record will not be converted by the utilities' conversion programs. Thus, adding field 987 to a record eliminates the potential for erroneous conversion.

Local system implementation

This local field will be implemented in the local library systems of the Library of Congress, OCLC, RLG, and others to store conversion status information until it is no longer needed. Because field 987 is a locally defined data element, all local systems are not required to implement this field. Its definition for use with the Pinyin Conversion Project may conflict with the local definition of field 987 in local systems not involved with the project. This should not be a problem for non-participants who do not receive cataloging records for Chinese materials from participants.

Field 987 definition

The table below defines OCLC-MARC field 987. **Note:** NR=nonrepeatable field.

First Indicator	
Undefined	
Second Indicator	
Undefined	
Subfield Code	
‡a	Romanization/conversion identifier (NR)
‡b	Agency that converted, created, or reviewed romanization/conversion (NR)
‡c	Date of conversion or review (NR)
‡d	Status code (NR) c Record fully romanized (by conversion or cataloger input) n Record processed but not converted (no eligible strings detected) r Record requires manual review to fully convert romanization
‡e	Version of conversion program used (NR)
‡f	Note (NR) [Free text note about status of the conversion of the record]

Field 987 examples

The table below contains some common examples of 987 fields added to converted bibliographic records.

Record whose romanization was fully altered by program
987 _ _ PINYIN ‡b CStrRLIN ‡c 20000619 ‡d c ‡e 1.0 ‡f [note on conversion]

Record reviewed by OCLC pinyin conversion software
987 _ _ PINYIN ‡b OCoLC ‡c 20000324 ‡d [code] ‡e 1.0 ‡f [note on conversion]

Record reviewed and status changed from 'r' to 'c' by LC cataloger
987 _ _ PINYIN ‡b DLC-R ‡c 20001010 ‡d c

Record created by LC cataloger during transition, following pinyin rules
987 _ _ PINYIN ‡b DLC-R ‡d c

Guidelines for manual input

The table below provides guidelines for manually inputting values in the subfields of field 987.

Subfield Code		Guideline
‡a	Romanization/conversion identifier (NR)	Type **_PINYIN._**
‡b	Agency that converted, created, or reviewed romanization/conversion (NR)	Use MARC Code (NUC Symbol) of institution performing manual review/conversion of record. MARC Codes may be found on the Web at < http://www.oclc.org/oclc/forms/pisearch.htm > or in your institution's primary record in the OCLC Name-Address Directory (IDENTITY subfield ‡n) **Note:** Do not use the OCLC institution symbol in this field.
‡c	Date of conversion or review (NR)	Use date format: YYYYMMDD **Note:** Do not use for new records.
‡d	Status code (NR) c Record fully romanized (by conversion or cataloger input) n Record processed but not converted (no eligible strings detected) r Record requires manual review to fully convert romanization	Use only Status code 'c' for new and manually converted records.
‡e	Version of conversion program used (NR)	Do not use for manual input.
‡f	Note (NR) [Free text note about status of the conversion of the record]	

More information

For the most current text of the Pinyin Conversion Marker for bibliographic records, see the LC Web site at < http://lcweb.loc.gov/marc/pinyin.html >.

6 Technical Bulletins Online

Introduction

This section covers retrieving Technical Bulletins in electronic form.

WWW

To access Technical Bulletins on the OCLC® Web site:

	Action
1	Type *www.oclc.org/oclc/menu/tb.htm* in the URL locator of your World Wide Web browser.
2	Click the name of the Technical Bulletin you want to view.

File Transfer Protocol

To retrieve Technical Bulletins via File Transfer Protocol:

	Action
1	Type *ftp.rsch.oclc.org* in the entry line of your FTP software and log on as anonymous using your e-mail address as the password.
2	Access the file at **/pub/documentation/tb/txt**.

Internet list

To subscribe to TECHBUL-L—a public list that e-mails an announcement of availability of new Technical Bulletins on the OCLC Web site:

Action
Type *www.oclc.org/oclc/forms/listserv.htm* in the URL locator of your World Wide Web software. Click **send** after you complete the form. Or Type *subscribe techbul-l* in the body of an e-mail message and send it to *listserv@oclc.org*. To unsubscribe, type *unsubscribe techbul-l* in the body of an e-mail message, and send it to *listserv@oclc.org*.

Listserv

To retrieve Technical Bulletins from the Internet list:

Action
Type *get tb_txt.[nnn]_txt* in the body of an e-mail message and send it to *listserv@oclc.org*. For example, type *get tb_txt.235_txt* to retrieve Technical Bulletin 235.

Comments

The OCLC Documentation Department welcomes comments about this document. Please send them via e-mail to *lois_yoakam@oclc.org*.

7 Current OCLC Technical Bulletins

List below current as of August 2000. Technical Bulletins not listed are obsolete.

No.	Date	Title	OCLC Manuals Affected
203	9406	Internet Access to OCLC	The relevant portion of this TB is available at < http://www.oclc.org/oclc/menu/tb203r.htm >
210	9511	PRISM Usage Stats	Name-Address Directory User Guide*, Union List User Guide*
217	9606	Cataloging Agent Authorization	Union List User Guide*
220	9702	OCLC-MARC Code Lists Changes	Bibliographic Formats and Standards*, OCLC-MARC Code Lists*, print editions; Web versions current
221	9703	UNIMARC Conversion Project	OCLC Selection User Guide*, OCLC-MARC Tape and Export Record Formats
222R	9807	OCLC Interlibrary Loan Enhancements	Union List User Guide*
223R	9802	OCLC ILL Management Statistics	Name-Address Directory User Guide*
225	9707	OCLC Product Services via the Web	
227	9809	OCLC-MARC Bibliographic Update 1998	Bibliographic Formats and Standards*, OCLC-MARC Code Lists*, print editions; Web versions current
228	9810	OCLC-MARC Authority Update 1998	Authorities User Guide*, OCLC-MARC Tape and Export Record Formats
230R	0001	OCLC Interlibrary Loan Service: New Request Status and Special Messages for ISO ILL	OCLC Interlibrary Loan Service User Guide*
231	9905	OCLC-MARC Authority Update 1999	Authorities User Guide*, OCLC-MARC Tape and Export Record Formats
232	9905	OCLC-MARC Field 994	OCLC-MARC Tape and Export Record Formats
233	9906	Interlibrary Loan: Supplying "Reasons for No" Responses	OCLC Interlibrary Loan Service User Guide*
234	9908	Changes to OCLC Interlibrary Loan and Direct Request Services	OCLC Interlibrary Loan Service User Guide*
235R	0008	Keyword Searching Changes in WorldCat; not yet implemented	OCLC Cataloging Service User Guide*, Searching for Bibliographic Records*, Selection User Guide*, Union List User Guide*
236	0003	OCLC-MARC Bibliographic Update 2000	Bibliographic Formats and Standards* (Web edition is current), OCLC Cataloging Service User Guide*, OCLC-MARC Code Lists*, OCLC-MARC Tape and Export Record Formats, Searching for Bibliographic Records*
237	0008†	Multiple-surname Coding Change [title changed]	Bibliographic Formats and Standards*, Authorities User Guide*, OCLC-MARC Tape and Export Record Formats, Searching for Bibliographic Records*
238	0005	OCLC Arabic Cataloging	Bibliographic Formats and Standards*, OCLC-MARC Tape and Export Record Formats
239	0005	Integration of the OCLC Cataloging Service and CORC	Bibliographic Formats and Standards*, OCLC Cataloging Service User Guide*, OCLC-MARC Tape and Export Record Formats, Guidelines for National Level Enhance Participants*
240	0008	Pinyin Conversion Project	Bibliographic Formats and Standards*, OCLC Cataloging Service User Guide*, OCLC-MARC Tape and Export Record Formats, Guidelines for National Level Enhance Participants*, Authorities User Guide*, Concise Input Standards*, OCLC Batchload User Guide*, Enhance Training Outline*

*Electronic versions of these documents are available at the OCLC® Web site: < http://www.oclc.org/oclc/menu/doc.htm >.

† Release of TB 237 was delayed from March 2000.