

**ALCTS Non-English Access Working Group on Romanization
Draft report and feedback received**

ALCTS Non-English Access Working Group on Romanization

Draft Report (Nov. 24, 2009)

Comments requested by **Tuesday Dec. 8, 2009**. Comments may be sent to rr2205@columbia.edu and will be forwarded to the Working Group.

A. Introduction

The ALCTS Non-English Access Working Group on Romanization was established by the ALCTS Non-English Access Steering Committee to implement Recommendation 10 of the report of the ALCTS Task Force on Non-English Access:

10. Examine the use of romanized data in bibliographic and authority records. Explore the following issues (including costs and benefits):

(1) Alternative models (Model A and Model B) for multiscript records are specified in the MARC 21 formats. The continuing use of 880 fields (that is, Model A records) has been questioned, but some libraries may need to continue to use Model A records. What issues does using both Model A and Model B cause for LC, utilities, and vendors?

(2) Requirements for access using non-Roman scripts (in general terms -- defining requirements for specific scripts falls under Recommendation 2)

(3) Requirements for access using romanization

The Steering Committee charged the Working Group as follows:

Reporting to the ALCTS Non-English Access Steering Committee, the Task Force on Romanization will examine the current use of romanized data in bibliographic and authority records, and make recommendations for best practices.

In particular, the Task Force will review Model A (*Vernacular and transliteration*) and Model B (*Simple multiscript records*) for multiscript data in MARC records (<http://www.loc.gov/marc/bibliographic/ecbdmulti.html>) and how these models are currently used

in library systems and catalogs, including the Library of Congress catalog and OCLC WorldCat. The Task Force should consider the needs of library users for search and retrieval of items and the impact that romanized data have on searches. The recent addition of non-Roman data to authority records and how library systems are using these records should also be considered.

The impact on library staff, including acquisitions, cataloging, circulation and interlibrary loan, should also be considered, particularly in situations where staff who are not language experts may need to process materials and requests.

The task force should address the following questions:

- Is romanization still needed in bibliographic records, and if so, in which situations and/or for which access points? Should best or different levels of practices be adopted for romanization?
- Can model A & B records coexist in library systems? If so, should guidelines for usage be adopted?

Time frame: The task force should complete a report by: December 15, 2009.

The Working Group has discussed whether to recommend continuing the use of Model A indefinitely, adopting Model B now, or adopting Model B at some point in the future when certain conditions are met. Related questions are whether catalogers could stop adding romanized parallel fields for some scripts but not others, and whether some libraries could stop adding them for some or all scripts while others working in shared databases continue to do so.

B. Model A and Model B

Two different models for multi-script bibliographic records can be followed in MARC 21: Model A (vernacular and transliteration) and Model B (simple multiscript records). In Model A, original-script fields are paired with corresponding transliterated fields. These are coded as 880 fields at the end of the bibliographic record, but in public display (and sometimes in staff display, as in OCLC Connexion) they display next to the corresponding transliterated field.

Model A

245 00 भारतीय गौरव के बाल नाटक / #c संपादक, गिरिराजशरण अग्रवाल.

245 00 Bhāratīya gaurava ke bāla nāṭaka / #c sampādaka, Girirājaśaraṇa Agravāla.

246 1# #i Title on t.p. verso in roman: #a Bharatiya gaurav ke baal natak

260 ## नई दिल्ली : #b डायमंड पॉकेट बुक्स, #c 2007.

260 ## Naī Dillī : #b Ḍāyamaṇḍa Pōkeṭa Buksa, #c 2007.

245 00 香港經濟日報 = #b H.K. economic times.

245 00 Xianggang jing ji ri bao = #b H.K. economic times.

246 31 H.K. economic times

260 ## 香港 : #b 港經日報有限公司,

260 ## Xianggang : #b Gang jing ri bao you xian gong si,

500 ## Description based on: 1992年8月15日; title from caption.

500 ## Description based on: 1992 nian 8 yue 15 ri; title from caption.

Model B

245 00 भारतीय गौरव के बाल नाटक / #c संपादक, गिरिराजशरण अग्रवाल.

246 1# #i Title on t.p. verso in roman: #a Bharatiya gaurav ke baal natak

260 ## नई दिल्ली : #b डायमंड पॉकेट बुक्स, #c 2007.

245 00 香港經濟日報 = #b H.K. economic times.

246 31 H.K. economic times

260 ## 香港 : #b 港經日報有限公司,

500 ## Description based on: 1992年8月15日; title from caption.

In addition to descriptive fields, headings may also appear in paired fields in Model A.

700 1# Βενιζελος, Ελευθεριος, #d 1864-1936.

700 1# Venizelos, Eleutherios, #d 1864-1936.

A system similar to Model B was used in North American card catalogs. Non-Roman descriptive elements were transcribed in their original script, and a "Title transliterated" (pre-AACR) or "Title romanized" (AACR) note was added at the bottom of the card, with a transliteration of the title proper only.

When library catalogs were computerized, at first only Roman script could be used, so both descriptive and access fields had to be entered in romanization only. In the 1980s OCLC and RLIN began to introduce character sets for major non-Roman scripts, enabling catalogers to transcribe bibliographic data as it appears on the piece in hand. Since then libraries have cataloged material in available scripts with full romanization and varying amounts of non-Roman script data in parallel fields (Model A). The amount of non-Roman script data appearing in these

records varies, but an attempt at standardization is now in progress, as a task force put together by the Program for Cooperative Cataloging is working on new draft PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets. For scripts not yet implemented in OCLC, such as Tibetan, romanization remains the only option.

Model B is currently used in East Asian online catalogs, i.e. no attempt is made to “transliterate” English or French text into Korean or Japanese script. But Latin script is much more widely known and used in East Asia than CJK scripts are in North America, so the use of Model B for Latin-script publications there does not have the same implications that the use of Model B for CJK publications would have here.

C. Questioning Model A

In the days of the card catalog, catalogers were able to enter original script in catalog records (Model B). That option was temporarily lost after the move to online catalogs, but catalogers have now resumed entering non-Roman script in catalog records, although they do so using a different model and retaining full romanization as well (Model A). It can now be questioned why we continue to romanize purely descriptive data. The cataloging rules for many years have had a rule (1.0E1) preferring transcription in the language and script in which they appear for certain elements. The adoption of Model B would result in simpler bibliographic records and more efficient cataloging. Romanizing takes time and can introduce errors. Romanization systems vary from country to country, and even the standard romanization systems we are supposed to use in North America can be difficult to apply consistently, unfamiliar to native speakers, and sometimes controversial (Persian, Greek).

C.1. Different romanization standards

Romanization is problematic when viewed from a global perspective. In North America, the ALA-LC Romanization Tables are an established standard for library cataloging, but libraries elsewhere in the world are more likely to use the various ISO romanization standards or a national standard. Often, different standards result in very different romanized strings that may, at best, look strange (and, at worst, not be recognizable) to a user accustomed to what is done in another country. They can also wreak havoc with attempts to match records. And MARC 21, unlike UNIMARC, has no way of indicating in the bibliographic record which romanization practice has been applied. (MODS, the XML schema based loosely on MARC 21 does have a type attribute for indicating both script and transliteration that can be added to any element. While a MARBI proposal to do so in MARC 21 may prove that it is too difficult to implement this in MARC 21 in ISO 2709, it may be possible in MARCXML.)

C.2. Romanizing unvocalized scripts

For many languages, even experts differ on the correct romanization of many words. Hebrew and Arabic are generally printed without vowels in the vernacular, so there is a certain degree of uncertainty in romanizing many words. In principle, standardized romanizations are selected by consulting specified dictionaries, but even standard forms that can be easily determined may seem arbitrary or controversial. For the Arabic word *نفط*, the standard romanization used by LC is *naft*, but many Arabic speakers might prefer *nift*. Romanization is in a sense playing favorites. It values one legitimate pronunciation over other equally legitimate pronunciations.

An additional complication with Hebrew and Arabic script is provided by “partially vocalized” title pages, where the publisher has provided the vowel marks usually seen only in sacred texts or works for children who are just learning to read. These marks are not normally included in original-script fields in cataloging records, but vowels must be included in the corresponding romanizations. The vowels provided on Hebrew materials are usually accurate, but those on Arabic materials often do not correspond to the vocalization recommended by standard sources. The Arabic word for “index,” *فهرس*, is often vocalized as *fahras* on title pages, but the standard romanization is *fihris*. Current practice in Arabic cataloging is to normalize the vocalization and use the standard form rather than transliterating the vowels actually indicated on the piece.

Romanization errors can occur when the cataloger misinterprets the romanization rules or is not deeply versed in the grammar of the language. In addition, personal names and nonstandard dialect words are particularly problematic when unwritten vowels must be supplied, and it can be difficult or impossible to find an authoritative source – or any source at all – for a “correct” romanization for these. Forcing catalogers to guess in cases like these slows down the cataloging process and serves no clearly useful purpose.

Entire romanization systems can be problematic. The ALA-LC Persian romanization system is frequently criticized by Persian speakers who say no one who knows the language would ever search by current romanizations. Romanizing Persian with the same three-vowel system used for Arabic ensures that most Persian words borrowed from Arabic are romanized in the same way as they are for Arabic text, facilitating romanized searches across languages, but this vowel system does not reflect the actual pronunciation of Persian in a way acceptable to most Persian speakers.

D. Advantages of romanization

The prospect of adopting Model B raises several concerns. A number of advantages of retaining Model A and romanization have been suggested.

D.1. Users who cannot read the original script

It is often suggested that romanization can help staff and patrons who cannot read non-Roman script work with library materials in these scripts for various purposes (acquisitions, ILL requests, storage retrieval requests, assembling bibliographies).

In principle, romanization seems to be of limited use to library staff unfamiliar with a non-Roman script. If a staff member is handling an item in non-Roman script and cannot read the original script, how does the romanization in the bibliographic record help the staff member match the item in hand with the bibliographic record? The romanized text in the record will not appear on the piece. These staff will be more likely to look for an ISSN, ISBN or call number to match up a book or serial issue with a bibliographic record rather than trying to use tables to transliterate a non-Roman script they do not know (and even that would be impossible for non-alphabetic or unvocalized scripts). However, not all titles have an ISBN or ISSN, and items not yet cataloged do not have a call number.

At some institutions the romanized title (and romanized enumeration/chronology if present) is written on the title page as part of the cataloging process, so for items that are already cataloged staff can retrieve them from the stacks and match the romanization in the bib. record against the form on the title page to confirm that they have the right piece. Items in non-Roman script may also arrive from the vendor with information in ALA-LC romanization attached, allowing, for example, serials check-in staff to match new serials items to the correct bibliographic record in their catalog.

For non-Roman titles, citations are often given in romanized form in western publications, and users may come to the library with these citations looking for help finding the material cited. With romanized records, public service staff can help them even if they do not read the script themselves. But a public services staff person who does not know the script can do little to help such a patron beyond simply typing the data in as it appears, which the patron could easily do themselves. And while the Chinese or Japanese transliteration systems used in libraries may be widely used in non-library contexts as well, for other scripts there is no widely accepted romanization system and any romanized data provided by a patron is unlikely to be in the system used by the library.

Romanization provides additional access points for those who might prefer to use them. For Chinese or Japanese, some catalog users may be non-native speakers who can read the original script to a limited extent but are more comfortable with transliteration. And in some cases a romanized search may be easier to input than an original-script one, even for users who can read the original script (see section E below).

D.2. Collocation of forms romanized the same way

Romanization provides collocation when the same word can be written in different ways in the original language. For example, *Han'guksa* ("history of Korea") can be written 韓國史 and 한국사 in Korean; *Zhongguo yi shu* ("Chinese art") can be either 中國藝術 or 中国艺术 in Chinese. Many of our systems are not yet sophisticated enough to treat these original-script forms as equivalent in their indexing (although WorldCat uses CJK mapping tables that allow traditional-character Chinese data to be retrieved when simplified characters are searched, and vice versa). And no system can automatically replace non-MARC 21 characters in users' searches with the equivalent MARC 21 forms (as given in LC's CJK Compatibility Database) that catalogers have to use to represent them in bibliographic records. But a search for the romanized form retrieves all these variants.

In Hebrew, many words can be written either with extra consonantal letters to flesh out the normal lack of vowel representation (full orthography), or without them (defective orthography). Without the item in hand, a librarian or patron cannot guess how many consonantal letters to include in a non-Roman search, and if the phrase to be searched includes several words which can be written more or less fully, the number of non-Roman searches needed to cover all possibilities can be quite high. The family name transliterated *Rozenberg* may appear as רוזנברג, ראזענבערג, רוזענבערג, or any of a number of other possibilities. *Yerushalayim* ("Jerusalem") may be spelled ירושלים or ירושלם, and the name *Aharon* may appear as אהרן or אהרון. The Hebrew or Yiddish spelling of a foreign name like "Lakewood, New Jersey" is even harder to predict. Catalogers transcribe these in original-script fields as they appear; they do not "normalize" the non-Roman spelling to one system, or enter multiple variants to account for possible spellings other than the one actually used. The presence of a romanized field which corresponds to all possible original-script orthographies provides a "normalized" spelling so that all variants are retrieved when a romanized search is performed. (But sometimes romanization has the opposite effect; see the end of section D.4. below.)

In Arabic and Hebrew, prepositions and the definite article are prefixed to the following nouns. The combination is presented as a single word in the non-Roman script. In ALA-LC

romanization, such particles are separated from their nouns by hyphens which have no equivalent in the non-Roman script. Thus a romanized search for the Arabic word *taqrīr* (“report”) will retrieve both records containing this word without an article (romanized as *taqrīr*) and records containing it with an article (romanized as *al-taqrīr*). The corresponding non-Roman forms (التقرير and تقرير) are indexed as single words and have to be searched separately.

D.3. Sorting

Doing a browse search for romanized text produces an alphabetical list of results in the OPAC that the user can scroll through with the expectation that specific results, if present, will be in predictable locations. Browse searches also appear to work well in most systems for the major non-Roman alphabetic scripts (Cyrillic, Greek, Arabic, Hebrew). But culturally-sensitive sorts have not yet been developed in library systems for non-alphabetic languages and scripts. For CJK, sorting by code point (the current effect of a browse search) does not produce acceptable results. The sorting orders that would be meaningful to native speakers are by radical and stroke number, or by Latin transliteration. The former would be difficult to implement; romanization provides the latter. However, in an online environment where many users increasingly rely on relevance ranking of keyword search results, this may not be important enough to be a deciding issue.

D.4. Added value

For some languages, romanization requires the cataloger to provide information about the standard pronunciation of script forms that are pronounced differently in different contexts. For example, romanization requires the cataloger to determine and indicate which of the many possible readings of a Japanese character is correct in the case being transcribed, for example whether 中 is pronounced *naka* or *chū* in a given context. (Japanese online catalogs such as NACSIS also indicate pronunciation, although they use Japanese syllabic characters rather than romanization to do this. In the NACSIS record for the title 日本漢学文芸史研究, the title proper is followed by its pronunciation spelled out in angle brackets: <ニホン カンガク ブンゲイシ ケンキユウ>, as is the corporate name in the added entry for the issuing body: 東京教育大学文学部 <トウキョウ キョウイク ダイガク ブンガクブ>.)

This effect of adding romanization to bibliographic records can be seen positively (providing “added value” by giving extra information about the readings of original characters) or negatively (sometimes Japanese catalogers need to spend a considerable amount of time researching the correct “readings” before they can enter them). For CJK, providing pronunciation-based access points can be useful for users who know the basics of a language but are not fully proficient in the

original script. From a public services perspective it could be a disservice to users to stop transliterating, especially for undergraduates or beginners, or researchers whose are not experts in these languages but need to work with materials written in them and have some ability to do so.

Although this sort of information is helpful to some users, it is not clear whether providing it should be seen as an essential function for a cataloger. It would certainly be simpler just to transcribe the original script as it appears on the piece. And users who search using romanization may sometimes have to do separate searches to account for differences in pronunciation in text strings that would be retrieved by a single search done in the original script.

E. Systems that do not support non-Roman script

Records with non-Roman script only are useless in systems that cannot handle non-Roman script at all, and while these are now probably rare in academic and research libraries they still be more common elsewhere.

A 2007 Cataloging Distribution Service survey related to character sets in MARC records found that a significant portion of their subscriber base was not yet able to handle UTF-8 records, i.e., they were generally limited to non-Roman scripts that are part of the MARC-8 repertoire (Chinese, Japanese, Korean, Arabic, Persian, Hebrew, Yiddish, Cyrillic, Greek). The differences vary system by system and may only be related to certain facets of a system functionality (e.g., import, export, input, display, indexing), some of which a system may accommodate while others it may not, making it all the more difficult to define “support” in this environment.

Another question is whether the software packages that libraries provide for creating bibliographies, labeling software, openURL resolvers, electronic resource management systems, etc. are able to deal with Unicode characters.

In addition, many languages and scripts outside of the MARC-8 repertoire of UTF-8 are currently impossible to input into LC’s local system due to a bug that renders their Microsoft IMEs unusable. Even if other libraries become UTF-8 compliant, romanization will be the only way to enter and distribute those records for some time to come.

CJK has some characters not covered by Unicode, so it is not yet possible to transcribe original script in every case.

Public library terminals (or users' personal computers) may not always allow non-Roman script input. Even if input is supported, different users might need a variety of keyboards, depending on what input method they normally use. For simplified-character Chinese, there are four different kinds of keyboard available in Windows. For traditional-character Chinese, there are eight. Not all of these may be available, even on the library's own terminals.

In some cases searching by script alone is completely unsatisfactory because of flaws (sometimes major) in the Microsoft IMEs used to input records or users use to input searches. The Microsoft Farsi (Persian) IME lacks some common and necessary characters, which must be created by workarounds in cataloging and cannot be input at all by searchers, so searching in non-Roman for strings containing these characters will always be largely unsuccessful.

F. Headings

Since headings are established in romanized form in the LC/NACO Authority File, they need to be entered in bibliographic records in romanization. Name headings may now have original-script references in the authority record, but subject headings do not. Current practice allows and sometimes requires the addition of parallel heading fields in bibliographic records in original script. For example, in current PCC documentation (now under review), parallel original-script heading fields are required for Arabic CONSER records with original-script descriptive fields, but for CJK CONSER records with parallel descriptive fields, parallel heading fields are optional.

Parallel fields for headings in bibliographic records are still necessary for keyword searching on script names. They are also necessary for a complete display in script of the basic bibliographic description for users who are unfamiliar with the romanization used (for example Cantonese speakers looking at Chinese records, where romanization is based on Mandarin pronunciation). They are essential when the romanized form is ambiguous, as it is for Chinese names where any romanized form could correspond to multiple names written with entirely different characters.

For some language/script cataloging communities, current guidelines attempt to ensure that headings are entered in a form that "corresponds" to the authorized romanized form, but there are still problems that prevent complete standardization. The same authorized romanized form may correspond to more than one original-script spelling (Ивановъ or Иванов for *Ivanov*, 中國 or 中国 for *Zhongguo*), and different practices exist for cataloger-supplied qualifiers (entered in the authorized romanized form, or in a "corresponding" original-script form, or omitted; this is a particularly difficult problem for right-to-left languages). So original-script headings, unlike romanized ones, are never completely consistent, and result in split indexes in the catalog.

One of the perceived advantages of adding non-Roman script references to authority records was the hope that, if they were added, it would no longer be necessary to provide non-Roman parallel access points in bibliographic records. However, when the project was undertaken to prepopulate the LC/NACO Authority File with non-Roman headings from OCLC, it became clear that providing users with full access to records without parallel fields for headings can work only if authority data is fully integrated into the searching process. If the system in use does not fully integrate authority data (and many systems do not), then access is lost if parallel fields are not maintained.

G. Automation of romanization

The effort required to provide romanization in bibliographic records can be reduced by automation, although some human checking for variations will always be necessary. LC has in production or is testing automatic transliteration for every language/script they provide except for Japanese. Conversion from original script to current transliteration schemes can be automated fairly easily for Cyrillic and (with the exception of the rough breathing) for Greek. Chinese and Korean transliteration tools are also being used or in development, and LC is hoping to identify groups that are interested in collaborating on Japanese transliteration along the lines of the recent efforts for Korean.

For scripts where vowels are not indicated in normal orthography, text cannot be automatically romanized from the original script. However, Arabic and Persian can have original script automatically reconstructed if the romanization is entered first. Since their romanization schemes match their scripts nearly character by character, automatic tools can be designed which make few errors. Hebrew presents more problems, since the romanization system contains many ambiguous signs and Hebrew orthography is not fixed. An automated process will not be able to tell from a string of text romanized according to the current ALA-LC Hebrew tables whether the publisher chose a "full" or a "defective" orthography for the original script.

H. Models A & B in one catalog

Models A and B can coexist in one catalog, and already do in OCLC WorldCat where many non-Roman-only records have been loaded by vendors and from the National Library of Israel. Non-Roman searches will retrieve records created under both models. But if libraries adopt Model B for future cataloging, their catalogs will still have (in addition to the existing Model A records) a large number of older bibliographic records cataloged using romanization only. In addition, there are countless records for Western language works containing romanized non-Roman words in their descriptive fields and headings. It would be very difficult to add non-Roman script to those

records. To convert them manually would require extensive resources, and for many scripts automated conversion would not provide even approximately correct results. If the records are left unconverted, original script searches would not retrieve pre-Model A records, and romanized searches would not retrieve post-Model A records. We will have permanently split catalogs for our non-Roman script materials.

I. Recommendations

1. A majority of the Working Group believes that the factors discussed in this report are significant enough to make a general shift to Model B in bibliographic records premature at this point. Some members of the Working Group feel that having romanized access points in records provides enough added value that their use should be continued indefinitely. Others believe that in an environment of shrinking staffs and production pressures we should anticipate future developments in making our decision and recommend a move to Model B sooner rather than later. However, most believe that although a gradual move towards the use of Model B for current cataloging is probable, we should continue current practice for some time longer as we prepare for the transition.
2. Further research is needed into the remaining obstacles so that we can identify decision points that will allow us to move beyond the status quo. We recommend that ALCTS sponsor a survey of libraries and library systems to better understand the status quo and possible future directions from a technical perspective.
3. Automatic transliteration software should be utilized to reduce time needed to create the romanization, when possible.
4. The amount of romanization in records could be reduced by limiting it to fields including key data for access (titles and headings).
5. Since different languages and scripts raise very different issues, some language/script cataloging communities may decide to move to Model B sooner than others. A coordinated decision to change practice within each community would be preferable to individual decisions to implement Model B in different libraries at different times.

Feedback received, Nov. 25-Dec. 9

1.

I agree with Recommendation 1:

"1. A majority of the Working Group believes that the factors discussed in this report are significant enough to make a general shift to Model B in bibliographic records premature at this point. Some members of the Working Group feel that having romanized access points in records provides enough added value that their use should be continued indefinitely. Others believe that in an environment of shrinking staffs and production pressures we should anticipate future developments in making our decision and recommend a move to Model B sooner rather than later. However, most believe that although a gradual move towards the use of Model B for current cataloging is probable, we should continue current practice for some time longer as we prepare for the transition."

As a video cataloger with only formal training in Romance languages, but who regularly works with non-Roman materials, I am not ready to give up romanization. I can interpret romanized text well, but it would be impossible for me to work in the original scripts, with the exception of Cyrillic. Furthermore, I still see plenty of video materials that have no identifying nos. on them.

----- Original Message -----

Having romanization in bib. records helps because often the acquisitions records (usually just includes title, possibly director and publisher) will be romanized. This gives me an access point to search on if there are no identifying nos. on the item. For example:

245 : Legenda o Tile (DVD PAL)
for the film Легенда о Тиле

I suspect the Acq. Dept. gets the title info from our selectors.

Note that [...] University has Russian, Arabic and 2 East Asian catalogers, but I only consult them when I cannot ascertain the descriptive elements myself or obtain them from another source.

Also, romanization helps in that I can determine the roles in the statement of responsibility. For instance if I see the following:

Ze dong dian ying gong si chu pin ; bian ju Jin Yong ... [et al.] ; dao yan Liu Zhenwei

I can determine that the first entry is for a corporate body (because of gong si), that the writer was Jin Yong (because of bian ju) and the director was Liu Zhenwei (because of dao yan). I really don't think that I could make the same determination without the romanization.

I hope that my explanation helps. Thanks for asking for my input.

2.

Ideally, all catalogs and cataloging should be in the vernacular. Unfortunately, I do not believe that the time has yet come for this, especially in an era of greatly reduced budgets and staffing. Even large institutions are not able to provide the access and support to non-English materials that they should.

For the time being, I would recommend that the Romanized forms be maintained, at least as alternate access points. My two primary reasons are as follows:

Even in an online environment, Unicode is required for appropriate display. This could mean no catalog access for significant portions of an institution's collection.

At "my library system, we do not yet have Unicode and our highest circulating non-English materials are in Russian and Chinese (Spanish has larger communities, but is a very distant third).

Most libraries do not have sufficient staff with a fluency in even one, let alone many non-Roman languages. Romanization is a boon to staff.

Romanization is the only way that the majority of staff can cope with these materials -- from entering barcodes to reader's advisory, to issues with circulation and fines.

3.

I don't work directly with non-roman script materials, but roman script catalogers cannot ignore the presence of non-roman script cataloging (and vice versa) in the context of the library catalog. I agree with the recommendation that the cataloging community stick with plan A. (parallel fields) for now. My problem with plan B (non-roman script only)? Roman script catalogers are generally

tasked with cataloging translations and roman script history/criticism of non-roman works. How do we propose to handle uniform titles and subject headings if the work cataloged is in roman script and the record for the original is non-roman? Without a heading and title in roman script, how do we determine shelf arrangement when cutting around call numbers assigned to records that are entirely in non-roman script? Isn't this a FRBR related issue? By which I mean, a certain category of bibliographic records cannot be viewed in isolation from the other bib. records in a catalog. There has to be some way that all record relationships can be made transparent to users with different language backgrounds--I don't see how this is possible without some linguistic crosswalks. Another thing to consider is the impact on public libraries. I think the reality is that they will need to collect publications for their various and changing linguistic communities, and few of them are going to have the wherewithal to train and employ specialist catalogers and acquisitions staff--they will rely on copy cataloging. But consider that if they are using vendor supplied transliteration, how will they be able to come close to matching an item to a record in the absence of a control number? What about videos in non-roman script languages? Would we want all the access points and description to be roman script only?

4.

Romanization is very definitely needed in catalog records for non-Latin script materials. Without it, one would not see these materials in a list while browsing the catalog, and students who are still learning the non-Latin scripts are aided by it. On top of that, it is still faster and more convenient to access these records through romanization because we all have QWERTY keyboards.

As for WHICH romanization system to use, we must bear in mind that romanization was invented to help non-native speakers learn the language. They are not served by systems which mangle the language's pronunciation. For example, the Japanese word "tsuchi" is best spelled, as I have just done, according to the Hepburn system. It does no good to dogmatically spell the syllabary as "ta ti tu te to" when this causes the student to pronounce "tsuchi" as "tooty (tuti)". I once hiked to the top of Mt. Fuji, but I have never seen Mt. Huzi.

Of course, when sounds exist in the target language that are not in the student's own, and when simplification is the goal, we are going to end up with spellings like pinyin's xi, ci, and qi, whose correct pronunciations are impossible to guess without help. I do prefer pinyin to the diacritic-laden and inconsistent Wade-Giles. The biggest problem for me is Korean, because the pronunciation of a consonant will vary depending on context. The national romanization system is

no help here. Its spellings are consistent, but often do not give the correct pronunciation. On the other hand, the McCune-Reischauer system is horribly complex and hard to follow. It gives the reader a good idea of correct pronunciation, if he has been schooled in the meanings of the diacritics, but it is very difficult for the writer to use.

I do not want to see romanization go away, by any means, and I wonder whether it is possible to devise software that will correctly romanize East Asian scripts according to context so that, for example, "kore wa pen desu" will not be rendered "kore ha pen desu" and so Korean words will always come out in correct McCune-Reischauer.

5.

First of all, what a great report. Many thanks to you and your committee. I loved the perspective of the historical background and how concise and clear the whole thing is. Really, really well written.

A tiny comment on recommendation 3: automatic transliteration can save time in either direction. It can be used not just for the creation of roman fields, but also for creating non-Roman fields.

A larger issue from section D1, paragraph 4: Public service users try to help library patrons with citations or searches for non-Roman materials. The argument is that even with romanized data, the staff person can do little more than retype the search, which a patron can easily do for themselves anyway. I don't agree. There is way more going on at the reference desk than that.

First, there is the whole issue of how much self service is too much. I am not always comfortable selling the idea of convenience and independence for the patron, when what is really happening is that we are cutting back service to save money or we don't have the resources or expertise to provide services in the first place. One of the issues with our DSpace repository is that we just upload data in whatever the original format is, and we do not provide tools or assistance to users who may wish to use that data. That's their problem, so far anyway. I don't want to abandon non-Roman users in the same way.

But say that we believe in providing reference service to users. They may come in person, or they may be asking for help remotely. Even when the search is for English language material and everyone involved is a native speaker of English, the library staff person is doing *much, much* more than just typing in the search again!

The staff's ability to reproduce a search and see what is going on, and make suggestions is pretty essential. Dealing with an unknown language (or even figuring out keyboarding when you do know the non-Roman script!) is a barrier. Romanization provides a kind of lingua franca (even given some of the shortcomings and controversies your report rightly highlights). Even our East Asian librarians sing the praises of romanization, much to my surprise, in the context of library services to their East Asian patrons. All the technical shortcomings that your report points out that are still holding us back with respect to non-Roman data and UTF8 in our systems also come home to roost at the reference desk as well.

6.

I agree with the draft recommendations - and in particular with the 1st sentence of the very first recommendation (which seems to me to be crucial to the whole discussion).

I hope the WG members won't be too upset if I offer a little criticism (and no, that's not why I'm mailing this on a US public holiday!!). The charge from the SC was quite specific - that you should examine use in both bibliographic AND authority records. The report pays only limited attention to authority data, and the concentration on bib records is apparent even in the subject line of the e-mail to which I am responding. Whilst there are doubtless very good reasons for trying to break the issues down into bite-sized pieces, I fear that the logical conclusion would be that nobody ever gets to look seriously at the relationship between the two. I wasn't a great fan of the implementation decisions taken for NACO non-Roman work - I would have much preferred alternate forms (i.e. scripts in this case) to have been entered in 7XX

fields, not 4XX., from which you could start to build better (proper?) relationships between authority and bib records' use of non-Roman data. But my point, in the context of your draft report, is that concentrating so heavily on bib data runs the risk of overlooking how information provided in authority records *might* help address some of the issues which you've (correctly, IMHO) identified as requiring more attention before we can start to make changes.

Perhaps I'm criticising you for failing to do something you never intended to do in the first place, but there seemed to me to be a gap between the charge and what you are planning to deliver. And I'm concerned that someone needs to address this (if not your WG, then you might reasonably ask the SC just who will).

----- Original Message -----

One relatively painless way out (!) might be to own up to the fact that you paid more attention to the things you paid attention to, but to justify that by saying that the relationship between authority and bib records is almost entirely a "systems" issue, and that recommending ways for exploiting the NR data now appearing in NACO records, whilst relevant to the work of your group, might be better tackled as a change to the SC on Automation.

This wouldn't be quite as much of an escape trick as it might be from the above brief statement. The decision on including NR data in NACO records was carried out pretty well independently of any consideration on how this data might be exploited. You're probably the first group to come up against this, and you can't really be expected to operate in a vacuum. Someone needs to start looking at how to make effective use of the growing amount of NR data in NACO (although, to my mind, we should have been considering what we wanted to do with the data *before* we decided how it should be represented in MARC terms - which is why I cling to my belief that we collectively backed the wrong horse in going for the 4XX approach, as it won't be able to deliver anything like the same sort of functionality I could envisage using 7XXs).

I wasn't aware that there was a PCC group looking at implementation and best practice for that NR data, but even that, I feel, misses the point that I'd hoped your WG might have tackled (but chose not to - and probably rightly) - let's call it the "exploitation" of that data.

Not sure if this helps (other than to soften what I realise, on re-reading, may have sounded more critical of your work than I'd intended at the time!).

7.

**Comments on the ALCTS Non-English Access Working Group on Romanization
Draft Report (Nov. 24, 2009)**

I would think the status quo works quite well in the North American setting where the catalogue information has to be used by people who know the original scripts and also by people who do not know the scripts. It is very true that in East Asia, romanisation does not serve much purpose, since East Asian characters take over there.

I have been using RLIN and now OCLC CJK for EA cataloguing. Their manuals usually suggest that, in a record, wherever original characters are used, their parallel romanised form need to be supplied in a parallel field. Thus, we have two 245s, two 490s + 830s, two 500s, etc.

The above practice seems to be able to cater to needs in most situations, regardless of the limitations or potentiality of the local systems used. Take the current display on WorldCat.org as an example: the notes fields (5xx) only display in romanised form, even though the full record in the WorldCat database contains the equivalent original scripts as parallel fields. If in the OCLC record no romanised parallel field is used together with the original script (that is, there is only a field for the original characters), then the display in WorldCat.org will not even show anything for that field.

The following 505 (contents) is taken from the WorldCat First Search which is different from WorldCat.org. First Search displays all fields:

目次: 引言 -- 第1章. 二王尺牘之特徵 -- 第2章. 二王尺牘之東傳日本與日本書紀之編撰 -- 第3章.
「東天皇敬白西皇帝」國書新考 -- 第4章.

日本「天皇」名號之成立與日本書紀所載國書之關係 -- 第5章. 結論 -- 跋文.

Yin yan -- di 1 zhang. Er Wang chi du zhi te zheng -- di 2 zhang. Er Wang chi du zhi
dong chuan Riben yu Riben shu ji zhi bian zhuan -- di 3 zhang. "Dong tian huang jing
bai xi huang di" guo shu xin kao -- di 4 zhang. Riben "tian huang" ming hao zhi cheng li
yu Riben shu ji suo zai guo shu zhi guan xi -- di 5 zhang. Jie lun -- Ba wen.

8.

It is good to learn that ALCTS Non-English Access Working Group on Romanization is working hard to present its final report to the Steering Committee very soon. It may be interesting for the Group to note that this company, D.K. Agencies (P) Ltd., New Delhi are originally a bookseller (Estd. 1968) but have been cataloging publications from India since 1973. So, we would like to share our experiences and efforts in regard to cataloging of Indian/South Asian vernaculars. At DK, we

- have been using Model A (vernaculars as well as transliteration)
- strictly follow ALA/LC Romanization Tables
- enter the transliterated form into the database
- automatically generate the original scripts of tags 245, 250 and 260 for tag 880 in respect of the following languages
 - o Assamese
 - o Bengali
 - o Gujarati
 - o Hindi/Nepali/Newari
 - o Kannada
 - o Malayalam
 - o Marathi
 - o Oriya
 - o Panjabi
 - o Sanskrit
 - o Tamil
 - o Telugu
 - o Tibetan
 - o Urdu

- use a software developed in house for the above conversion
- generate the original script of tag 100 too, if asked for by any library

Incidentally the Hindi book example listed in Model A of the draft report is a creation of DK only (liNdDKA)

We have also started developing softwares to support auto transliteration (ALA/LC Romanization) from the original scripts of Indian languages. So far we have launched Hindi and Tamil language scripts only. Relevant screen shots are sent herewith.

You may like to make use of this information while finalizing the report. In case anybody has any questions please contact us off-line at docinfo@dkagencies.com

9.

Congratulations on a difficult job well done; I have been following this of course, as one of the writers of the original charge.

My own views would differ more in emphasis than content; some issues I find most important are mentioned only in passing, and would have liked to have them get a slightly stronger statement; but I am sure that's always the case.

However, there is one perspective, that of the user's need in certain languages, which I find has not been addressed sufficiently; and one possible alternative to romanization which has not been mentioned at all.

One major reason users have for checking catalogs is to find material cited in bibliographies in Western sources, works on people mentioned in newspapers, etc. The reality is that most of these bibliographies do not include original scripts; and of course, no newspapers do. Now, for some languages, the step from a transliterated title (whether using the library or other transliteration standards) to original script is relatively small, given the training such users would have received; even a relative beginner in Arabic, Hebrew, or Russian, given any original title or author, could search a catalog using the original script with minimal guesswork. However, in other languages (CJK) this is completely impossible: even an expert would not know which characters are behind a particular transliterated name; if s/he were required to search in an original-script only database, he would first having to go somewhere to any work which includes both (and gives enough information to decide between the many alternatives). Currently, such a user would go to

the catalog first, even if his final search will take place elsewhere, in a database, since of all possible reference works, a catalog using Model A best solves this issue. If model B were to be adapted, one would have to use other reference sources, each of a much more limited scope, to go from a transliterated author or title to the original script form.

Indeed, in this reference librarian's perspective, a frequently received complaint from users about using the increasing number of electronic resources originating in China, Japan or Korea, is that these are not usable for beginners and intermediate users at all, and can be used by expert users only in conjunction with other reference works (or, in practice, the current catalog!) At a recent session of teachers of Japanese subjects with librarians, this issue was paramount: they wanted librarians to lobby with Japanese database providers to make romanized access available.

And indeed, some providers of such databases, therefore, if they wish to increase their sales to the US market, increasingly add phonetic readings to authors and titles to provide increased access; after all, the problem also to some extent exists for users in those countries. "Phonetic" in this sense does not necessarily mean in romanized form: romanized pinyin is common for Chinese, but for Japanese kana readings are more frequently used, and hangul ones for Korean. And indeed, library catalogs in those countries sometimes chose similar solutions, since also in those countries there is a need to go from reading to characters, and to find all different spellings for the same word in one search. It would be ironic if the increased demand for such added entries in such newer media and in those countries would be paralleled with their abolition in the US, where the need is even greater.

One option which could be investigated, therefore, is whether for Japanese and Korean the reading access generally provided in those countries (kana and hangul respectively) could replace romanization in the strict sense. In a world where exchange of records with those countries is increasingly common, it is much more likely to find Japanese records with added kana, and Korean records with added hangul fields, than romanized fields. (For Chinese, romanized pinyin remains the preferred way.) If library systems could be updated in such ways, perhaps both access for Western end users, the need for which cannot be underestimated, and exchange with records created in other countries, can be served.

10.

Thank you for a good analysis of the problem, and solid recommendations. I concur with your recommendations, especially I.1—to maintain the current practice of model A as we prepare for

the transition to model B. You have adequately identified the main problem in section H, that a complete switch to model B would permanently divorce any records which have been entered without the original script (model A without the parallel fields). I believe, as you have indicated, that this requires a careful review, lest we end up with permanently split catalogs for non-Roman script materials.

One consideration that I did not see addressed was the need for smaller libraries without specialized language skills to enter records into WorldCat. It is increasingly common for public libraries to purchase materials in non-Roman scripts, because their patrons request them. While the libraries may not have language-capable catalogers, they will quite likely have patrons who can provide a transliteration of the data from the material in hand. The cataloger (who may have no idea how to use an IME) can then easily enter the item into WorldCat. If we drop model A completely, we would have records with “[Title in Japanese]” in the 245—an old practice to which I do not wish to return. A basic record in romanized form (following model A) can then later be upgraded by a library with additional language capabilities.

11.

Thank you for your diligent pursuit of the Romanization issue. I think the Romanization should always stay. In my judgment, the vernacular is only a helping tool to add more access points to the record and to clarify some of the ambiguities in Romanization. At any rate, due to the absence of the vowels in the Arabic script, Romanization gives clearer idea on how the words are pronounced. On the other hand for classification and indexing purposes the vernacular script “only” will create unsolvable problems. There is no need for the vernacular in 300 500 and 650. Otherwise, all the Romanized fields can be turned to the vernacular in OCLC with the Romanization button with minimal editing and only sometimes. As for the dates, etc., I do not think this is a problem. The dates in the vernacular are in the Arabic numerals (i.e. western numerals). Even the Indian numerals in the record are turned to western numerals. Our own materials are not segregated by languages as is the case in some ME libraries. Also, for NACO purposes we are now adding the vernacular to the records with dates and abbreviations. Things seem to work well so far. Having said this I am with model A. In our last MELA conference no one mentioned anything about getting rid of Romanization. As for saving money, I do not think that much money can be saved by getting rid of Romanization; the money saved will be spent on typing the vernacular. To my knowledge there is not yet any technology which can scan the info from the book and transfer it to the bib record. As I said: Progress and civilization cannot be built

12.

Thank you to you and to the Working Group for the draft report on romanization. We agree with the conclusions of the report and would like to submit suggestions for your consideration. Under the section on Model B in section B, it would be more useful not to include examples with publisher-supplied romanized data or parallel titles in English. In other, words it would be best to keep the examples as simple as possible. Also, including full records as examples for Model A and Model B would be more useful than including portions of records.

13.

Romanization has the added benefit of allowing acquisitions personnel to communicate with publishers if there are problems with the book coming in, being damaged, etc. If they have a question, they can either type or even say (on the telephone) “Do you have Bhāratīya gaurava ke bāla nāṭaka?” without needing to be an expert in Hindi. If all they have on their page is “भारतीय गौरव के बाल नाटक”, it makes it more difficult to communicate.

14.

Overall a very interesting report. These issues are not easy and I do not think there is yet an obvious solution.

Overall, parts of the collection where issues of romanisation occur fall into:

- materials entirely or primarily in the language which is written in a non-Roman script. In these materials the title page and contents are in the script, and chances are that name access points which are relevant would have a preferred form in that script. A library that has a substantial collection in a language or script is likely to have at least some expertise available.

Bilingual or multilingual materials, which may have parallel titles or may not. These materials may also be bought by libraries who are primarily interested in the roman script content. I have

seen materials with a parallel title in non-Roman script, but essentially no text in that script (such as translations), or whose content is mainly pictorial (exhibition catalogues).

Unilingual materials in a Roman script language where relevant access points have been established in romanisation. Such materials include translations of works, works about people and events, works sponsored/created by corporate bodies. Such materials may be held by libraries which have no collections written in the relevant non-Roman script. Currently rules such as 22.3C2 choose as a heading a form in common usage in the language of the catalogue, if headings are instead established in the original script, the library may find that the headings on such records are the only part of their catalogue which is in a non-Roman script.

The needs of libraries with these different collection profiles need to be accommodated, particularly with respect to the copy being made available for derived cataloguing.

C.1 and C.2: excellent points.

D. Another issue is display: not all systems can display characters (both original scripts and the diacritics needed in certain transliteration schemes) adequately, particularly web OPACs, and this can be an issue both within the library and with user's own PCs. For example, the empty boxes found in some LC authority records.

Another use for transliteration is cuttering: chances are that the author/main entry cutter still needs to be done in romanisation, as I cannot see shelving staff sorting the books otherwise.

D.1 another category of use of transliteration: Yiddish text in transliteration can be quite intelligible to readers which some knowledge of German, who cannot read it in Hebrew script. This can be useful for cataloguing staff. Often the transliteration is most useful to cataloguing/database maintenance staff who are working with bibliographic records to complete tasks not requiring a knowledge of the language and not with the documents themselves.

D.2: The searching of words joined to articles is also an issue for other languages. In French for example, l' is an article causing keyword searching problems for many words.

D.3: In a previous position I worked with an ILS that provided both stroke order and pinyin order sorts for Chinese. We implemented pinyin, but it is not 100% satisfactory as the sort can only recognize a single pinyin value for each character to sort on, thus some characters are not sorted where users expect in certain titles. Also, as several characters can have the same pinyin, they may be interfiled in the browse sort.

E. A review of what it means for a system to support a script may help. It needs to include both staff and user input of the script (for record creation and searching), display, printing, indexing, import/export of records, and products. When a script is fully supported, it should not be a second class citizen in any respect. This is almost never the case.

Paragraph 1: This report should consider a wider range of libraries and collections than just major academic and research libraries. I suspect that very many public libraries or school do not have such support, and that many others may be using a system that theoretically supports a variety of scripts, but that they do not have the expertise to set them up.

F. non-Roman script access for subject headings (if we mean topical subjects, 150/650) is an entirely different issue than for name headings. What would be needed is an actual subject scheme in the language/script in question. Other than the display issues, the effort is similar to adopting any additional subject scheme.

The point made that parallel fields in bibliographic records are generally still needed to allow keyword searching in the script, even though the non-roman references in authority records should replace them for browse searching, is well taken and easily forgotten. Another issue is the interaction with authority control systems, as when the parallel bibliographic field matches an authority 4xx, the system may try to "correct" it with the authorized, transliterated version which is already present.

H. It may not be desirable to all libraries to add script headings to records for western language works which happen to have a name heading that might exist in script. It might prevent any access at all in a small public library to their books about Mao Zedong for instance. If we do this for country names, it would be quite dramatic.

Recommendations

1. I agree that a shift to Model B is premature and that romanization of the name access points (establishing them in romanization) is added value. I also think that title proper in romanisation is worthwhile. Maybe it is less needed for some quoted notes.

3. Automatic translation is good, but realistically will be used by large libraries with large collections in the language in question. Small libraries, or when dealing with a language collection that is small for the library, will not see this as a solution.

4. Yes, we will need to collectively decide what data elements to transliterate once we abandon the idea that the transliteration is the full record.

5. I agree that a phased approach will be wise.

15.

Hi, good report. Just one minor thing about the sentence:

".2. Romanizing unvocalized scripts

For many languages, even experts differ on the correct romanization of many words. Hebrew and Arabic are generally printed without vowels in the vernacular, so there is a certain degree of uncertainty in romanizing many words."

This is a minor pet peeve of mine :) In most cases it is only the short vowels that are not vocalized. The long vowels are usually fully vocalized with alif, vav, and y in Arabic.

16.

I think that overall the draft does a good job of explaining the issues within the scope of the committee's charge. I do have a couple of notes to add.

In section E, systems not supporting non-Roman are still quite common in the African context, even for academic and research libraries, where UNESCO's CDS-ISIS is still in wide use. This is starting to change, but slowly. To the best of my knowledge, the only ILS's that facilitate e-mailing of non-roman script in records are Aleph and possibly Koha. I don't know of any Voyager sites that are able to send non-roman script in their e-mailed records.

I would highlight for this reason the case of Ethiopic, where LC romanization is not widely known to the user community, yet most systems in use in Ethiopia and Eritrea do not offer good support for the non-roman script either. The worst case scenario would be general implementation for all scripts toward Model B before systems are in place to handle it, although some pressure in that direction may actually be helpful. The Ethiopic scenario falls under the case cited in Section B with regard to Tibetan, minus the advantage of a widely accepted romanization method.

If I understand correctly, the dropping of a requirement for parallel romanized fields as 880's would not necessarily preclude use of alternate or repeatable fields, such as the 246 and 700's, and parallel access could still be added in alternate scripts, such as katakana. If so, I don't see the advantage of dropping the ability to link parallel fields. I would be in favor of considering a modified version of model B—whether based on retaining 880's but making them optional, or counting them as core-level perhaps; or replacing the requirement to romanize with one to provide parallel access in some form, regardless of script.

17.

B. last sentence of first paragraph: suggest ending as "...can be displayed next to the corresponding transliterated field." [Some systems actually give choices as to whether you display the non-Roman data corresponding to the regular MARC field, or to give as a block together, either before or after the regular MARC field block]

B. examples: should probably indicate that you are using a shorthand display like used in OCLC, not the actual MARC data (which would have 880 fields, \$6 data, \$a used), presuming you don't want to reformat as true MARC. Lord knows the real data is much more complicated!

B. second to last paragraph: probably should indicate that this paragraph refers primarily to North American libraries-- some other paragraphs in this section refer to NA or East Asia, and one reader sometimes got confused as to what was being referred to.

B. last paragraph: I think you were pretty careful to refer to non-Roman instead of non-Latin, except in this paragraph.

C. first sentence: again, I think you mean North American catalogers?

C.2 last sentence of first paragraph: two misspellings of pronunciation

C.2 last paragraph-- we should note that the Persian romanization table is under review/revision.

D.1, second paragraph, first sentence: "In principle" wasn't understood by one reader-- suggest lopping it off, as it would still work without it.

D.1, fourth paragraph, last sentence: maybe soften to say "...by a patron *may not be* in the system used by the library", as it would vary by script, availability of ISO standard, etc., so "unlikely" may be a bit harsh.

D.2, third paragraph, first sentence: one reader was confused with end of sentence, could we reword as "...and the definite article are prefixes to the nouns."

D.3., last sentence: instead of "this may not be important enough ..." could we say "sorting is not an important issue for some types of keyword indexing"?

D.4, second paragraph, last sentence: "whose" should be "who"

E., first paragraph: think there's a word missing at end "...in academic and research libraries, *but* they *may* still be more common elsewhere."

E., third paragraph: instead of "provide" could we say "use or provide"; also put a comma after "etc." (Sorry, this is bad wording I think I gave you!)

F. second paragraph, first sentence: can we add to the end of that sentence "...in systems that do not provide keyword searching of references in linked authority records."

F. last paragraph: could we add one more sentence at the end, more or less stating an 'ideal' situation, like: "Ideally, most heading variations could be handled in authority records, leaving redundant use of original script data in parallel bibliographic headings to the few special cases where it may continue to be necessary."

G., first paragraph, second sentence: one reader thought it would be better to name all the

languages where we do make use of automatic transliteration-- that would be: Chinese, Korean, Arabic, Persian, Urdu, Yiddish, Hebrew, Russian and some other Slavic languages. Then the last sentence could be simplified by eliminating "Chinese and Korean transliteration tools are also being used or in development, and" and just having "LC is hoping to"

G., last paragraph: while this section points to a number of problems with automatic transliterations, it shortsells the significant productivity gains that even imperfect transliteration is providing us-- we suggest adding a final sentence at the end: "Even with some of the shortcomings mentioned above, automated transliteration techniques still provide a significant benefit to the cataloger and greatly increase cataloger productivity."

I., last paragraph, sentence that begins "The amount [sic] of romanization in records could be reduced ..." Since there is no real discussion of this topic in the report, maybe this needs to be rephrased as a recommendation to see if an investigation into reducing the amount of romanization could be proposed.

I., last paragraph: is it worth trying to mention the languages/scripts that are more likely to be able to move to Model B sooner, or do we not really know that yet? [...] also suggests that a recommendation to consult with library system vendors and developers of other tools used (e.g., Microsoft) would be useful-- we complain about their products to each other, but are we complaining to them?

18.

Automatic transliteration is not yet being used at LC for Hebrew or Yiddish, though we are working very hard on proposed methods. I wouldn't want to raise the possibility that Hebraica catalogers out there might start holding their breaths at this point.

Because the Hebrew ALA/LC romanization system does not provide a one-to-one character match with Hebrew script text (far from it!), it seems likely to me that a significant amount of "massaging" by catalogers will always be necessary when automatic transliteration finally becomes available. From what I hear, though, Hebrew will provide the low position on the spectrum--other languages don't seem to have nearly as many problems. I would rewrite the end of [...]s last sentence above as "(greatly) increase cataloger productivity, especially for certain languages and scripts."