6-1-2002

The Aftermath of the Chinese Pinyin Conversion Project

Rosina Leung

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

BYU ScholarsArchive Citation
Available at: https://scholarsarchive.byu.edu/jeal/vol2002/iss127/3

This Article 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.
THE AFTERMATH OF THE CHINESE PINYIN CONVERSION PROJECT

Rosina Leung

University of Toronto

Introduction

Starting from the summer of 2000, the Library of Congress, RLG and OCLC began the Chinese Pinyin Conversion Project to change the romanization of Chinese records in RLIN from Wade-Giles to Pinyin. A special computer program was developed to perform the conversion. This program was a universal program designed to deal with all the Chinese language records submitted by all of the libraries in the North America. Therefore it could not handle those situations that occurred only in the records of individual libraries such as typographical errors and the misuse of diacritics. This being the case, it was known from the outset that records in the RLIN database could not be converted completely, and manual adjustments to partially converted records would be required. This paper describes how such institution-specific errors in the converted database were identified and cleaned up at the University of Toronto Cheng Yu Tung East Asian Library.

When the conversion program was being developed, the Cheng Yu Tung East Asian Library (EAL), whose holdings in RLIN include around 130,000 Chinese records, was one of the libraries invited to submit a sample of records for testing the program and to provide feedback on the converted records. After the tests were completed, the results of the final conversion of our sample records were used for identifying the types of errors that could possibly occur in our database. Although the size of this sample was too small to enable us to identify all the errors in the converted database, it nevertheless did provide us a good understanding of the errors that could occur. Hence, EAL used the results of the sample to formulate the strategies for the clean up plan as described below.

I would like to give special thanks to Ms. Anna U, Director of the Cheng Yu Tung East Asian Library, University of Toronto and Mr. Jack Howard, Librarian of the H. H. Mu Far Eastern Library, Royal Ontario Museum, for kindly reviewing this article. They have given me valuable comments on the contents as well as on the English writing of this article.

Analysis Of The Test Samples

As mentioned above, EAL submitted 100 sample records to RLG for testing the conversion program. The sample included records with potential for different kinds of problems. These included regular records without any particular problem, old error records with obsolete or wrong diacritics, personal name inconsistencies, inconsistent aggregations, mixed Wade-Giles and Pinyin subject headings, and mixed texts (foreign language characters, English and other non-Chinese strings, etc.). The converted sample data was analyzed and the types of errors were identified. These errors, listed in Appendix B, were used for two purposes: (a) to develop sub-
plans to deal with the different types of errors identified; (b) to train the staff involved in each sub-plan as described below.

Based on the result of our analysis of the sample records, the clean up plan was separated into four sub-plans as follows:

Sub-plan 1: Error Records
Sub-plan 2: Old Records
Sub-plan 3: Manual Review Records
Sub-plan 4: On-going Correction upon Discovery

**Characteristics Of Records In Each Sub-Plan**

Sub-plan 1 deals with the batch of error records marked with code “e” in the fixed field ST in the RLIN database. These error records were identified and marked code “e” by RLIN when they were uploaded from UTLAS to RLIN before 1983. Correction of the errors has continued since that time. However, they could not all be corrected before the conversion took place. The batch size of these remaining “e” records is 11,209. During the conversion process, the conversion program converted all the fields in the records except for the fields marked with the error code. The conversion program assigned a marker 987‡dc to these records if the rest of the record was correct. Sub-plan 1 was designed to (a) identify these records with errors which have been marked with 987‡dc during the conversion of the database in RLIN and (b) correct them.

Sub-plan 2 involves a batch of records coded with 01-01-01 in the fixed field UD. Similar to the error records in Sub-plan 1, these records were coded with 01-01-01 by RLIN when they were uploaded from UTLAS to RLIN before 1983. These were the old records without Chinese characters. In order to facilitate the identification of these records so that Chinese characters could be added subsequently, RLIN was requested at that time to code these records with 01-01-01. This batch of records contains both records with or without the fixed field ST coded “e” and may have marker 987‡dc or 987‡dr depending on whether there are any other errors in the record. The batch size of these old records is 38,164.

Sub-plan 3 deals with the records marked by the conversion program for manual review during the conversion. This batch of records contains the marker 987‡dr. The batch size of the manual review records is 15,986.

Finally, there may be some errors that have not been identified at this moment because they were not included in the sample records submitted to RLIN for the test run. Therefore, Sub-plan 4 was established to address this particular problem. Since we do not know the characteristics of these errors, we have had to develop a more detailed plan designed to clarify the definitions of the errors, and the responsibilities and the procedures to be performed by each staff member involved in this Sub-plan, as described in Appendix A.

The following table summarizes the four Sub-plans:
<table>
<thead>
<tr>
<th>Sub-plan</th>
<th>Record Type</th>
<th>Number of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Error Records</td>
<td>11,209</td>
</tr>
<tr>
<td>2</td>
<td>Old Records</td>
<td>38,164</td>
</tr>
<tr>
<td>3</td>
<td>Manual Review Records</td>
<td>15,986</td>
</tr>
<tr>
<td>4</td>
<td>On-going Corrections Upon Discovery</td>
<td>Not Known</td>
</tr>
</tbody>
</table>

**Implementation**

The following sections describe the planning and the workflow in each Sub-plan.

**Sub-plan 1: Error records (11,209 records)**

As mentioned above, this batch of records was loaded from UTLAS into RLIN in 1983. Most of the errors are in tag 100, and 245 with incorrect indicator(s) or subfield code(s).

It was desirable clean up this batch of records prior to the conversion. Otherwise, the tag(s) marked with error(s) would not be converted by the conversion program and would have to be manually converted subsequently. However, this could not done because of limited resources available and time constraints. We noticed that most of these error records had no Chinese characters. This means that upgrading them will be quite time consuming. As a result, the following alternative to cleaning up the batch prior to conversion was adopted.

**Workflow (Planning)**
- Retrieve the error records in a batch with the RLIN command FIN ST E;
- Print out all of the error records;
- Sort them into 12 piles within two groups;
- Group 1: with Chinese characters: (1) 1xx, (2) 2xx, (3), 4xx/8xx, (4) 6xx, (5) others, and Group 2: without Chinese characters: (6) 1xx, (7) 2xx, (8), 4xx/8xx, (9) 6xx, (10) others, (11) Japanese language, and (12) Korean language. We anticipate that most of the error records belong to this latter group without Chinese characters.

<table>
<thead>
<tr>
<th>MARC tag</th>
<th>With Chinese Characters</th>
<th>Without Chinese Characters</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1xx</td>
<td>(1)</td>
<td>(6)</td>
<td></td>
</tr>
<tr>
<td>2xx</td>
<td>(2)</td>
<td>(7)</td>
<td></td>
</tr>
<tr>
<td>4xx/8xx</td>
<td>(3)</td>
<td>(8)</td>
<td></td>
</tr>
<tr>
<td>6xx</td>
<td>(4)</td>
<td>(9)</td>
<td></td>
</tr>
<tr>
<td>Other tags</td>
<td>(5)</td>
<td>(10)</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>(11)</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td>(12)</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Workflow (Editing)**
- Work according to the priority set as above from pile number (1) to (10). Piles number (11) and (12) are to be handled by the Japanese and Korean section;
- Edit the error field(s) and quickly review the record to see if there are any other errors to be corrected;
• Change the fixed field “ST E” to “ST P”;
• Make sure that the fixed fields “RTYP C” and “ST P” are in the record. When EAL purchases the snapshot file from RLIN, EAL will only purchase the records with “RTYP C and ST P” (processed catalogued recorded), and will not purchase Order Records (RTYP O), In-process Records (RTYP I) and Unprocessed Catalogued Records (RTYP C and ST S);
• Add Chinese characters to the record if necessary;
• Retrieve the book from the stacks to verify against the information in the record;
• Check the label on the book against the call number in the record to ensure that the book-in-hand corresponds to the record;
• Search in RLIN to see if there is an appropriate record to copy;
• If there is one, overlay our record with the record in RLIN. Check carefully the publisher, edition and the publication year to ensure that the correct title is copied;
• If an appropriate record cannot be found in RLIN, the project staff will manually enter the Chinese characters and edit the error field;
• Go to DRA (the in-house system currently used at the University of Toronto) to retrieve the corresponding record (still in Wade-Giles). Manually add RLIN record ID in DRA in the 035 field. (This step is to prepare for a successful loading of the snapshot file into DRA after the manual review is done in RLIN);
• Do not edit the corresponding record in DRA because it will be overlaid later. The only exception would be if there is a major error in the record that will affect the user access;
• Change 987‡dr to 987‡dc and delete the ‡f after reviewing the record.

Sub-plan 2: Old Records, UD 01-01-01 (38,164 records)

Planning
The characteristics of records in this batch are quite similar to those of the error records in Sub-plan 1. However, this batch of records is coded with 01-01-01 in the fixed field UD. As mentioned above, this batch of records was marked when the old records were uploaded from UTLAS into RLIN in 1983 and the adding of Chinese characters had not been completed. This batch of records may contain both records with or records without the fixed field ST coded “e”. Also this batch of records may have marker 987‡dc or 987‡dr depending on whether the conversion program could detect any other errors in the record.

Sub-plan 2 will begin after both Sub-plans 1 and Sub-plan 3 have been completed. This is because the error records in Sub-plan 1 and 3 are retrieved in a batch from RLIN rather than individually. Retrieving records by batch is much more cost effective than retrieving each record individually. Since there is no command in RLIN to retrieve the 01-01-01 records in a batch, each record will have to be retrieved by its individual record ID. The record ID can be obtained from RLIN. As mentioned above, the 01-01-01 records originally in RLIN before our clean-up project began may well include records that will be dealt with under Sub-plans 1 and 3. Hence, the number of records that need to be retrieved under this Sub-plan 2 most likely will be reduced significantly after Sub-plans 1 and 3 have been completed.

An initial survey of the records indicates that many of these 01-01-01 records are rare book titles. We foresee that more time will be required to upgrade these records.
Workflow
- This batch of records cannot be retrieved as a batch because there is no searching command in RLIN as FIN UD 01-01-01;
- The library will obtain a list of record IDs from RLIN for this batch and store them in the Word file;
- Each record has to be retrieved individually by its record ID with the RLIN command FIN ID XXXX;
- The workflow of editing is similar to that of Sub-plan 1.

Sub-plan 3: Manual Review Records (987%dr) (15,986 records)

These records are marked for manual review by the conversion program. On May 14, 2001, all of our converted Chinese records in RLIN numbered a total of 129,175 records. After the conversion, 15,986 records (12.38% of the total) were marked for manual review (please refer to Appendix B for the detailed listing of each type of errors that were identified during the sample test run). This batch of records is summarized as below:

<table>
<thead>
<tr>
<th>File code in RLIN</th>
<th>File type</th>
<th># of records</th>
<th>Priority**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bks</td>
<td>Book</td>
<td>15,428*</td>
<td>7 &amp; 8*</td>
</tr>
<tr>
<td>Mdf</td>
<td>Computer file</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Map</td>
<td>Map</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Rec</td>
<td>Recording</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Sco</td>
<td>Scores</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>Ser</td>
<td>Serials</td>
<td>526</td>
<td>5</td>
</tr>
<tr>
<td>Vim</td>
<td>Visual Materials</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>15,986</td>
<td></td>
</tr>
</tbody>
</table>

*In the book file, approximately 9,000 records marked for manual review have Chinese characters (Priority 7), while 6,000 records do not (Priority 8).
**The priority was determined based on the number of records in each file. Higher priority was assigned to those files with fewer records so that a smaller number of files needed to be opened at one time.

Workflow
The workflow of this Sub-plan was divided into 5 stages according to the different levels of knowledge and experience of the project staff. Stage 1 was the training period for the staff. At this stage, the objective was to get them familiar with the procedures, specifications, rules, romanization guidelines and types of errors identified, etc. Procedures in this stage were more specific. As the knowledge and experience of the staff improved over time, procedures were simplified to increase the efficiency.

A graduate student was hired to work part-time on the project (8 hours per week). The librarian spent approximately one hour per day on the project.
Records were classified into different priorities as listed in the above table. Files with higher priority were to be worked on first.

Workflow
- Print out a batch of 100 records in the file at a time starting from the beginning of the batch by using the following commands:
  - Use RLIN command (FIN PY R NOT ST S) to retrieve the records marked for manual review;
  - Use the RLIN command (TYP 1-100 FUL HOL) to print the 100 records required;
- The librarian reviewed the records and marked the correction on the printout according to the information in tag 987xf. At this point, the librarian would review the whole record quickly to see if there are any cataloguing errors or mistakes that have not been identified by the conversion program;
- The student, following the instructions on the printout, performed online editing by using the following commands:
  - Use the RLIN command (FIN PY R NOT ST S, FUL XX) to retrieve the above records for editing.
- Edit only the records in the RLIN database; records in DRA will be changed later.
- After editing, change 987xfdr to dc and delete xf if there is one.

Production Rate
Around 10 records per hour were produced in the beginning and these increased to 20 records per hour after a month.

Stage 2  April 2001 – June 2001

At this stage, the project staff already knew how to correct the errors. Therefore, he was assigned the task of marking the corrections on the printout independently. The librarian reviewed all of the corrections.

Workflow
- Similar to Stage 1 except that the corrections were marked by the student on the record printout and reviewed by the Librarian
- The RLIN commands used were the same as those in Stage 1.

Production Rate
The production rate at this stage was increased significantly because of the simplification of the workflow procedures and the advancement of the student’s experience. The production rate was around 30 records per hour at the beginning and increased to 40 records per hour after another month.

Stage 3  July 2001 – Sept. 2001
The hours of the student were increased to 20 hours per week.
The workflow at this stage changed significantly. Records were not printed in order to save time and paper unless he came across any difficulties and problems. The librarian would review all printed records and provided feedback. In order to ensure the quality, the student was asked to write down the record ID of all the records corrected by him. The librarian randomly checked the student’s edited records online.

Workflow
- Similar to Stage 1 except that the corrections were made by the student online and reviewed by the Librarian on a random basis;
- The RLIN commands used were the same as those in Stage 1.

Production Rate
The production rate was around 40 records per hour at the beginning increasing to 60 records per hour after one month.

Stage 4 Oct. 2001 – Nov. 2001

Workflow, commands used and production rate were similar to those in Stage 3. However, the student was not required to write down all the record ID edited by him except for records with problems.

By this stage, files with priority numbers 1 to 7 were completed.

Stage 5 Dec. 2001 – Present
Stage 5 deals with the file with priority number 8. At this stage, the remaining batch contains only records without Chinese characters. Hence, the student is required to retrieve the book from the stacks to verify the contents of the record. It is important to ensure that the correct book is used for the verification. Emphasis is also made to ensure that an appropriate record with the same author, publisher, publication date and edition is copied. The RLIN record ID is added to tag 035 of the corresponding record in the DRA. Please note that this is the only stage in Sub-plan 3 where the records in the DRA database are modified.

Workflow
- The workflow is similar to those in Sub-plan 1 & 2.
- The RLIN commands used were same as those in Sub-plan 1 and 2.

Production rate: 6-8 records per hour.
The production rate is affected by the following factors:
- Longer verification time is needed as the records have no Chinese characters.
- Books have to be physically retrieved from the stacks.
- If there is no appropriate record that can be copied in RLIN, the Chinese characters have to be input manually;
- Need to go into the DRA database to retrieve the corresponding records and to add the RLIN record ID;
- There are numerous rare book titles in this batch of records. Therefore, it is more time consuming to perform the verification and the editing.
Remarks
For the Sub-plans 1, 2 & 3 (Error Records, Old Records, Manual Review Records, respectively) described above, there are several principles that were followed:

- For Sub-plan 3 stages 1 to 4, use the RLIN command FIN PY R to retrieve the records by batch. By doing so, all the records required in each editing session could be retrieved in one search. An editing session is the time required to perform the online editing before leaving RLIN. By the following this principle, searching costs were minimized.
- Save times and paper by reducing the number of printouts. Only print when necessary.
- For Sub-plans 1, 2 and stage 5 of the Sub-plan 3, ensure that RLIN record ID (for the records without Chinese characters) is added to every corresponding record in DRA. This will facilitate the loading process in the future.

Sub-plan 4: On-going Corrections upon Discovery
This Sub-plan deals with the errors that are identified by the library staff while performing their regular duties. In other words, there are no specific extra procedures or strategies to search for the errors. When anyone of the library staff comes across an error an RLIN record, he or she is automatically expected to follow the precise guidelines under this Sub-plan (refer to Appendix A). The guidelines in Appendix A will specify how the errors should be recorded, reported and corrected depending on the job description of the library staff.

**TIMELINE OF THE SUB-PLANS**

<table>
<thead>
<tr>
<th>Sub-plan</th>
<th>Type</th>
<th>No. of Records</th>
<th>Status</th>
<th>Estimated timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Error Records</td>
<td>11,209</td>
<td>452 records done, 10,857 records left</td>
<td>No resources available</td>
</tr>
<tr>
<td>2</td>
<td>Old Records</td>
<td>38,164</td>
<td>300 records done, 37,864 records left</td>
<td>No resources available</td>
</tr>
<tr>
<td>4</td>
<td>On-going</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Resources For The Sub-Plans**

The resources available for the project vary since June 2000.

*From June 2000 – Jan. 2001*
After the test run, 15% of one full-time librarian was assigned to be responsible for the project.
From Feb. 2001 – May 2001
A graduate student was hired on a part-time basis to work on the project: 8 hours per week; librarian’s time allotment remained at approximately 15%

From June 2001 – Present
The graduate student's hours increased to 20 hours per week, with librarian's time remaining the same as above.

Conclusions

The success of this record clean-up project relies on the knowledge and the experience of the project staff. The more experience the project staff has, the simpler the workflow of the sub-plans can be. Streamlined workflow procedures will reduce the total time required to complete the project. Good knowledge of the project staff about the characteristics of the database, specifications of the conversion and the romanization rules will enable the project staff to identify the errors in a record more precisely. This will improve the quality of the database.

The clean up of the converted bibliographic records in the RLIN database is only one third of the whole conversion clean-up project. After all the records in the RLIN database have been edited, a tape containing all these edited records will be purchased from RLIN and loaded to our DRA database. Certain problems will be encountered during the transfer of these records from RLIN to the University’s DRA system. Matching of the corresponding records in the RLIN and DRA database, for example, may be problematic because some of the records in DRA may not have the RLIN record ID (in the 035 field) required for matching. The final component of the clean-up project is to update the authority records in DRA. This will be another major and time-consuming step in the completion of the clean-up project. The discussion of the other two parts of the clean-up project does not come into the purview of this paper and will be described separately.

This article is intended to share our experiences of the clean-up project. Comments and suggestions are welcomed and appreciated.
APPENDIX A

Definition of Manual Review
There are two circumstances where a staff member needs to manually review a record for errors. In the first situation, whenever a staff member is displaying a record on the screen or looking at a record printout while performing his/her regular duties and finds errors, the record should be reviewed. The second situation is whenever a staff member is performing the above projects and is retrieving a record to correct particular errors. Under both situations, the staff should follow the “What to be aware of when reviewing a record” guidelines (see below).

Who should review a record
Staff who encounter the East Asian Library’s error records will automatically become "reviewers." See “What to be aware of when reviewing a record” below for guidelines. Staff who only review (i.e. cannot edit) records are called Reviewers.

Who should edit a record
Staff who edit records are an "Editors." These individuals should edit only the portion of a record that falls into their area of responsibility. For example, when performing the regular duties, acquisition staff should edit the acquisition fields only. The cataloguing staff should be able to edit the whole record. Those responsible for in-process records should only edit the portion of in-process fields. All staff should follow “What to do when coming across an error record” guidelines.

Acquisition staff
Acquisition staff members are responsible for creating and deriving on-order records, i.e. records with record type RTYP “o”. The fields under his/her responsibility are variable fields: 245, 260, 700, 987 and PVT. These staff members should edit these fields if they come across an error in a record. If they find any errors in the other fields in the same record, they should print out the record, mark the error(s) and send it to the Chinese Cataloguing Librarian.

In-process staff
In-process staff are responsible for creating and deriving in-process records, i.e., records with record type RTYP “i”. They are responsible for editing variable fields: 245, 260, 700, 987 and PVT. They should edit these fields if they come across an error in a record. If they find any errors that are beyond their responsibility in the other fields in the same record they should print out the record, mark the error and send it to the Chinese Cataloguing Librarian.

Cataloguing staff
Cataloguing staff are responsible for editing all the errors in the records.

Project staff
Project staff are assigned by the Chinese Cataloguing Librarian to perform the projects as defined above. Each project staff member will be assigned responsibilities as either a reviewer or an editor.
Other staff
All the staff not included in the above 4 categories will be considered as "other staff." They should only review the records.

What to be aware of when reviewing a record
When reviewing a record, the reviewer/editor should generally focus on:
• Any Wade-Giles string that should have been converted to Pinyin but was not.
• Any Wade-Giles string that was converted to Pinyin incorrectly.
• Any non-Wade-Giles string that should not have been converted to Pinyin but was.
• Any romanization string that is not complied with the LC’s New Chinese Romanization Guidelines
• Any typos or cataloguing errors existing in the record.

What to do when coming across an error record
When coming across an error record, staff should follow the following procedures according to their area of responsibility.

Reviewer
• Print out the error records.
• Mark on the printout the error(s).
• Mark on the printout your comments or suggestions of how to correct the error(s).
• Date and initial the printout.
• The reviewer should not edit the record.
• Send the marked printout to the Chinese Cataloguing Librarian as soon as possible.
• The Chinese Cataloguing Librarian will edit the error record and keep statistics.

Editor (Follow the procedures of Editing a Record immediately below)

Procedures of Editing a Record
• Print out the error records before editing, as the Cataloguing Librarian needs to know what types of errors exist in the database to plan for the Local Conversion Project.
• Mark on the printout what the error(s) is/are.
• Date and initial the printout.
• Edit the record online according to LC rules (if in doubt, consult the Chinese Cataloguing Librarian).
• If staff correct the errors only in part, skip the following two steps.
  * Change the marker 987×dr to 987×dc if ALL the error(s) is/are corrected.
  * After correcting all the errors, delete subfield ×f of 987 tag if there is one.
• Save the record.
• Mark “DONE” on the printout to indicate which error(s) has/have been corrected.
• Mark on the printout error(s) NOT corrected.
• Do not print an edited record.
• Sent the printout to the Chinese Cataloguing Librarian as soon as possible.
APPENDIX B

Types of Errors Identified
As part of its participation in the Conversion Test Run Project led by RLG the East Asian Library submitted 100 sample records that potentially had different kinds of problems. These included ordinary records without any particular problem, old error records with obsolete or wrong diacritics, personal name inconsistencies, inconsistent aggregations, mixed Wade-Giles and Pinyin subject headings, and mixed texts. Based on the results of five-test runs on the sample done by RLG, we identified several errors that need to be addressed as part of our Chinese Pinyin Conversion Project as listed below. These errors either need to be corrected manually or by computer program. The list of errors below is not meant to be comprehensive because the sample was small and unlikely to contain all types of errors.

Some of the following errors can be retrieved by searching, but many of the errors cannot be retrieved, e.g., the batch of records coded with fixed field ST “e” can be retrieved by the command “fin st e”. The batch of records with marker 987×dr can be retrieved by the command “fin py r”. Another batch of records with fixed field UD coded “01-01-01” can only be retrieved by the command “fin id xxx”. There is, however, no specific way to retrieve the rest of records with the errors. They can only be corrected as staff come across them or by some kind of local conversion program.

Error records with the fixed field ST coded “e” and marker 987×dc
Field(s) marked with “?? SI 10” were not converted. Other fields in the same record were converted if the Wade-Giles romanizations of those fields were correct.

E.g. ONTG1035791-B and ONTG1066669-B, field 100 and 245 are not converted. But the record has marker 987×dc.

Hyphen
- Place name without hyphen in the original record
If the place name in the record was input without a hyphen, the conversion program will not join the syllables together during the conversion.

E.g. ONTG1035791-B, field 260 “Chung kuo” was input without a hyphen and was converted into “Zhong guo” instead of “Zhongguo”.

- A hyphen added to Wade-Giles incorrectly will lead to error after conversion.
E.g. ONTG1191769-B, field 260 “Min-Kuo” should have been input as “Min kuo” and it would then have been converted into “Min guo” instead of “Minguo.”

E.g. ONTG92-B2737, field 245 “ta-kang” should have been input as “ta kang”. Because of the error, it converted into “da gang” instead of “dagang”. But the record has marker 987×dc.

Record with 987 marker ×dc carries with it unconverted subfield(s)
Some records are marked “correctly converted” but the subfield(s) is/are in fact not converted.
E.g. ONTG1035791-B
Chung kuo without a hyphen was converted into Zhong guo while it should be Zhongguo.

E.g. ONTG106669-B
The conversion program did not detect the subfield containing “jén” which was not converted.

E.g. ONTG98-B11463, field 245 “Diquan” should be converted into “qi quan”

E.g. ONTG98-B1905, field 260 “Hui-pei” should be converted into “Hubei”

E.g. ONTG98-B6192, field 440 “Gu” should be converted into “gu”

Records with 987 marker ±dr but in fact there is no error in the record
E.g. Records contains foreign language
ONTG90-B2457 group contains Tibetan and English.

Mixed Text
When a string contains a “Mixed Text”, the conversion program will not convert it. The “Mixed Text” may include:

- Presence of a “foreign syllable” (other than those in the exclusion list: i. e., et al, et al., etc., etc.)
- Presence of an English string where there is no punctuation ‘:’ and ‘=’ or there is a space before “;”

E.g. ONTG91-B6152, field 500 (1) “By Ch’eng Ying-liu” should be converted into “By Cheng Yingliu”

E.g. ONTG91-B6152, field 500 (2) “Pei-ching ch’u” should be converted into “Beijing chu”

E.g. ONTG99-B13501 “Collins Ying Han Shuang … “ English mixed with Wade-Giles.

E.g. ONTG92-B2737, field 260 “[Taipei?]” is treated a mixed text and will not be converted into “[Taipei?” The record has marker 987±dc.

E.g. ONTG180819-B, field 245 “A Q ch’eng ch’uan” is treated as a mixed text and will not be converted.

- A typo in the original Wade-Giles string
- A Pinyin word within a Wade-Giles string

E.g. ONTGS48244-B
Typographical errors
When there is a typo in the string, the conversion program will not convert the Wad-Giles right after the typo. The Wade-Giles string before the typo in the same string will be converted.

E.g. ONTG97-B9648, field 245 field: “pein” should be input as “pien”. It should be converted into “bian”.

E.g. ONTG98-B273, field 260 “hsüehch’u” should be input as “hsüeh ch’u”. It should be converted into “xue chu”. The record has marker 9874dc.

E.g. ONTG33857-B, field 700 name should be input as “Feng, Yün-hao” with dieresis above u”. It should be converted properly into “Feng, Yunhao”.

E.g. ONTG548233-B, field 490 “wen I” should have been input as “wen cung” and would be converted into “wen cong”.

Diacritics
- Ayns, apostrophes
  If ts’an is input as tsan in Wade-Giles, it will be converted into zan, a completely different Pinyin romanization. For example, CHING TAI (it should be CH ING) will be converted to JING DAI, not QING DAI.

  If the original input omitted a diacritic, the resulting Pinyin will not represent the proper romanization of a character.

- Hàček
  The conversion program will not convert the string if the Hàček diacritic is used instead of the breve diacritics above u.

E.g. ONTG689888B

- Dieresis on u
  The conversion program will not convert the string if a dieresis is input when it should not be there. The record will be marked for review. Also the conversion program will not convert the string when the dieresis is missing above u.

E.g. ONTG98-B7052
ts’un was inputted incorrectly as tsun

Geographic locations Shi, Xian, Sheng
According to LC, the city, county and province in headings are usually capitalized. However, the program does not capitalize these three terms when they appear as part of a uniform title.

E.g. ONTG98-B10802
All the “shi” are not converted into “Shi”

503, and 546 fields
The conversion program will not convert 503 and 546 fields. Therefore, we need to edit these fields in each individual record.

E.g. ONTG1127523-B, field 503 is not converted.

Redundant Pinyin Title Entry
After the conversion process, there are records that now contain redundant Pinyin title entries.

E.g. ONTG89-B146, the record contains:
245 Liaoning ribao =‡bLiaoning ribao.
246 Liaoning ribao.

E.g. ONTG98-B9612, the record contains:
245 Zhongguo yu fa xue shi =‡bZhongguo yufaxue shi.
246 Zhongguo yufaxue shi

E.g. ONTG94-B2277, the record contains:
245 Zhongguo di tu ji =‡bZhongguo ditu ji.
246 Zhongguo ditu ji

Cataloguer-supplied imprint place
The conversion program will not convert a cataloguer-supplied imprint place.

E.g. ONTG89-B4243
260 [Peking] will not be converted.

Dynasty
The conversion of Dynasty ignores the dates in order to reduce the number of non-converted records. When we review our records in the future we should aware that there can be “date typos” for dynasty.

The conversion program converts dynasty reigns ONLY if they are in subfield y of field 650, not subfield x. If the dynasty is in a different subfield, it won’t be converted.

E.g. ONTG712673-B, field 650 “Sung dynasty” is input under subfield x instead of subfield y. If input correctly, it would have been converted into “Song dynasty”. The record has marker 987‡dc.

Quotation marks
Wade-Giles text using quotation marks instead of dieresis will not be converted.
Records without a language code “chi”
Only records with a language code “chi” are converted; otherwise they have to be manually reviewed. (AFTER all Chinese-language bibliographic records are converted, RLIN will schedule converting records with a language code of “JPN” or “KOR” or a “chi” in the 041 tag).

E.g. ONTG97-B3065 The record has language code “eng”. The conversion program will not convert it. However, it has “chi” code in 041 field. It will be converted in the next schedule.