

ALCTS Task Force on Non-
English Access

Comments Received and Their Disposition

This compilation is based on comments received on the Task Force's Report via the form posted on the ALCTS website https://cs.ala.org/alcts/non-english_comment_form/ starting from October 12, 2006 and ending on the closing date for comments, December 1, 2006.

Comments from institutions and individuals received directly via electronic mail are also included. The compiled comments are arranged in order of the sections in the report. General comments or comments not categorized by the poster are grouped at the end.

Feedback via ALCTS site as of December 1, 2006:

Total number of comment submissions received to date: 35

Total number of individuals submitting comments: 7

When a submitted comment addressed more than one section of the report, it was split into multiple comments, necessary because this document is arranged in the order of the sections of the Task Force's Report.

Contents

I. Executive Summary	2
II. Recommendations	3
IV. Overview of the Current Situation	11
Appendix A: Charge	19
Appendix B: Machine-Readable Bibliographic Information (MARBI) Committee	20
Appendix C: Committee on Cataloging: Description and Access (CC:DA)	21
Appendix D: ALCTS Committee on Cataloging: Asian and African Material (CC:AAM)	21
Appendix E: Other ALCTS Groups	22
Appendix F: Library of Congress	22
Appendix G: Program for Cooperative Cataloging (PCC)	23
Appendix H: OCLC	23
Appendix I: RLG	23
Appendix J: ILS Vendors/Local Systems	23
Appendix K: Authority Control Vendors	24
Appendix L: Reports from Other Groups	24
Bibliography of Unicode™ in the Library Environment	24
General Comments	25
Kudos	27
ANNEX: Task Force Responses to Questions About Romanization	27

I. Executive Summary

3rd paragraph

Supporting access to non-English materials in the context of an English-language catalog requires significant additional resources. To that end bibliographic utilities and integrated library systems are making substantial progress in adding scripts that have not previously been supported in the bibliographic environment. This task is complicated by the need to agree on the technical standards that are vital if record retrieval and exchange is to occur without data corruption. Name Authority Cooperative Program (NACO) institutions are adding non-Roman data to authority records. The bibliographic utilities are acquiring non-English language records from book trade companies, and are also providing cataloging services to libraries overseas.

Comment # 1

Name: James E. Agenbroad

Institutional Affiliation: Retired

In the third paragraph (p. 3) of the executive summary it says, "Name Authority Cooperative Program (NACO) institutions are adding non-roman data to authority records."

The next paragraph says, "The current NACO file provides name/title access in languages written in Roman script. Addition of non-Roman script capability will provide for access in other languages."

Is nonroman access already possible and allowed or not?

Comment # 2

Name: Library of Congress

In the Executive Summary, 3rd paragraph, the statement is made "Name Authority Cooperative Program (NACO) institutions are adding non-Roman data to authority records," but in the 4th paragraph, "The current NACO file provides name/title access in languages written in Roman script." In fact, the NACO file also includes romanized access for non-Roman script entities. We also hope over the next year or two we will enable adding cross-references in non-roman scripts to provide even greater access.

Disposition of comments #1 and #2 [Editorial change]

Error in text.

Delete the sentence

Name Authority Cooperative Program (NACO) institutions are adding non-Roman data to authority records,

and substitute:

The principal institutions involved in NACO ("NACO nodes") are working on procedures to allow the addition of non-Roman data to authority records."

4th paragraph

There are several additional requirements to make library resources available to a person whose first language is not English. The most easily achieved requirement is the addition of user instructions and online help in additional languages to library systems. Many integrated library systems already provide for this. The ability to search in a language other than English requires either that records include multilingual access points, or alternatively that searching is redirected through the use of multilingual thesauri and authority files. Noting that catalogs are intended to meet the needs of their respective user communities, provision of non-English access points and other information will be a library-specific decision. Record acquisition by the utilities from foreign sources will certainly facilitate such cataloging. The current NACO file provides name/title access in languages written in Roman script. Addition of non-Roman script capability will provide for access in other languages. Thought must also be given to providing multilingual and multiscript subject access. The need to present search results in language-specific sort order must also be carefully evaluated.

Comment #3

Name: J. McRee Elrod

Institutional Affiliation: Special Libraries Cataloguing

[Continuation of posting]

The statement that decisions are "library-specific" moves away from IFLA's idea of UBC and international standardization. Standardization of records, and of search procedures, insofar as local resources allow, should be sought.

Disposition of comment #3

Apparent misunderstanding on reader's part. While multilingual/multiscript data may be available (in accordance with the principles of UBC), the decision on what records are appropriate for a library's users is up to each library. "Search procedures" will be "standardized" by the use of software and input equipment from major IT companies.

Comment #4

Name: Library of Congress

In the Executive Summary...in the 4th paragraph, "The current NACO file provides name/title access in languages written in Roman script." In fact, the NACO file also includes romanized access for non-Roman script entities. We also hope over the next year or two we will enable adding cross-references in non-roman scripts to provide even greater access.

Disposition of comment #4 [Editorial change]

Error in text.

Delete the phrase

languages written in

and add "only" at the end, so that the sentence reads:

The current NACO file provides name/title access in Roman script only.

Comment #5

Name: James E. Agenbroad

Institutional Affiliation: Retired

[Continuation of posting]

It would also be useful to clarify what is meant by "multilingual and multiscript subject access". It could mean either: a) subject access for readers of foreign languages/scripts to library materials in foreign languages they could read via terms in their respective languages/scripts, or b) Access for readers of foreign languages/scripts to ALL library materials via terms in their languages/scripts. In the first case Spanish, Chinese, etc. readers would get access via Spanish, Chinese or other terms to resources they could read only. In the second case they would get subject access to all library resources on topics of interest to them--even those they could not read.

Disposition of comment #5

No change to the Executive Summary needed. The phrase covers both cases.

II. Recommendations

Recommendation 1

Convert the Task Force's discussion list <NONENGLISH@ala.org> to an open list available to the library community to facilitate continued discussion of non-English language access issues.

Comment #6

Name: James Agenbroad

Institutional Affiliation: Retired

The first recommendation is to open the Task Force list to the public. Instead a new list for the public was formed. This deprives the public of access to earlier discussions. The new list has just become active; I wonder if it is rehashing some of the TF's earlier discussions.

Disposition of comment #6

Setting up a separate list was an ALCTS decision, and is consistent with what the Task Force intended. Everything significant that the Task Force discussed on their list is presented in the Report.

Recommendation 2

Establish a working group whose task is to define requirements for the support of each script and language in library computer applications. These requirements will facilitate library system development and evaluation, and will provide guidance to system implementers. It is recommended that the working group's charge include the following points:

- Determine the scripts to be documented. Include Latin as one of the scripts. (This is necessary because Unicode includes additional Latin script characters outside of ASCII and ANSEL.)
- Create a checklist that supports all scripts and languages, and that can be the basis for requirements for specific scripts or languages.
- Obtain the assistance of appropriate language experts and organizations (including those outside of the United States) when defining the requirements for a specific script.

Comment #7

Name: Library of Congress

The second recommendation to establish a working group to define requirements for support of each script and language in library computer applications should result in a very useful checklist. May we offer that the Library of Congress be added to the "Sources of expertise" given our long-term interest, commitment, and involvement in these issues?

Disposition of comment #7

[Editorial change to Recommendation 2]

Add "Library of Congress," **immediately after the subheading** "Sources of Expertise:".

Comment #8

Name: James Agenbroad

Institutional Affiliation: Retired

The working group is an excellent idea. To whom would it report? Google and other commercial firms are also doing more with multilingual access and might be consulted. I suspect they can afford to hire language expertise. What arrangements would let them share what they learn I do not know.

Disposition of comment #8

The reporting structure is unknown at this time.

Comment #9

Name: Keiko Suzuki

Institutional Affiliation: East Asia Library, Yale University Library

#2&5: Does this working group include members from vendors (, which are at least included in Interested parties)? If it does, that would be excellent. I believe the working group needs at least one liaison from vendor group to facilitate the discussion as well as put pressure on improvements.

Disposition of comment #9

The composition of working groups depends on the appointing agency (such as the ALCTS Board) and the availability of volunteers. An effort is usually made to include vendor representation on projects likely to affect implementations.

The appropriate place for libraries to request specific changes to vendor systems is through their vendor's customer services or through related user groups. ALA does not address specific vendor offerings except through the promotion of community standards.

Comment #10

Name: Martin Heijdra

Institutional Affiliation: Princeton University

The Council on East Asian Libraries, (CEAL) has formed a Special Committee on CJK Capabilities in Local Systems which may be related to Recommendation 2.

Disposition of comment #10

Noted. This recommendation includes working with appropriate organizations when defining requirements for specific scripts.

Recommendation 3

Charge CC:AAM and/or the Committee for Cataloging: Description and Access (CC:DA) to work with the Program for Cooperative Cataloging (PCC) to review and update the core level supplement on "Guidelines for Multiple Character Sets". (see: <http://www.loc.gov/catdir/pcc/bibco/coreintro.html#9>).

Comment #11

Name: Keiko Suzuki

Institutional Affiliation: East Asia Library, Yale University Library

#3: I think this is great idea. Even though this is more regarded to CC:AAM responsibility, CC:AAM might have hard time to find volunteer members for the work. Thus, it's good idea to expand the charge to CC:DA or other CC:DA liaison groups (ACRL SEE, AJL, PLA etc.)

Disposition of comment #11

[Editorial change to Recommendation 3]

Add the following text:

Parties charged with carrying out this recommendation are permitted to request assistance from organizations with language expertise, both within and outside ALA.

Comment #12

Name: Library of Congress

Similarly for recommendation 3 to develop the "Guidelines for Multiple Character Sets," may we offer LC expertise, as it is the mission of the Cataloging Policy and Support Office to provide such documentation for our own staff and to assist with the PCC documentation (indeed also writing some of it)?

Disposition of comment #12

[Editorial change to Recommendation 3]

In the subheading "Sources of expertise", change "PCC" to "Library of Congress and PCC representatives,".

Recommendation 4

As *Resource Description and Access* (RDA) is developed, it is recommended that CC:DA and CC:AAM consider and comment on any impact that the new rules will have on cataloging non-English materials. This review should be referred to appropriate liaisons and groups when appropriate language expertise is lacking.

Comment #13

Name: J. McRee Elrod

Institutional Affiliation: Special Libraries Cataloguing

[Continuation of posting]

Influencing RDA should be made High, not High/Medium. Although it moves beyond access points, substituting words in the language of the catalogue for the present Latin abbreviations of AACR2 is a move away from equal access to multilingual materials.

[Other portions of this posting appear as comments in *Recommendations for Additional Recommendations* and in *General Comments*.]

Disposition of comment #13

The priority reflects the opinion of Task Force members (in this case, a divided opinion). Opinions on specific aspects of RDA should be communicated through the existing channels. The Task Force has no direct involvement with RDA development.

Comment #14

Name: Keiko Suzuki

Institutional Affiliation: East Asia Library, Yale University Library

#4: As a liaison of CC:AAM to CC:DA, I have been involved this process personally, and trying to get as many volunteers as possible from ACRL SEES, AJL, and other area studies organizations to review RDA. And JSC is also up to the task of internationalize RDA with LC's numerous proposals called "5JSC/LC/5 Internationalization". I hope we have opportunities to be able to discuss these efforts with other national library associations in near future.

Disposition of comment #14

No action required. Recommendation #4 appears to be already in effect.

Recommendation 5

Analyze the need for rules for the sorting of bibliographic entries in the Unicode environment. This analysis should take into account current cataloging rules, in particular the rules for heading creation, and the capabilities of online library catalogs. If it is determined that there is a need for such rules, the Unicode Collation Algorithm (<http://www.unicode.org/reports/tr10/>) should be considered during development.

Comment #15

Name: Library of Congress

For recommendation 5, we suggest that the "NACO nodes," (Library of Congress, The National Library of Medicine, the British Library, and OCLC (and formerly RLG) in consultation with the Library and Archives Canada), who coordinate implementation of changed practices involving the authority record format and character sets, should be included in this analysis of rules for sorting bibliographic entries in the Unicode environment. This group is also working on plans for implementing full Unicode characters (not just the MARC-8 repertoire) as well as more specific plans to include non-roman references in authority records in the NACO environment.

Disposition of comment #15

[Editorial change to Recommendation 5]

In the subheading "Sources of expertise", insert "libraries that set national policies," between "... Unicode Collation Algorithm," and "LITA".

Note: The Task Force considers "OCLC (and formerly RLG)" to belong in the "Technical experts" category.

Comment #16

Name: James Agenbroad

Institutional Affiliation: Retired

Library rules for sorting now interfile headings, titles, etc. in various roman script languages. The alternative of separate sorting for Polish, Czech, Norwegian, Swedish, etc. seems undesirable, especially since MARC does not identify the language of a field or text string, but only just the language of the text. I believe this approach would be useful for other scripts used by several languages such as Cyrillic and Arabic. CJK sorting is best treated as a separate issue.

Disposition of comment #16

Noted.

Comment #17

Name: Keiko Suzuki

Institutional Affiliation: East Asia Library, Yale University Library

#2&5: Does this working group include members from vendors (, which are at least included in Interested parties)? If it does, that would be excellent. I believe the working group needs at least one liaison from vendor group to facilitate the discussion as well as put pressure on improvements.

Disposition of comment #17

The composition of working groups depends on the appointing agency (such as the ALCTS Board) and the availability of volunteers. An effort is usually made to include vendor representation on projects likely to affect implementations.

The appropriate place for libraries to request specific changes to vendor systems is through their vendor's customer services or through related user groups. ALA does not address specific vendor offerings except through the promotion of community standards.

Recommendation 6

Assign the ALCTS Cataloging and Classification Section (CCS) to work with the Public Library Association (PLA) Cataloging Needs for Public Libraries Committee to plan joint programs, preconferences, and continuing education on the cataloging needs for libraries with multi-lingual user populations.

No comments received.

Recommendation 7

Continue to sponsor programs, preconferences, and continuing education on multilingual access, including Unicode implementation, standards, best practices, and user interface issues, collaborating with LITA and other groups, where appropriate.

No comments received.

Recommendation 8

In keeping with ALA's commitment to a diverse library workplace, work with ACRL, PLA, and other organizations to recruit library workers and specifically catalogers who are expert in one or more non-English languages.

Comment #18

Name: Satia Orange

Institutional Affiliation: Office for Literacy and Outreach Services (OLOS), American Library Association
#8: Recruitment falls in HRDR (ALA Office for Human Resource Development & Recruitment). HRDR recruits and runs placement centers and services, along with American Libraries and ACRL. Spectrum has a network of scholars as possible recruiters and applicants. Spectrum staff also communicate with other agencies and organizations networking with new librarians.

Disposition of comment #18

Not all organizations could be listed in the report, but this is useful information for those who may have to follow up on this recommendation if the ALCTS Board approves it.

Recommendation 9

Assign the CCS Cataloging of Children's Materials Committee to work with American Association of School Librarians (AASL) and PLA to identify specific needs for non-English speaking children in regards to library information.

Comment #19

Name: Joanna F. Fountain

Institutional Affiliation: Department of Library Science, Sam Houston State University (Tex.)

I am concerned that the medium/low priority assigned to Recommendation 9 will not result in anything of substance for several years. I believe we need information sooner than that. If possible, it would be good to move this up so that the 3 organizations - each of which has its own procedures - can begin to work on a

plan for a study. Then past the plan to the study - how long? - and agreeing on and issuing a report. Thank you for considering this!

Disposition of comment #19

Although Recommendation 9 has medium/low priority, the starting date is given as Midwinter 2007.

Speeding up the process is up to the interested parties. For example, the Cataloging of Children's Materials Committee can ensure that CCS and ALCTS issue the desired directives as soon as possible to sanction CCMC activity.

Comment #20

Name: Satia Orange

Institutional Affiliation: Office for Literacy and Outreach Services (OLOS), American Library Association
#9: ALSC (Association of Library Services to Children) is also an appropriate resource for cataloguing children's materials.

Disposition of comment #20

[Editorial change]

ALSC should be added to this recommendation.

Additional Recommendations from the Task Force

Two comments in particular (one representing input from a group of librarians) questioned the use of romanization in general, and the use of 880 fields in MARC 21 records. The section "Multiscript Records" in each MARC 21 format describes records with and without 880 fields (designated as Model A and Model B). When 880 fields are used (Model A), non-Roman script data is in the 880 fields, and its romanized equivalent is in the regular fields.

Comment #21

Name: Frances McNamara

Institutional Affiliation: University of Chicago Library

[Recommendation 2] Sounds good but would this group give guidance on changes to current practice. In particular use of 880 vs. use of vernacular in the actual fields like 245.

[Remainder of posting asked about transliteration of Cyrillic – see below]

Comment #22

Name: Ohio State University Libraries (Magda El-Sherbini, compiler)

The question was raised -- why are we continuing to romanize? Users of the materials know the language. Physical control is by barcodes. We don't romanize web sites and users still can search the web in foreign languages. Etc, etc. The practice of romanization developed so that cards could be filed in card catalogs. Users, however, find romanization confusing -- that is one reason why they want to browse the physical collection -- since there are many different systems. In the case of Japanese, for example, students learn a different romanization in the classroom than the one the library uses. The library's system -- which involves use of diacritics -- can only be displayed, not input by the user. Furthermore, the libraries are using inconsistent romanization -- the rules have changed several times, for one thing. For example a search on the keyword "Shimbun" in our local system (Oscar) produces 219 matches, while "Shinbun" produces 551 - - those are the same word in Japanese, romanized in two different ways. This example illustrates how romanization gets in the way of efficient retrieval of information. Again, why are we continuing to romanize? If we continue to romanize, why don't we monitor discrepancies that impede access (eliminate shimbun/shinbun, etc)?

The Forum on Non-English Access (held Jan. 20, 2007) included a slide "What We Have Learned" that noted romanization as an issue.¹ Part of the Forum was devoted to a mini-

¹ The other issue "Not just ILS—other library software (ILL, bibliography)" was based on comments on the discussion list, which pointed out that the overview section did not specifically mention library software products other than ILSes, and the surveys were limited to ILSes and authority control software. This is not

discussion on romanization; questions about romanization that were asked at the Forum are listed as an annex (at the end of this document). Some of the issues posed in Comment #22 were also raised at the Forum.

Comment #23

Name: Frances McNamara

Institutional Affiliation: University of Chicago Library

[Continuation of posting]

I'm looking for guidance about taking advantage of programs to convert transliterated Russian to Cyrillic and what the MARC records will look like at the end of the process.

Disposition of comment #23

A paper that might be helpful to you is cited in the bibliography: Jacobs, Jane W., Ed Summers, and Elizabeth Ankersen. "Cyril: Expanding the Horizons of MARC 21." *Library Hi Tech* 22, no. 1 (2004): 8-17.

OCLC documentation on its Connexion cataloging service <http://www.oclc.org/connexion/about/features/nonlatin/default.htm> says:

"Additional transliteration macros, created by Joel Hahn, Niles Public Library District, are available for use with Connexion client to transliterate Cyrillic, Greek, and Hebrew to and from Latin script."

Romanization is clearly an issue of concern to the library community. The Task Force therefore proposes an additional recommendation to address the community's concerns:

Recommendation 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

Priority: High

Sources of expertise: Library of Congress, other libraries that set national policies, ALCTS, PCC representatives.

Interested parties: Catalogers and public services librarians

Contingencies: Inability of library systems to deal with both types of MARC record

Timeline: Beginning Summer/Fall 2007

Comment #24

Name: James E. Agenbroad

Institutional Affiliation: Retired

On page 3 of the Executive Summary, the last paragraph, "Thought must be given of providing multilingual and multiscript subject access." I agree and would add that classification schedules with their captions and index terms are also part of subject access. An additional recommendation is needed so this gets not just thought but action.

an issue, however, since Recommendation 2 applies to "library computer applications." That is, Recommendation 2 applies to all library computer applications, including those that were not mentioned/surveyed (e.g., ILL systems, bibliography formatting systems, federated search software, etc.) as well as those that were.

Disposition of comment #24

No change to the Executive Summary needed.

Note that the Library of Congress is adding text in non-Roman scripts to selected classification schedules and other tools, and this fact is now mentioned in IV. I *Conclusion*.

Recommendation 11

Assign the CCS Subject Analysis Committee (SAC), working with appropriate library organizations, to study the needs of library users for multilingual subject access in the appropriate script(s), and to propose steps to address those needs.

Priority: Medium

Sources of expertise: CCS Subject Analysis Committee (SAC); ALA committees specializing in languages other than English; IFLA Classification and Indexing Section; Library organizations in nations and regions where a language is used

Interested parties: Libraries providing service to multilingual populations

Contingencies: None identified

Timeline: Beginning Summer/Fall 2007

Other Comments Recommending Additional Recommendations

Comment #25

Name: James Agenbroad

Institutional Affiliation: Retired

Recommendation Number: New

Add a recommendation: "Urge MARBI, via its two ALCTS representatives, to expand the character repertoire for Unicode (UTF-8) encoded records to include all Unicode characters. In 2005 MARBI responded (p.37-38) to a motion to this effect from CC:AAM that doing this must wait until resolution of technical issues involving OCLC and RLG. Since OCLC now uses Unicode encoded Thai and Tamil characters (p. 45) not yet allowed in MARC and other South Asian scripts are expected in 2006, and RLG now uses Unicode control characters not allowed in MARC (p. 48) this expansion is now appropriate.

Priority: High

Source of expertise: MARBI and members of the Unicode-MARC list

Interested parties: Library users, MARC software vendors

Contingencies: Is a MARBI resolution sufficient or is a proposal needed?

Timeline: 2007 Midwinter for a resolution or a proposal or Annual at the latest."

Comment #26

Name: J. McRee Elrod

Institutional Affiliation: Special Libraries Cataloguing

Comment on Specific Section:

There should be an additional recommendation that MARBI be urged to expand the approved character set.

[Remainder of posting included a comment on Recommendation 4, *which see*]

Disposition of comments #25 and #26

Moot. The Library of Congress is in the process of revising the *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media* (as announced by Sally McCallum at the January 2007 MARBI meeting).

IV. Overview of the Current Situation

A. Cataloging Practices

For many years libraries in the United States have assumed that the language of the user was English. For this reason the language of the catalog record and the language of the catalog interface have generally been English. Although many libraries built collections in foreign language materials, the language of the catalog including subject access points has been English. Cataloging for works in non-Roman scripts was done through transliteration of those scripts into a representation of the text in the Roman alphabet. Even where card catalogs were able to represent a number of different scripts for a multitude of languages, representation of these same scripts in electronic records that could be shared among different library catalogs has been a more difficult challenge to overcome. Groups addressing cataloging of non-Roman script materials focused on the problems of transliteration and romanization of scripts and the merits of various schemes.

Comment #27

Name: Library of Congress

Although we understand that the language of cataloging and the language of the catalog apply to only some of the elements in the bibliographic record, it might be beneficial to clarify that indeed some transcribed information in bibliographic description will be in the “original” language found on the item being cataloged, even if in some cases that additionally is transliterated or romanized.

Disposition of comment #27

[Editorial change to Report]

Replace current paragraph with the following two paragraphs:

For many years libraries in the United States have assumed that the language of the user was English. For this reason the language of the catalog record and the language of the catalog interface have generally been English. Although many libraries built collections in foreign language materials, the language of the catalog including subject access points has been English. Where possible, the actual language and script has been used to transcribe particular descriptive information. AACR2 Rule 1.0E says, for specific areas, “give information transcribed from the item itself in the language and script (wherever practicable) in which it appears there.”

The earliest automated library systems were limited to Latin script (except for three Greek symbols). Cataloging for works in non-Roman scripts therefore required romanization of any text in non-Roman script(s). With the addition of support for non-Roman scripts to MARC, cataloging records with non-Roman data created by US libraries generally have both transcription of the actual script(s) and the romanized equivalent. The inclusion of romanized equivalents allows all libraries to utilize the records, even when a system lacks non-Roman script capability.

Comment #28

Name: Library of Congress

You may also know that the early catalog cards provided by the Library of Congress gave transcription in original scripts with a filing title in romanized form. It was with the introduction of the MARC Format and system limitations of the time that robbed us of non-roman access. Such background may be useful to include in your report.

Disposition of comment #28

Covered in the LC report (Appendix F).

Comment #29

Name: Library of Congress

It may be useful to use the term, “romanized” in place of “transliterated” to be more inclusive for the US environment of conversion to the Latin script.

Disposition of comment #29

[Editorial change]

Comment applied in the rewording of this section (Disposition of comment #27).

B. MARC 21 and Unicode

4th paragraph

Extension of the MARC 21 character repertoire beyond what is available in “MARC-8” was discussed in two reports presented to the MARBI committee in 2004 and 2005. MARBI is currently addressing an issue in Unicode to MARC-8 conversion: how to indicate the presence of a character that cannot be mapped to a MARC-8 equivalent in the MARC-8 record.

Comment #30

Name: James Agenbroad

Institutional Affiliation: Retired

In section B, the fourth paragraph of the overview should be revised to show that MARBI has approved two procedures for dealing during conversion from Unicode to MARC-8 with characters that have no MAR-8 equivalent. I believe this justifies expansion of the MARC's Unicode character repertoire to all of Unicode 5.0.

Comment #31

Name: Library of Congress

You undoubtedly have already received comments to update this section to reflect the latest MARBI activity.

Disposition of comments #30 and #31

[Editorial change]

The statement was accurate at the time of publication.

Change

MARBI is currently addressing an issue in Unicode to MARC-8 conversion: how to indicate the presence of a character that cannot be mapped to a MARC-8 equivalent in the MARC-8 record.

to

MARBI has approved techniques to indicate the presence of a character that cannot be mapped to a MARC-8 equivalent in the MARC-8 record.

See also comment #54 on Appendix B: Machine-Readable Bibliographic Information (MARBI) Committee.

C. Non-English Support in Library Systems

1st paragraph

In the 1980s and 1990s, the bibliographic utilities developed methods to represent non-Roman scripts in bibliographic records. Local library systems were slower to adopt the technology that would allow search and display in these scripts. The development of the Unicode standard led to the hope that all languages could be represented fully in their proper character set in library systems.

Comment #32

Name: Library of Congress

It would be more accurate to say in the first sentence that the bibliographic utilities with the Library of Congress developed methods to represent non-roman scripts in bibliographic records.

Disposition of comment #32 [Editorial change]

Agreed. Change will be made.

Comment #33

Name: James Agenbroad

Institutional Affiliation: Retired

The first paragraph of section C of the overview, needs to be updated. "The development of the Unicode standard led to THE HOPE THAT all languages could be full represented ..." Given that OCLC has implemented Thai and Tamil scripts and the British Library has implemented Armenian, could "the hope that" be replaced with "the demonstrated reality that"?

Disposition of comment #33

No change needed.

The paragraph begins "In the 1980s and 1990s," so it is clearly referring to the past.

4th paragraph

A questionnaire was sent at the end of 2005 to 20 companies that market ILSes or OPACs to determine support for languages other than English and for non-Roman scripts in each company's system. Seven responses were received, and are reported in Appendix J. The 30.5% response rate represents a reasonable profile of the companies queried. Four companies reported using Unicode to provide script support; another company has a Unicode-based system under development (71% of respondents). Note that Unicode is used not just for non-Roman scripts, but also for *all* scripts that are supported, including Latin script.

Comment #34

Name: James Agenbroad

Institutional Affiliation: Retired

In the fourth paragraph of section C, of the overview, the number of responses seems sparse. Perhaps the ALA Library Technology Project could undertake a fuller exploration of vendor's nonroman capabilities. A minor matter, isn't seven responses from 20 vendors over a third, not 30.5%?

Disposition of comment #34

The purpose of the survey was to obtain an overview of the marketplace. The Task Force is satisfied that this goal was accomplished.

The Library Technology Project no longer exists. It was dissolved in 1972 on the recommendation of the Committee on Program Evaluation and Support (COPES).

http://arc.cs.odu.edu:8080/dp9/getrecord/oai_dc/findingaids.grainger.uiuc.edu/oai:findingaids.grainger.uiuc.edu:ALA/15/1/2

[Editorial change] "30.5%" to "35%".

D. Non-Roman Scripts in the Bibliographic Utilities

Comment #35

Name: James Agenbroad

Institutional Affiliation: Retired

Overview section D does not address the need to download records containing nonroman scripts for use in local OPACs for local searching and display.

Disposition of comment #35

No change needed.

A primary function of the bibliographic utilities is to provide records that their customers require (including records that contain non-Latin scripts).

Comment #36

Name: Martin Heijdra

Institutional Affiliation: Princeton University

Section Number: IV.C [as this is a comment about OCLC, it applies to Section D]

As some participants have remarked, OCLC has gotten ahead of the game in allowing now Thai, Tamil, Greek, Devanagari and Bengali script records into WorldCat. However, in as far as I can see, the indexing is somewhat odd and seemingly against Unicode principles: if I understand the documentation correctly, glyphs such as say Devanagari or Tamil ki are indexed as *glyphs*, and not as character combinations, which seems to go against the basic principles of Unicode (there may have been reasons for it, of course,

but it may be a not-well thought-out choice). I would hope that discussions of matters such as these would find a place among topics to be discussed among any follow-up groups to be formed after this Task Force.

Disposition of comment #36

Moot. OCLC reports documentation is the problem.

Comment #37

Name: Martin Heijdra

Institutional Affiliation: Princeton University

Section Number: IV.D

There was one important issue that I think was not really mentioned in the report: between the limitation to MARC8 scripts and expansion to include new scripts there is for CJK users the issue of expansion of characters within Chinese, Japanese and Korean: the use of the scripts is old, but many characters are currently not allowed: ever since RLG implementation of CJK we have all been waiting to be able to include other non-MARC8 CJK characters, and we still seem far away from that goal. Of course, including them would also require much better cross-indexing than currently available: expansion to fully "new" characters is less problematic than increased allowance of variant characters which should cross-index with already available ones.

Disposition of comment #37

Misunderstanding on reader's part. Expansion of character repertoire is not limited to addition of new scripts.

Note that the issue of expansion of the MARC 21 character repertoire applies not only to the utilities but to any MARC 21 application]

E. Non-Roman Scripts in Authority Files

Comment #38

Name: Library of Congress

We are glad to see mention of our early use of the HKCAN authority file, this section should also note the LC-led efforts towards implementing Unicode in bibliographic and authority records that are shared among the NACO nodes (noted above).

Disposition of comment #38

This section deals with authority records, so discussion of use of Unicode in bibliographic records as well as authority records does not belong here.

The use of Unicode in MARC 21 records is addressed in Section B, *MARC 21 and Unicode*.

Library of Congress report (Appendix F) says:

“Since 2002 LC has been exploring ways to restore non-Latin script access in both bibliographic and authority records for all scripts by creating authority records and bibliographic records in the original scripts, transcribing information in the script of the manifestation. One goal is to enable the use of cataloging copy from any source worldwide, capturing the records from the Web or using Z39.50 or other protocols to incorporate those records into the Voyager system if they are useful copy.”

and the sentence

“In the standards area, LC continues to work with the community on the introduction of the use of full Unicode in MARC records.”

is followed by a summary of progress where LC's role was critical.

F. Staffing and Economic Issues

2nd paragraph

The Library of Congress has been distributing all the Arabic, CJK, and Hebrew script books it catalogs in the RLG Union Catalog through the MARC Distribution services. Additionally, CONSER serial records

that include scripts have also been distributed through LC's MARC Distribution services. RLG and OCLC have been exchanging CJK script book records since 1987.

Comment #39

Name: Library of Congress

Second paragraph, please clarify that "The Library of Congress has been distributing MARC records for all the Arabic, CJK, and Hebrew script books it catalogs in RLIN through LC's MARC Distribution Service." It also would be more accurate to explain that LC distributes MARC records for the Arabic, CJK, and Hebrew script serials created in OCLC as well as the CONSER serial records. LC is not distributing the books, which is how the sentence now reads.

Comment #40

Name: James Agenbroad

Institutional Affiliation: Retired

In the second paragraph of section F, "The Library of Congress has been distributing all the Arabic, CJK and Hebrew script books ..." I think this means distributing MARC records for these books but it says LC distributes the books themselves.

Disposition of comments #39 and #40

[Editorial change]

Change

The Library of Congress has been distributing all the Arabic, CJK, and Hebrew script books it catalogs in the RLG Union Catalog through the MARC Distribution services.

to

The Library of Congress has been distributing MARC records (through its MARC Distribution Service) for all the Arabic, CJK, and Hebrew script books it catalogs in RLIN.

Change

Additionally, CONSER serial records that include scripts have also been distributed through LC's MARC Distribution services.

to

Additionally, the CONSER serial records that are distributed through LC's MARC Distribution Service may include non-Roman scripts.

G. Non-Roman Collections

Collections of books and materials in non-Roman scripts at large research libraries such as the Library of Congress, the New York Public Library, and major academic libraries grew significantly in the mid-20th century through the Farmington Plan and the PL-480 Program. Research, and the Title-VI campus programs, permitting students to be trained in research using primary sources in their original languages.

Most public and school libraries did not have to face the issue of building collections of material in foreign languages (some in non-Roman scripts) until the 1990s, when multiculturalism and the need to address the needs of multilingual and immigrant communities began to emerge. Once strategies of acquisitions were set in place, and materials in other languages began to flow freely into libraries, the matter turned to how to get these materials into the hands of users.

Comment #41

Name: James Agenbroad

Institutional Affiliation: Retired

In the first paragraph of section G of the overview needs revision. Many public libraries have provided immigrants with nonroman script library resources since early in the 20th century.

Disposition of comment #41

The comment refers to the second paragraph.

[Editorial change to report]

Replace current sentence with the following:

Except for places with a significant multilingual population, public and school libraries did not have to face the issue of building collections of material in foreign languages (some in non-Roman scripts) until the last few decades, when multiculturalism and the need to provide for people whose primary language is not English were recognized.

H. Unicode and Libraries

Comment #42

Name: James Agenbroad

Institutional Affiliation: Retired

Section H should probably be updated to show MARBI approval in 2006 of proposals dealing with conversion of MARC records from Unicode to MARC-8 encoding and implementation by OCLC of Thai and Tamil with more expected soon and Armenian by the British Library.

Disposition of comment #42

No change needed. The information is in the relevant sections:

- MARBI activity is covered in Section B: *MARC 21 and Unicode*.
- OCLC's script additions are covered in Section D: *Non-Roman scripts in the bibliographic utilities*.
- The British Library's use of Armenian script (mentioned on p. 49) does not belong in this section.

I. Conclusion

1st paragraph

In the context of English as the language of the catalog, provision for all languages and scripts includes:

1. inclusion of foreign language text in its correct script in bibliographic records;
2. authority control mechanisms to provide access by names and titles written in non-Latin scripts to the AACR2 form;
3. authority control mechanisms to provide access by foreign language subject terms in any script to the English language (e.g., LCSH) equivalent;
4. source records for cataloging of foreign language material, especially material in non-Latin scripts.

First Listed Item

1. inclusion of foreign language text in its correct script in bibliographic records;
[Editorial change to Report]

Replace the current wording for this item with:

For foreign language text, use of the script(s) in which the language is conventionally written.

Third Listed Item

3. authority control mechanisms to provide access by foreign language subject terms in any script to the English language (e.g., LCSH) equivalent;

This item appears to place limits on the exploration sanctioned in the new Recommendation 11:

Assign the CCS Subject Analysis Committee (SAC), working with appropriate library organizations, to study the needs of library users for multilingual subject access in the appropriate script(s), and to propose steps to address those needs.

Furthermore, there may not be an exact English language equivalent for certain concepts in other languages.

[Editorial change to Report]

Delete the final clause:

to the English language (e.g., LCSH) equivalent

Comment #43

Name: Library of Congress

In the first paragraph of provisions, the third item indicates “authority control mechanisms to provide access by foreign language subject terms in any script to the English language (e.g., LCSH) equivalent.” The Library of Congress has been exploring the addition of non-roman script references in *LCSH*, and indeed already is including such access and captions in the *Library of Congress Classification* schedules and in the *Classification Web* product (so far Arabic in KBP, Greek in P-PA, Chinese in Buddhism BQ, and soon Hebrew in Jewish Law KBM). We also have been exploring links to other translations of *LCSH* (French and Spanish as a start).

Disposition of comment #43

[Editorial change]

On page 16, **delete the sentence**

Thought must be given to provision of multilingual and multiscript access for subjects.

and insert a new paragraph:

The Library of Congress has been exploring addition of non-Roman script references in *LCSH*, and links to translations of *LCSH*. Text in non-Roman scripts appears in certain *Library of Congress Classification* schedules and in the *Classification Web* product.

Comment #44

Name: James Agenbroad

Institutional Affiliation: Retired

In the third item of section I, foreign language subject access is most ambitious. Since many foreign libraries rely on classified catalogs for subject access it should include use of the captions and index terms as a means of access to classification numbers.

Disposition of comment #44

No change needed.

Addressed in the fourth paragraph of this section:

“Access by foreign language terms for subjects will require the relevant experts within ALA and other library groups to agree on the subject headings list or thesaurus to be used for each language, with development of a list or thesaurus for a language if necessary. The experience with multilingual subject access of libraries in other countries should be studied.”

Comment #45

Name: James Agenbroad

Institutional Affiliation: Retired

The fourth numbered item, "source records for cataloging of foreign language material, especially material in non-Latin scripts" might be clearer if revised to say "records for cataloging of foreign language material (especially non-Latin material) from their country of origin."

Disposition of comment #45

No change needed.

The proposed rewording is too restrictive. Source records for the cataloging of foreign language material may not necessarily be obtained from the country of origin of the material. For example, RLG added CJK records from a French library as RLIN resource records.

3rd paragraph

The principal institutions involved in NACO are at work on the addition of non-Roman data to authority records.

Comment #46

Name: Library of Congress

The third paragraph mentions the “principal institutions involved in NACO” – which more accurately are the NACO nodes mentioned above.

Disposition of comment #46

[Editorial change]

Add a second sentence:

These institutions (the “NACO nodes”) are the Library of Congress, The National Library of Medicine, the British Library, and OCLC (and formerly RLG) in consultation with Library and Archives Canada.

4th last paragraph (on page 16)

The ability to search in a language other than English will require either that there are records with access points in the language, or that the search may be redirected to the AACR2 or LCSH equivalent through use of an authority file.

2nd last paragraph (on page 16)

The current NACO file provides for name/title access in languages written in Latin script.

Addition of non-Roman script capability will provide for access in other languages.

Thought must be given to provision of multilingual and multiscript access for subjects.

Comment #47

Name: Library of Congress

On page 16 following the “ideal requirements” mention is made of using authority files. It also should be mentioned that the Virtual International Authority File proof of concept project (with the Library of Congress and the German National Library and technical expertise from OCLC) is also exploring how to enable the provision of the user-specified language/script for the names of persons (eventually to expand to geographic names and potentially to corporate names). The longer-term vision for this VIAF system is to enable users to see the script and language they want, with the authority records serving as a building block for the future Semantic Web to enable the switching of displays for such capability.

Disposition of comment #47

[Editorial change]

Change

through use of an authority file.

to

through use of an authority file, or, in the future, multiple files linked as the Virtual International Authority File (VIAF).

Additional Issues Not Covered in Overview

The next three comments mention problems with specific library products. Please bear in mind that the appropriate place for libraries to request specific changes to vendor systems is through their vendor's customer services or through related user groups. (ALA does not address specific vendor offerings except through the promotion of community standards.)

Comment #48

Name: Martin Heijdra

Institutional Affiliation: Princeton University

While the report went into some detail on cataloging practices, support in library systems, the bibliographic utilities and authority files etc., I think the net could have been cast a little bit more widely. ILL systems are left out, for example, while these are hopelessly behind in their support of other scripts: support in the leading OCLC-supported system ILLiad is simply abysmal, it still does not manage to talk properly with OCLC even with MARC-8 diacritics, and of course it can't handle 880 fields nor normal fields in other

scripts, etc. etc. Part of the problem seems to be that the z39.50 protocol is insufficient for character encoding specification. I would propose including ILL systems in any further market researches.

Disposition of comment #48

Recommendation 2 states: “Establish a working group whose task is to define requirements for the support of each script and language in library computer applications.”

The Task Force focused on ILS systems because they are fundamental to library service, and many offer a choice of languages for operation. Some ILS systems allow users to initiate ILL requests. ILS systems are also likely to have the broadest support for non-Latin scripts.

Authority systems were examined because the addition of non-Latin scripts to authority files is critical for the provision of adequate library service in many widely-used languages.

Librarians interested in ILL may wish to undertake their own survey of the language and script capabilities of ILL systems, and share the results via the Non English Access discussion list (as well as ILL-specific lists).

The Z29.50 protocol provides for (a) communication between systems about their character set capabilities, and (b) identification of use of UTF-8. Problems encountered with specific systems may be due to the absence of these more recent Z39.50 features or failure to implement them correctly.

Comment #49

Name: Martin Heijdra

Institutional Affiliation: Princeton University

Another problem with bibliographic records is interchange with bibliography software; even when both local systems and the usual bibliographic software are Unicode-savvy, interchange between them of bibliographic records is problematic because of the way bibliographic records are structured. Records with interleaved parallel fields have difficulty because there is no way to differentiate between a 245 romanized, a 245 parallel original script, or a 245 original script-only field or subfield (generally only a and b are needed), since they are all called 245, yet all may have unwanted subfields c. Records using 880 fields cannot easily be interchanged at all to bibliography software, since the 880s just become one blob. I am not sure where the problem lies, but the fact is that currently even with perfect Unicode records and Unicode-capable software such as Endnote and RefWorks, it is basically impossible to import those records from our OPACs in the way users would want to: Romanized Author (,) AUTHOR IN ORIGINAL SCRIPT, Romanized Title (,) TITLE IN ORIGINAL SCRIPT, etc.

Disposition of comment #49

The inability of bibliography systems to interpret the structure of MARC 21 records is unrelated to the Task Force's charge.

Comment #50

Name: Martin Heijdra

Institutional Affiliation: Princeton University

Yet a third item could be federated search software; because of the limitations of z39.50, searches across Unicode and non-Unicode databases, OPACs etc. are difficult to combine.

Disposition of comment #50

The inability of federated search software to deal with diverse databases is unrelated to the Task Force's charge. The issue might be brought to LITA's attention.

The Z29.50 protocol provides for (a) communication between systems about their character set capabilities, and (b) identification of use of UTF-8. Problems encountered with specific systems may be due to the absence of these more recent Z39.50 features or failure to implement them correctly.

Appendix A: Charge

Comment #51

Name: Library of Congress

Topic: Scope of the group's charge

There is some ambiguity in the charge as to whether the "all language and scripts" applies to the library resources or to the enabling structure – indeed both surely apply. Users wanting the non-English access will hope to find the library systems in their language/script (if they are the target users of that library) and hope also to find materials in their language/script in the collections. We also hope there would be links in future systems to enable such users access to materials in other libraries and other information provider locations (Amazon, publishers, etc.). However, we want to caution that users may indeed be looking for or delighted to find non-English materials on a topic of interest, even if they cannot read that language or script.

As one of our staff noted, "So, for subject access, getting a vendor record that provides subject terms in Spanish or Italian as well as those we add in English (or adding them ourselves) is an enhancement for users operating in Spanish or Italian, but not so much for those operating in Thai or Dutch. Piecemeal solutions may add value that addresses needs in specific circumstances and may be worthwhile on that basis; but they may not be as satisfactory if we have broader expectations for resource sharing."

Disposition of comment #51

Noted.

Appendix B: Machine-Readable Bibliographic Information (MARBI) Committee

Comment #52

Name: [posted anonymously]

Institutional Affiliation:

In appendix B, MARBI (page 20) quoting from the MARBI UTech Task Force report, "One of the principal motivations for adopting a UCS [i.e. Unicode] encoding is to facilitate expansion of the USMARC character repertoire." and "restriction of characters to those listed in the USMARC to UCS mapping are viewed to be operative only until such time as proposals concerning expansion are submitted by interested parties and adopted by MARBI." It unclear to me whether or not character repertoire expansion to allow the rest of Unicode to UCS encoded MARC records still requires MARBI approval. In any case recommendation to MARBI that it authorize such expansion would seem in order. Surely the Task Force is an interested party. 1. The Task Force report (p. 45) documents that OCLC has implemented Thai and Tamil script and expects other scripts this year. 2. The report also documents (p.49) that the British Library has implemented Armenian. Thesescripts are not now allowed in MARC. 3. The Library of Congress cataloging policy position, cited on page 66 of the Task Force report, in the section "Opportunities for the future" it says, "The Library will also be an interested participant in the expansion of the MARC repertoire of Unicode to scripts that are not part of that subset [i.e. the MARC-8 subset] today." I hope the Task Force report will urge/recommend MARBI seize this opportunity.

Disposition of comment #52

See disposition of comments #25 and #26.

Comment #53

Name: James Agenbroad

Institutional Affiliation: Retired

In appendix B, The sentence that begins "In 2004, two reports ..." should begin "In 2004 and 2005 ..." since later in the same paragraph it says part 2 is dated June 2005.

Disposition of comment #53 [Editorial change]

Make the recommended correction to the paragraph on page 20.

Comment #54

Name: James Agenbroad

Institutional Affiliation: Retired

Revise appendix B, page 21 to show that in 2006 MARBI approved two proposals to deal with unmappable characters during conversion of Unicode records to MARC-8 encoded records.

Disposition of comment #54

[Editorial change]

Replace the final paragraph

Proposal 2006-04, "Technique for conversion of Unicode to MARC-8," is based on discussions at that forum, and will be discussed by MARBI in January, 2006, in San Antonio.

with

In January, 2006, MARBI passed proposal 2006-09 which defines a lossless technique for conversion of Unicode to MARC-8 that can be used for communication of records during the time that library systems are in transition between the two character sets.

Appendix C: Committee on Cataloging: Description and Access (CC:DA)

Comment #55

Name: James Agenbroad

Institutional Affiliation: Retired

In the section of appendix C on rules internationalization, item 4, what is the current situation regarding CC:AAM's concerns about RDA and romanization? Did CC:AAM prepare a discussion paper for CC:DA? If so is it in the bibliography?

[Remainder of comment added below under Bibliography]

Disposition of comment #55

Item 3 (not item 4) describes CC:AAM's concerns about the application of cataloging rules to the languages of the regions for which CC:AAM is responsible. (For items 3 and 4, see page 27 of the Report.)

Item 11, at the end of the section "*Internationalization of AACR2 and RDA*" (pages 28-29), describes efforts to "make RDA open to use by any community with a context other than English language, ..." thus addressing the concerns expressed in item 3.

CC:AAM did not prepare a formal discussion paper.

CC:AAM reviewed the RDA draft. CC:AAM comments on RDA Part 1, and the Library of Congress proposal on RDA Part I internationalization were incorporated into the ALA responses (available on the JSC Web site).

RDA Part I: <http://www.collectionscanada.ca/jsc/docs/5rda-part1-alaresp.pdf>

LC Proposal: <http://www.collectionscanada.ca/jsc/docs/5lc5-alaresp.pdf>

Comments from the Council for East Asian Libraries were also incorporated into ALA's comments on RDA.

Note: CC:AAM documents are available on the CC:AAM site <http://www.ala.org/ala/altscontent/catalogingsection/catcommittees/catalogingasiana/catalogingasian.htm>

Appendix D: ALCTS Committee on Cataloging: Asian and African Material (CC:AAM)

Comment #56

Name: James Agenbroad

Institutional Affiliation: Retired

In appendix, page 36, what did the CC:AAM report recommend regarding the word "Oriental"?

Disposition of comment #56

The minutes of CC:AAM's 2003 Midwinter meeting (available at <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/catalogingasiana/2003midwinter.htm>) document the outcome of CC:AAM's discussion of this issue.

In the review comments on RDA Part I from the Council on East Asian Libraries, the issue of "Oriental" was brought up again. The Library of Congress proposal notes that both "oriental" and "vernacular" may be sensitive terms, and suggests alternative wording to eliminate their use. <http://www.collectionscanada.ca/jsc/docs/5lc5rev.pdf>

Comment #57

Name: James Agenbroad

Institutional Affiliation: Retired

On page 38, were the prioritized list, literature study and questionnaire prepared? If so could they be cited in the bibliography?

Disposition of comment #57

CC:AAM has not yet published anything.

Note: CC:AAM documents are available on the CC:AAM site <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/catalogingasiana/catalogingasian.htm>

Appendix E: Other ALCTS Groups

No comments received.

Appendix F: Library of Congress

Non-Latin Script Resources

The Library of Congress (LC) has produced cataloging records for non-Latin script resources for over 50 years. Before automation the bibliographic data on cards was produced by GPO in vernacular scripts. Transliteration was also provided for the access points by the catalogers. LC continued to produce the printed non-Latin script cards for a period even after automation and in parallel with MARC record distribution. In the 1980s LC joined RLG to enable production of MARC records with the vernacular for Chinese, Japanese and Korean. Over the next few years LC stopped producing vernacular cards. At that time, LC produced MARC records with transliteration for all scripts except for CKJ, adding vernacular records for Arabic and Hebrew and Yiddish when those scripts were implemented in RLG. The activity described above was largely for monographs. LC produced MARC records only in the Latin script for other forms of material until OCLC offered non-Latin script input. LC then started making serial records in CJK using OCLC, adding Arabic when it became available.

Comment #58

Name: James Agenbroad

Institutional Affiliation: Retired

In appendix F, the second sentence [in the paragraph shown above on Non-Latin Script Resources] would be more accurate if revised, "Before automation some bibliographic data on cards was produced by GPO in vernacular scripts and some the the LC field offices and then reproduced by GPO."

Disposition of comment #58

No change needed. GPO produced the cards regardless of where the typesetting was done.

MARC Standard

In the standards area, LC continues to work with the community on the introduction of the use of full Unicode in MARC records. Initially a mapping of MARC 21's 16,000 Latin, Chinese, Korean, Japanese, Arabic, Hebrew, Cyrillic and Greek characters was made. Then to enable expansion of the repertoire to all of the Unicode characters, a technique for conversion of full Unicode to MARC-8 was developed and approved by the community. This was to enable continued record exchange between systems that had and had not implemented Unicode. Two techniques are being standardized, a lossy one that uses a substitute

character for a Unicode character outside of the MARC-8 mapping range and a lossless one that substitutes a numeric character reference for the out of range characters. The latter technique allows a return to Unicode and restoration of the Unicode character. These techniques are expected to take the community through what may be a long period of conversion of systems to be able to handle Unicode.

Comment #59

Name: James Agenbroad

Institutional Affiliation: Retired

In appendix F, page 43, [in the paragraph shown above] on the MARC Standard, begin the penultimate sentence: "Except when an extremely long field is truncated the latter technique allows ..."

The sentence in question, about use of numerical character references (NCRs), reads:

"The latter technique allows a return to Unicode and restoration of the Unicode character."

Disposition of comment #59

The sentence is correct as written. It says that when an NCR is present in a MARC-8 encoded record, the Unicode equivalent of the NCR can be identified and restored. There is no loss of data in this operation. Loss of data can occur in the earlier conversion process, from Unicode to MARC-8, whenever excessively long fields or records result from use of NCRs. Such truncation would prevent reconstruction of the original Unicode record in its entirety. Note that the substitute character option is preferred by the user community that still requires records encoded in MARC-8.

Appendix G: Program for Cooperative Cataloging (PCC)

No comments received.

Appendix H: OCLC

Comment #60

Name: James Agenbroad

Institutional Affiliation: Retired

Could the statistics on page 46 of appendix H be updated?

Disposition of comment #60

Updated statistics on the language and script aspects of OCLC WorldCat are available at <http://www.oclc.org/worldcat/statistics/default.asp>.

Appendix I: RLG

Comment #61

Name: James Agenbroad

Institutional Affiliation: Retired

In appendix I, the last paragraph on p.48, have RLG's plans for 2006-2007 to add support for character repertoires beyond the MARC-8 repertoire changed due to its merger with OCLC?

Disposition of comment #61

The footnote on page 3, annotating the second paragraph of the Executive Summary, reads:

Note: At the time this report was written (June 2006), OCLC and RLG were separate organizations. No attempt has been made to revise the report as it will be affected by the up-coming integration of RLG network resources into OCLC.

Appendix J: ILS Vendors/Local Systems

No comments received.

Appendix K: Authority Control Vendors

No comments received.

Appendix L: Reports from Other Groups

Comment #62

Name: James Agenbroad

Institutional Affiliation: Retired

In appendix L, other groups, there is no mention of ACRL's Asian, African and Middle Eastern Section. Should there be?

Disposition of comment #62

[Editorial change]

To clarify the content of Appendix L, **insert the following paragraphs** on page 58 immediately below the heading and before any individual reports:

Appendix L is a collection of reports provided by various groups that have an interest in non-English access. These reports come from ALA Committees, Sections, and Working Groups, and from organizations outside of ALA.

This Appendix should not be regarded as a comprehensive listing of all organizations or groups that have an interest in non-English access.

Comment #63

Name: James Agenbroad

Institutional Affiliation: Retired

On page 62 of the same appendix, on MELA, at 1988, the first, and only, volume of the Near East Union List was published that year but by the Library of Congress. Funds for the project may have come from MELA. Though the data was all romanized, the filing was quite advanced. For example, the Arabic initial article, "al-" was ignored at the beginning of personal names, at the beginning of corporate names and, for government agencies, after the place \ name, e.g., 110 1 \$a Egypt.\$b al-[Ministry of Agriculture]. Similar problems can be anticipated with sorting Arabic script names as part of recommendation 5.

Disposition of comment #63

Noted.

Comment #64

Name: James Agenbroad

Institutional Affiliation: Retired

In appendix L, page 64, MELA, if Arabic script "best practices" have been compiled they should be of interest to the recommendation 2 working group, and would seem a useful addition to the reports bibliography.

Disposition of comment #64

No action needed. The text on page 64 actually says that the MELA Committee on Cataloging **intends** (*emphasis added*) to compile "best practices" for bibliographic records with Arabic script.

Bibliography of Unicode™ in the Library Environment

Comment #65

Name: James Agenbroad

Institutional Affiliation: Retired

In the Bibliography (appendix M?) my article, "Romanization is not enough" (Cataloging & Classification Quarterly 42:2 (2006) p.21-34) would seem a timely addition.

Comment #66

Name: James Agenbroad

Institutional Affiliation: Retired

In the section of appendix C on rules internationalization, item 4, what is the current situation regarding CC:AAM's concerns about RDA and romanization? Did CC:AAM prepare a discussion paper for CC:DA? If so is it in the bibliography? If not my recent paper in *Cataloging & Classification Quarterly* (42:2 pp.21-34) "Romanization is not enough" might be usefully added.

Disposition of comments #65 and #66

This selective bibliography was compiled in May 2006. The only changes that have been made are to correct erroneous or outdated information that could mislead readers.

It is not possible for the Task Force to undertake additional literature searches to augment the bibliography. Mr. Agenbroad has notified the library community about his paper by citing it on the Non-English discussion list.

Comment #67

Name: James Agenbroad

Institutional Affiliation: Retired

In entries on page 68 by Wen, Denyang and Ibrahim, Amed Elhaz the date, pagination or both seem odd.

Disposition of comment #67

RELATED TOPICS—TERMINOLOGY, CHARACTER SETS, ISSUES

Editorial Change To Wen Citation:

Delete

(Nov 2001, 2001).

and substitute the following text:

Metadata - from the Dublin Core 2001 Conference, Article No. 64, 2001-11-06.
<http://jodi.tamu.edu/Articles/v02/i02/Wen/>

The Wen document does not have page numbering.

SPECIFIC LANGUAGES AND SCRIPTS—ARABIC, CJK, INDIC, ETC.

Editorial change to Ibrahim citation:

Delete second occurrence of "2005" and unnecessary comma and space preceding it.

The pagination is correct.

General Comments

Comment #68

Name: James E. Agenbroad

Institutional Affiliation: Retired

General Comments:

On page 162 of *Flavor of the Month* by Joel Best (University of California Press, 2006): "Talking about the march of progress implies that change occurs at a steady pace, and always forward. That's wrong. Change is messy. It involves lots of false starts, uncertainty, and confusion. Sometimes people make unwise choices—they guess wrong about how things will change, and they scramble aboard bandwagons headed nowhere." The Task Force report does not reflect that the path to non-English access has also had some difficulties. Some MARBI Task Forces have been slow and not completed their charge.

[Remainder of posting appears in next comment.]

Disposition of comment #68

Noted.

Issue: Some MARBI Task Forces have been slow and not completed their charge.

If there were problems with MARBI Task Forces, they should have been brought to MARBI's attention at the time. The fact that MARBI Task Forces did not appear to move rapidly is due to both the complexity of each project and the care that had to be taken.

Comment #69

Name: James E. Agenbroad

Institutional Affiliation: Retired

General Comments:

[Continued from preceding comment]

There was no follow-up on a British Library suggestion for revision of a proposal for an AACR rule change to allow original script access. The Task Force reports fails to mention these unfortunate facts. It should do so to deter their repetition.

Disposition of comment #69

This comment refers to a proposal submitted by Mr. Agenbroad to CC:DA in 2000, to add a new rule to AACR2 to sanction the use of access points in non-Roman scripts in bibliographic records. <http://www.libraries.psu.edu/tas/jca/ccda/agenbr2.html>

ALA submitted the proposal to JSC. The ALA representative to JSC reported to CC:DA in 2001:

All constituent responses to this document have been negative, suggesting that nonroman access should be accommodated via authority records rather than added entries." <http://www.libraries.psu.edu/tas/jca/ccda/jsc0104.html>

The British Library's suggestion was to modify AACR2 Rule 26.2A2 (which deals with references for different forms of name), not support for the rule change put forward by ALA in its proposal.

No follow-up on the JSC response has been possible because the NACO authority file does not yet have the capability to include non-Roman script data in records. (For further information, see E. *Non-Roman scripts in authority files* and I *Conclusion* in IV *Overview of the current situation*, and the Library of Congress report in Appendix F.)

Comment #70

Name: James Agenbroad

Institutional Affiliation: Retired

Earlier today there was a comment on the 'nonenglishaccess' list about two topics the report doesn't mention. Since time is short I will mention them though I have no familiarity with either one:

1. Interlibrary Loan systems are not discussed in the report and there is room for improvement in how ILL systems deal with requests for nonroman script items.
2. Loading foreign language and nonroman script records into bibliographic systems is unsatisfactory too.

For more details I refer you to an entry on today's list with the subject: Non-English access beyond bibliographic records and local systems.

Disposition of comment #70

See dispositions of comments #48 through #50.

Comment #71

Name: Ohio State University Libraries (Magda El-Sherbini, compiler)

We would like to see more discussion and ways of involvement for people outside the Task Force group.

Disposition of comment #71

There will be opportunities as groups are formed to address issues identified in the report. See also Recommendation #7.

Kudos

Comment #72

Name: Keiko Suzuki

Institutional Affiliation: East Asia Library, Yale University Library

General Comments:

It's very comprehensive and well thought-out report. Kudos to all the TF members!

Comment #73

Name: Library of Congress

We applaud ALCTS for taking this step to recommend ways to advance non-English access to information (both in terms of making non-English resources available to users through acquisitions and in terms of enabling bibliographic access to such materials through bibliographic and authority records as used in library systems).

Thanks for the praise!

ANNEX: Task Force Responses to Questions About Romanization

Forum on Non-English Access (Jan. 20, 2007)

Responses given during the Forum on Non-English Access (held during the ALA Midwinter Meeting on) with further additions from the Task Force.

Q. Why do we continue to Romanize?

Glenn Patton: Romanization still needed as long as NAF does not allow non-Roman scripts.

Ann Della Porta: The Library of Congress has initiated discussions with OCLC, the British Library, Library and Archives Canada, and the National Library of Medicine to outline the steps necessary to provide non-roman data in authority records issued as part of the LC/NAF (Name Authority File). An early agreement has been reached to use the regular MARC 21 tags for including non-roman data (e.g., 4XX, 7XX) in authority records, rather than paired regular and 880 fields, which is the current model for bibliographic record exchange. A proposed model for when and how to record non-roman forms of established headings, and a timeline for including the data in NACO distributions are currently under discussion. LC and the NACO partners will release information on this timeline as it becomes available. A presentation will be made on Sunday during the LITA/ALCTS Authority Control Interest Group.

[Additional TF response]

There are many reasons for the use of romanization in bibliographic and authority records. One reason has to do with choice of access point, and the need to collocate records. (See also comment about collation below.)

Q. Will the information about adding non-Roman scripts to authority records be included in the final version of the report?

A. We will not update the report to be current as of this writing.

Ann Della Porta (Library of Congress): This is the first venue where this [adding Non-Roman scripts to authority records] has been announced; the actual decision was made at LC just over a month or so ago. Sunday's LITA/ALCTS presentation will appear on the LITA web page.

<http://www.ala.org/ala/lita/litamembership/litaigs/authorityalcts/authoritycontrol.cfm>

COMMENT: I prefer Romanization because I am not confident about the scripts.

[Additional TF response]

If you do not read a particular language well, romanization can be helpful. But the cost/benefit question is: should cataloger time and effort be devoted to romanization to meet the needs of such users? Even when pronunciation is provided via romanization, there is no guarantee that the user will know what the word means. If romanization was not included in the record, the user would have to make use of a

dictionary or an online translation device (just as they would have to do if they were reading anything else written in that language).

COMMENT: Romanization fills in the gap for collocation.

[Additional TF response]

The long-standing convention in catalogs created in the English-speaking world is to provide access points in Latin script, regardless of the language of the publication being cataloged, in order to collocate entries for all works by a particular author regardless of language and script.

AACR2 is silent when it comes to provision of access points written in other scripts. Furthermore, conflicts in orthographic conventions based on language or writing traditions would have to be decided.

COMMENT: Romanization sometimes contains extra information about how to read a particular character (Japanese). Scholarly enhancement by the cataloger.

[Additional TF response]

Readings are available elsewhere, e.g., in OCLC's CJK E-Dictionary (assuming that what is described as "Modified Hepburn input code" in the OCLC documentation means "Modified Hepburn romanization.") See also "not confident about the scripts" comment above.

COMMENT: Ease of Romanization depends on the language; many users do not know the LC/ALA tables.

[Additional TF response]

It doesn't matter whether the romanization is easy or not. Romanization is unnatural for a person who reads a language that is properly written in a non-Roman script.

Libraries should, of course, make the ALA/LC tables available as part of their "how to use the catalog" instructions.

COMMENT: Tibetan ALA tables are not popular with scholars due to diacritics.

[Additional TF response]

No romanization scheme is perfect, and there is often more than one romanization scheme for a particular language. Even the library community has developed international standards (published by ISO) that differ from ALA/LC schemes.

The ALA/LC romanization table for Tibetan uses 8 diacritics, plus apostrophe and soft sign. If experts in Tibetan would prefer a different romanization scheme, they should contact the Library of Congress.

COMMENT: With Devanagari scripts; 100 per cent reverse Romanization is possible.

[Additional TF response]

The ALA/LC *Romanization Tables* have tables for these languages that are written in the Devanagari script: Hindi, Marathi, Prakrit and Sanskrit (the latter two in a single table listed under Sanskrit). There are additional tables for other languages of the Indic subcontinent that are written in other scripts.

This comment is supported by the Wikipedia entries for IAST (International Alphabet of Sanskrit Transliteration) <http://en.wikipedia.org/wiki/IAST> and the National Library at Kolkata romanization "also known as *Library of Congress*" (that is, ALA/LC) http://en.wikipedia.org/wiki/National_Library_at_Kolkata_romanization

Q: Are there plans to survey other library systems used in other countries to see if they have solutions from which we can benefit?

A. No plans.

[Additional TF response]

Romanization is a feature of catalogs developed for use in communities with languages written in Latin script. Where the predominant or some languages are written in non-Latin scripts, records are designed for the speaker of a particular language using the original script. Note that MARC 21 is an international format, and can accommodate either cataloging practice.

COMMENT: Please consider not just OPACs but also ILL systems, bibliographic tools (EndNote, etc.) and use of non-Roman scripts.

[Additional TF response]

This is met by Recommendation #2 "Establish a working group whose task is to define requirements for the support of each script and language in library computer applications." There is no limitation to any specific type of application. See also the dispositions for comments #48 through 50.

Q. Do library systems allow a choice to add Romanized or non-Roman data only?

Glenn Patton (OCLC): OCLC's Connexion has allowed to creation of script-only records for several years. There are Arabic records that only use scripts (no Romanized fields) and Connexion will give you the option of exporting only the Romanized fields or only the non-Roman fields from parallel field records.

COMMENT: However, if you have script-only records, it might be difficult for the user who in theory, has to perform every search twice to ensure that maximum retrieval is made.

[Additional TF response]

If user searches using script, and records have only romanization, nothing is retrieved. If romanization input by user does not match ALA/LC romanization, nothing is retrieved. Does a user want maximum retrieval, or only enough to satisfy need? Also depends on what is used for search, on structure of "script only" record, and on authority file support.

COMMENT: OPACs ignore definite articles and have a list of stopwords in Latin-script languages. I suggest adding a recommendation that would eliminate non-Roman script articles from OPAC searches.

[Additional TF response]

Which languages? Who specified the list of stopwords? Are they defined by each library for its own OPAC? Were they developed by a vendor for a particular product?

RLG considered elimination of particles (articles and article-preposition combinations) in word indexing of Hebrew and Arabic, but user experience did not reveal a need for this feature.

Closing Remarks at the Forum, Bruce Johnson [ALCTS President]: This is not a simple issue that will be fixed tomorrow; all of them have dollar considerations. ALCTS is interested in serving as a catalyst to address these issues, but do not go away thinking that these will be solved in 6 months' time. The comments that you all are contributing are very important for consideration in future discussions.

Other Questions/Issues on Romanization from Comment #22

Q/I: The practice of romanization developed so that cards could be filed in card catalogs. Users, however, find romanization confusing -- that is one reason why they want to browse the physical collection -- since there are many different systems.

Recognition of the problems caused by romanization was the impetus for adding support for East Asian and other scripts to the utilities beginning in the 1980s.

Q/I: In the case of Japanese, for example, students learn a different romanization in the classroom than the one the library uses. The library's system -- which involves use of diacritics -- can only be displayed, not input by the user.

Presumably the library educates the faculty and students about the differences in the romanization schemes so that they can substitute the library's "spelling" conventions from what is used in class.

In library systems, diacritics are normally ignored in searching, so inability to enter a diacritic should not be a problem for the user. When diacritics are ignored, you get the same result whether the search includes a macron (ALA/LC) or a circumflex (Nihon-shiki, Kunrei-shiki) to indicate a long vowel, or is simply left out.

Q/I: Furthermore, the libraries are using inconsistent romanization -- the rules have changed several times, for one thing. For example a search on the keyword "Shimbun" in our local system (Oscar) produces 219 matches, while "Shinbun" produces 551 -- those are the same word in Japanese, romanized in two different ways. This example illustrates how romanization gets in the way of efficient retrieval of information. Again, why are we continuing to romanize? If we continue to romanize, why don't we monitor discrepancies that impede access (eliminate shimbun/shinbun, etc)?

There are numerous examples of alternative spellings in our catalogs: consider British versus American conventions for English!

If out-dated romanization practices cannot be corrected, users should at least be told about the problem and, if a workaround exists, how to resolve it.