Association for Library Collections and Technical Services
Task Force on Non-English Access

Complete Report

September 18, 2006

Chair:
Beth Picknally Camden

Members:
Joan Aliprand
Diana M. Brooking
Karen Coyle
Ann Della Porta
Shi Deng
William J. Kopycki
Heidi G. Lerner
Kristin Lindlan
Sally H. McCallum
David N. Nelson
Glenn Patton
Karen Smith-Yoshimura

Table of Contents

I. Executive summary

The Association for Library Collections and Technical Services (**ALCTS**) Executive Committee established the ALCTS Task Force on Non-English Access in October 2005 with the following charge:

> The Task Force shall examine ALA's past, present, and potential future roles in enabling access to library resources in all languages and scripts and in addressing the needs of users of materials in all languages and scripts through the development of library standards and practices.

The task force gathered information from ALCTS and American Library Association (**ALA**) groups, other library organizations, the Library of Congress (**LC**), the bibliographic utilities (OCLC and RLG[1]), and library system vendors. The task force found that substantial activity has taken place over many years to address the complex issues associated with multiscript and multilingual access. A summary of findings is included below and in more depth in the *Overview* section. Detailed reports are attached as appendices.

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.

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

---

[1] 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.

multiscript subject access. The need to present search results in language-specific sort order must also be carefully evaluated.

The following recommendations address specific actions to move this work forward, as well as ongoing needs for education, communication, and staffing.

II. Recommendations

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.
    Priority: High
    Sources of expertise: Owner of NONENGLISH list
    Interested parties: Library community
    Contingencies: Release of report
    Timeline: October 2006 (at the same time as publication of the report)

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.

    Priority: High
    Sources of expertise: OCLC; Area Studies library associations (including Committee for Cataloging: Asian and African Materials (**CC:AAM**) and Association of College and Research Libraries' (**ACRL**) Slavic and East European Section (**SEES**)); Vendors; ALCTS; Library Information and Technology Association (**LITA**)
    Interested parties: Vendors; Purchasers/users of library systems
    Contingencies: Some output will be used in Recommendation 5
    Timeline: Beginning Winter/Spring 2007

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)
    Priority: High
    Sources of expertise: PCC assisted by CC:AAM and/or CC:DA
    Interested parties: PCC participants using one or more non-Latin scripts; other catalogers using non-Latin scripts
    Contingencies: None identified
    Timeline: Beginning Winter/Spring 2007

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.
Priority: High/Medium
Sources of expertise: CC:DA; CC:AAM; SEES; other liaisons and groups as needed
Interested parties: Joint Steering Committee for Revision of Anglo-American Cataloguing Rules (**JSC**) and national agencies represented on JSC; other national library associations and their cataloging experts; all prospective users of RDA.
Contingencies: RDA deadlines
Timeline: As soon as possible for initial review; ongoing during implementation and beyond[2]

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.
Priority: High/Medium
Sources of expertise: Technical experts familiar with Unicode Collation Algorithm; LITA; ALCTS; CC:AAM
Interested parties: Vendors; Purchasers/users of library systems; Language service librarians; ultimately, Library users
Contingencies: Recommendation 5 is contingent upon output from Recommendation 2 with respect to any requirements for sorting. As a preliminary to considering the need for library-specific rules, information could be gathered on searching and sorting in current integrated library systems and database services used by libraries.
Timeline: Beginning Spring/Summer 2007

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.
Priority: Medium
Sources of expertise: CCS; Cataloging Needs for Public Libraries Committee; Association for Library and Information Science Education (**ALISE**)
Interested parties: Library community, especially catalogers and language service librarians
Contingencies: None identified
Timeline: Begin 2007 to plan for 2008 programs

---

[2] **ALCTS Board note**: CC:AAM already has a sub-group working on precisely this issue.

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.
   Priority: Medium
   Sources of expertise: CC:AAM; SEES; Area Studies Library organizations; LITA
   Interested parties: Library community, especially vendors, language service librarians, and systems librarians
   Contingencies: None identified
   Timeline: Begin 2007 to plan for 2008 programs

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.
   Priority: Medium
   Sources of expertise: ACRL; PLA; Area Studies library associations; Library schools
   Stakeholders: Libraries providing service to multilingual populations; Research libraries
   Contingencies: Recruitment for specific positions within libraries is contingent on the needs and budgets of the individual institutions.
   Timeline: Ongoing

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.
   Priority: Medium/Low
   Sources of expertise: CCS (Cataloging of Children's Materials Committee); AASL; PLA
   Interested parties: Children's librarians; School librarians; language service librarians; Teachers; Parents and their children
   Contingencies: None identified
   Timeline: Begin Midwinter 2007

III. Terminology

The report of the Task Force (**TF**) uses terms that may not be familiar to all members of the library community.

Definition of terms

Language: a system of communication between humans that may be oral or written. Text is a language represented visually according to the writing system for that language.

Multilingual: more than one language. The writing method -- whether the script(s) conventionally used for the language, or romanization, or even a mixture -- is irrelevant.

Script: a collection of the symbols used in the writing system(s) of particular language(s). Most scripts are used for more than one language. Some languages are written using a combination of scripts.

Multiscript: more than one script.

Romanization: use of Latin script to write a language that is conventionally written in some other script or scripts.

JACKPHY: acronym standing for the languages Japanese, Arabic, Chinese, Korean, Persian, Hebrew, and Yiddish. These were the languages for which the Library of Congress continued to produce catalog cards until their scripts were implemented in machine-readable form.

MARC-8 scripts: The scripts covered by the individual character sets sanctioned for use in MARC 21 records, namely, Latin, Chinese ideographs, Japanese hiragana and katakana, Korean hangul, Arabic, Cyrillic, Greek, and Hebrew.

Examples of terms

English is written in Latin script. French is written in Latin script. Text containing English and French is multilingual, but is not multiscript (both languages are written in Latin script).

English is written in Latin script. Russian is written in Cyrillic script. Text containing English and Russian is both multilingual (two languages are present) and multiscript (two scripts are used).

Japanese, a single language, is written using several scripts: hiragana, katakana, ideographs (*kanji*) and occasionally Latin letters (*romanji*). Text containing only Japanese is not multilingual but is multiscript.

Text containing a language written in its true script and also in romanized form is not multilingual but is multiscript. (There is only one language even though it is written in two different scripts).

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.

B. MARC 21 and Unicode

The MARC 21 Specifications[3] dictate computer encoding of the data in bibliographic and authority records for exchange. From its inception, MARC allowed the use of special characters and diacritical marks needed for other languages written in Latin script and for the romanization of languages written in non-Roman scripts according to the ALA-LC Romanization Tables. Over the years, support for a number of non-Roman scripts was added using the "MARC-8" technique modeled on International Standard ISO/IEC 2022: individual character sets identified with distinctive "announcers." Unicode (in the UTF-8 encoding form) was subsequently adopted for all MARC 21 data (including Latin script) as an alternative to the "MARC-8" technique.

The character encoding used in a MARC record is identified in Leader byte 9. If blank, the encoding scheme is "MARC 8;" if the value is "a", the record is in Unicode (that is, in UTF-8).

Committees and Task Forces appointed by the Machine-Readable Bibliographic Information (MARBI) Committee developed mappings for all of the individual MARC 21 character sets. A separate Task Force was appointed to define the attributes of a MARC record containing Unicode data.

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.

---

[3] *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media*, Library of Congress, Cataloging Distribution Service, 2000.

A list of MARC Proposals, Discussion Papers, and reports related to character sets and the use of Unicode in MARC 21 records is at the end of Appendix B.

C. Non-English Support In Library Systems

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.

Implementation and migration to local multiscript systems has begun, but chiefly at the larger libraries. Most smaller libraries, including public and school libraries have not yet addressed the needs of their non-English speaking users for access to materials in other languages.

In the absence of "best practices" for script support in library systems, the language experts in each library will have to define requirements and develop tests to evaluate how well a library system supports desired scripts.

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.

All except one company reported multilingual capability in their products, and most had multiscript capability as well. The product without multilingual or non-Roman script capability is designed for libraries with fewer than 50,000 items.

All companies except one reported that user dialog with the system could be conducted in multiple languages. The number of languages other than English ranged from 3 ("French, Spanish, Italian") to "up to 115." Each company used a difference resource for preparation of the user dialog text.

Searching with languages other than English was covered by questions about multilingual subject access. (If a system has multilingual subject access, it is certain to have multilingual access by author and title as well.) All companies except one provide subject access in languages other than English in their ILS, provided (as was pointed out by two respondents) that records contain such data.

Of the systems providing multilingual subject access, five have authority control for subject access points, the other responded "this is up to the library." Four systems had all

languages in a single combined file (in one case, extraction by language was supported); one system has "multiple subject thesaurus authority files," the remaining respondent said "this is up to the library."

Integrated library systems also support non-Roman scripts. The products of five companies support non-Roman scripts, in one additional product, support of non-Roman scripts is under development, and one product does not have support for non-Roman scripts.

Of the five available products that support non-Roman scripts, four support "all other Unicode-based, Microsoft-supported scripts" (as one respondent expressed it), including all of the "MARC-8" scripts. The other system has only Cyrillic and Greek, plus "Central European" (possibly a Latin script character set for languages such as Czech and Polish).

Because all scripts are equal in Unicode, the multilingual features listed above are not necessarily limited to languages written in Latin script. For example, in response to the question about the multilingual user interface, respondents mentioned Arabic, Chinese, Greek, Hebrew, Hindi, Inuktitut, Japanese, Korean, Russian, Thai, and Ukrainian,

OCLC provides interfaces (i.e. menus, error messages, etc.) for its services in multiple languages and scripts:
- for cataloging in Chinese (traditional and simplified), Japanese, Korean, English, German and Spanish;
- for FirstSearch in Arabic, Chinese (traditional and simplified), Japanese, Korean, English, French and Spanish;
- for Open WorldCat in Dutch, English, French, German and Spanish.

D. Non-Roman scripts in the bibliographic utilities

Facilitating access to non-Roman scripts within libraries in North America has been contingent on developments in three areas:
(1) the ability to create machine-readable records in non-Roman scripts;
(2) the ability for libraries to exchange these records with one another via shared bibliographic utilities; and
(3) the ability for patrons and staff to search, retrieve and display records using these scripts with local library OPACs.

Developments within the bibliographic utilities RLIN and OCLC led the way. Chinese, Japanese and Korean scripts (CJK) were implemented on the RLIN cataloging system in 1983, and by OCLC in 1986.

RLG subsequently implemented Cyrillic script in 1986, Hebrew script in 1988, and Arabic script in 1991.See the RLG report (Appendix I) for additional information.

OCLC added Arabic-script cataloging to its cataloging software in 2000, and by 2005, Cyrillic, Greek and Hebrew scripts were also introduced. OCLC added the Thai and

Tamil scripts as of spring 2006. This is significant development because these are the first scripts that are not also encoded as individual "MARC-8" character sets implemented on a bibliographic utility. See the OCLC report (Appendix H) for additional information.

OCLC also provides interfaces for its services in various languages and scripts (described in the preceding section).

E. Non-Roman scripts in authority files

Bibliographic records, however, are only one part of the puzzle. There is still a need to develop a solution for non-Roman script authority records. Since 2005, the NACO nodes (British Library, Library of Congress, OCLC, National Library of Medicine, and RLG) have been working together on adding the "MARC-8" non-Roman scripts (Arabic, CJK, Cyrillic, Greek, Hebrew) to the LC/NACO authority records.

A significant development in authority work was creation of the multiscript Chinese Name Authority File (HKCAN) by the institutions of the Joint University Librarians Advisory Committee (JULAC) of Hong Kong. Records are in MARC 21 (using the Model B structure), and are based, where possible, on records from the LC/NACO file. The file initially contained about 120,000 records and grows at the rate of 8,000 to 10,000 records a year. The Library of Congress began trial use of the Chinese Name Authority File in January 2003. OCLC began to provide access to the Chinese Name Authority File in July 2005.

Appendix K contains the responses to the *Survey on Non-English Support by Authority Control Vendors*. Three of seven vendors responded (42% response rate).

The ALCTS/LITA Authority Control in the Online Environment Interest Group (ACIG) has sponsored a series of conference programs on non-Roman scripts in shared authority records. For details, see Appendix E.

F. Staffing and economic issues

Cataloging and sharing of bibliographic records containing non-Roman scripts through the bibliographic utilities RLIN and OCLC is discussed above under *Non-Roman scripts in the bibliographic utilities.* Creation of a core of MARC records enriched with non-Roman scripts took place through the activity and efforts of administrators and staff of large academic and research libraries.

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.

Not all non-Roman script content in contributed records is at the same level. This was (and still is) due to a variety of reasons, including labor costs, need, and managerial issues. Some institutions decided that all available scripts were to be input into their records and made every effort to do so. Others made the decision to input some scripts (such as East Asian) but not all (Arabic, Cyrillic). There were those who opted not to include any scripts at all. On the other hand, some institutions took advantage of specially-funded recon projects to increase efforts to contribute records in non-Roman scripts.

In short, the pool of available non-Roman script records has been created largely through the efforts of participating libraries that have been willing to take the risk and effort to contribute such records. The question remains as to where the increase in this pool will come from, as more and more libraries, both public and academic, show interest in access to non-Roman script materials yet are unable to contribute to this pool.

Software packages now recreate original script from romanized text, or vice versa. Even though the output must be reviewed for correctness, this is a useful addition for cataloging. OCLC's Connexion client v. 1.50 onwards supplies Arabic script; the Queens Borough Public Library's "Cyril" program supplies Russian in Cyrillic script, and Joel Hahn of Niles Public Library has created transliteration utilities to and from romanization for Cyrillic, Greek, and Hebrew (see: http://users.rcn.com/aardy/oml/connex.html).

Different projects to introduce vendor records in non-Roman scripts into the bibliographic utilities have taken place; for example, the addition to RLIN of records from A.I. Weinberg Book Agency and from the Toshokan Ryutsu Center. Such projects are usually done in partnership with other libraries and institutions.

Another source could come through encouraging the participation of academic and national libraries overseas; for example, Asian libraries are contributing non-Roman script records to OCLC's WorldCat.

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.

H. Unicode and Libraries

The Unicode Standard is at the core of all modern software, and required by recent IT standards. Using Unicode gives the potential to implement any of its scripts, but actual support for a script requires software for input, indexing, and presentation of the script as well as suitable fonts, and putting this all together takes time.

Similarly, defining the specifications for use of Unicode in MARC 21 records has been a complex task, and is not finished yet. The ability to interchange data reliably requires agreed-upon answers to technical issues followed by implementation in software.

Membership of the Unicode Consortium has a strong library-oriented component. A significant number of corporations providing services to libraries are members of the Consortium:
- the utilities OCLC and RLG
- eight library systems suppliers: Endeavor Information Systems, Ex Libris, Innovative Interfaces, LIB-IT Bibliotheks EDV-Systeme, The Library Corporation, SIRSI, Talis Information, and VTLS
- Google, parent company of the Library Project.

Academic institutions that are Unicode members include two universities, Columbia University and The University of California at Berkeley, and one library, Bibliothèque Universitaire des Langues et Civilisations.

I. Conclusion

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.

The bibliographic utilities and integrated library systems are making substantial progress in adding scripts that have never been supported in the bibliographic environment. The delay in implementation is due to the need to agree on the technical standards that are vital if record retrieval and exchange is to occur without corruption of data.

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

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.

The utilities are making efforts to acquire source records for foreign language material from companies in the book trade. By providing cataloging services to libraries overseas, the utilities will obtain more contributed records for material from those locations.

Ideal requirements to make library resources available to a person who uses a language other than English are:

1. bibliographic records where the language of the catalog is not English but the user's language;
2. authority files to support access in languages other than English;
3. library systems with operating instructions and help in the user's language;
4. the ability to search with the user's language against the library's catalog and databases;
5. presentation of search results ordered by the conventions for the user's language.

The most easily fulfilled requirement is the addition to library systems of operating instructions and help in additional languages. Integrated library systems already provide for this.

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.

Provision of access points and other information in non-English languages is a library-specific decision. The acquisition of records from foreign sources by the utilities will facilitate such cataloging.

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.

The need to present search results in the exact sort order of the user's language should be carefully evaluated.

Appendix A: Charge

Association for Library Collections and Technical Services
ALCTS Executive Committee

Task Force on Non-English Access
Charge
(October 7, 2005)

The Task Force shall examine ALA's past, present, and potential future roles in enabling access to library resources in all languages and scripts and in addressing the needs of users of materials in all languages and scripts through the development of library standards and practices.

In particular, the TF shall examine MARBI, CC:DA, and CC:AAM actions related to non-English language resources, and their role in encouraging and facilitating access to materials in all languages and scripts. The TF shall also examine efforts by online vendors and bibliographic utilities to facilitate balanced access to non-English language materials within their areas of responsibility.

The Task Force shall provide a written report to the ALCTS Executive Committee at its Spring 2006 meeting. This report will include information on access to non-English language materials in library records, catalogs, online systems, and bibliographic utilities, as well as recommended actions for ALCTS to consider for encouraging timely development of non-English language material access and support.

Roster:

Chair:
Beth Picknally Camden
Director, Goldstein Information
Processing Center Van Pelt-Dietrich
Library Center
University of Pennsylvania
phone: 215-746-4149
e-mail: bethpc@pobox.upenn.edu

Diana M. Brooking
Cataloging Librarian (Slavic)
University of Washington Libraries
PO Box 352900
Seattle, WA 98195-2900
work: (206) 543-8405
fax: (206) 685-8782
e-mail: dbrookin@u.washington.edu

Members:
Joan Aliprand
10109 Oakleaf Place
Cupertino, CA 95014
phone: (408) 253-0249
fax: (408) 253-3245
e-mail: jaliprand@pobox.com

Karen Coyle (MARBI)
Digital Library Consultant
Kcoyle.net
work: (510) 540-7596
fax: (510) 848-3913
e-mail: kcoyle@koyle.net

Ann Della Porta
Library of Congress
Washington, DC 20540
work: (202) 707-4761
e-mail: adel@loc.gov

Shi Deng
Head, CJK Acquisitions/Cataloging
UCSD Libraries
9500 Gilman Drive
La Jolla, CA 92093-5004
work: (858) 822-2104
fax: (858) 822-0349
e-mail: sdeng@ucsd.edu

William J. Kopycki (CC:AAM)
Middle East Studies Bibliographer
University of Pennsylvania Libraries
3420 Walnut Street
Philadelphia, PA 19104
work: (215) 898-2196
fax: (215) 898-0559
e-mail: kopycki@pobox.upenn.edu

Heidi G. Lerner (CC:DA)
Hebraica/Judaica Cataloger
Stanford University Libraries
Catalog Dept.
Stanford, CA 94305-6004
work: (650) 725-9953
fax: (650) 725-1120
e-mail: lerner@stanford.edu

Kristin Lindlan
Head, Serials Cataloging Section
University of Washington Libraries
PO Box 352900
Seattle, WA 98195-2900
work: (206) 543-4675
fax: (206) 685-8743
e-mail: klindlan@u.washington.edu

Sally H. McCallum
Chief, Network Development and
MARC Standards Office
Library of Congress
Washington, DC 20540
fax: (202) 707-0115
work: (202)707-5119
e-mail: smcc@loc.gov

David N. Nelson (CC:AAM)
South Asia Bibliographer
University of Pennsylvania Library
3420 Walnut St.
Philadelphia, PA 19104-6278
work: (215) 898-7460
fax: (215) 573-9610
e-mail: nelsond@pobox.upenn.edu

Glenn Patton
Director, WorldCat Quality
Management
Online Computer Library Center
6565 Frantz Road
Dublin, OH 43017-3395
work: (800) 848-5878 or (614) 764-6371
fax: (614) 718-7187
e-mail: pattong@oclc.org

Karen Smith-Yoshimura
Manager, Bibliographic Services
Research Libraries Group
2029 Stierlin Court, Suite 100
Mountain View, CA 94043-4684
work: 650-691-2270
fax: +1.650.964.0943
e-mail: kss@rlg.org

Appendices B-L: Reports

Appendices B-E: ALCTS Sections, Committees and Task Forces

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

The MARC format has accommodated non-Roman character sets since the early 1980's using a technique that is peculiar to MARC, called MARC-8. In 1996, with the advent of Unicode and the Universal Character Set (UCS), work was done to map the MARC-specific character set for the Arabic, Cyrillic, Hebrew, and Latin scripts, as well as Greek symbols, superscripts, and subscripts to UCS/Unicode. Proposal 97-10, "Use of the universal character code set in MARC records," contained four parts, one of which was:

"- Establish that USMARC records employing character codes from the universal
        character set would use only those listed in the USMARC to UCS mapping…"

Although the part of Proposal 97-10 relating to the treatment of "ASCII clones" was approved (as per the updated proposal and the minutes of that meeting), it does not appear that other statements in the proposal were approved, including that of limiting USMARC to the UCS-mapped characters. However, there seems to have been an assumption that such a decision had been made. For example, a later discussion paper (DP 2002-06) states:

> "One of the working principles of the Character Set Subcommittee stated in
> Proposal 97-10 is that round-trip mapping would be provided between the
> MARC-8 set and the accepted repertoire in UCS. This has the effect of providing
> UCS as an alternate encoding without approving any repertoire expansion."

It is unclear whether this is the actual decision of the committee at that time. The minutes state the motion as:
> "Paul Weiss … moved that we take the other three recommendations of Proposal
> 97-10, and turn them over to a technical working group with recommendations
> due in 1998."

> "ACTIONS TAKEN:
> -- Option 1 passed to deal with the ASCII clone issue.
> -- Technical working group will be established to deal with the other issues."

In 1997 a task force was formed to establish the mapping between the MARC East Asian Coded Character set (which cover Chinese, Japanese, and Korean scripts) and UCS/Unicode. These mappings were approved by MARBI in 2001 (DP 2001-09).

In 1998, MARBI approved Proposal 98-18 which defined the method of encoding UCS/Unicode in a MARC format record. This proposal created "leader position 09 as 'Character coding scheme,' with two values defined: blank to signify USM-94 [MARC-

8], and value a to signify Unicode." The proposal also established some key technical points relating to the use of UCS/Unicode:

1. MARC records would use UTF-8 encoding of Unicode
2. A MARC record would use only one encoding, and a MARC file would use only one encoding
3. Restricts values in the Leader to ASCII characters
4. Because of the variable length nature of characters in UTF-8, the proposal defines "character position" in the MARC record format as 8 bits, and deals with non-filing and positional data elements.
5. Requires combining characters in Unicode encoded records to follow those they modify

At the time, the task force stated (Section 1.1):

> "One of the principal motivations for adopting a UCS encoding is to facilitate expansion of the USMARC character repertoire and once this encoding has been specified there will be considerable pressure to use additional characters in USMARC records. The specifics of such expansion lie outside the charge to this Task Force, but 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."

The first instance of such a proposal to expand the USMARC character repertoire was Proposal 2002-11, "Repertoire Expansion in the Universal Character Set for Canadian Aboriginal Syllabics." The CAS does not have an equivalent in the MARC-8 environment. This proposal was approved by MARBI in June of 2002.

In 2004, two reports relating to UCS/Unicode and MARC were presented to the MARBI committee: "Assessment of Options for Handling Full Unicode Character Encodings in MARC 21 -- *Part 1: New Scripts* (January 2004) " and "Assessment of Options for Handling Full Unicode in Character Encodings in MARC 21 -- *Part 2: Issues* (June 2005)." These reports presented options for record sharing in a mixed environment (MARC-8 and expanded UCS/Unicode) as well as technical issues in the adoption of an expanded UCS/Unicode repertoire. These issues include the treatment of UCS/Unicode characters when received by systems using MARC-8, indexing and searching of the expanded repertoire, sorting of fields, record matching, and display.

In 2005, the MARBI committee was presented with a letter from CC:AAM with a motion that passed that committee urging MARBI to allow use of full UCS/Unicode in MARC 21 records. MARBI's chair, Adam Schiff, responded:

> "MARBI considered this motion at its midwinter meeting in Boston in January 2005. The Committee agrees with the intent of this motion and supports the goal of inclusion in the character repertoire of the Unicode Universal Character Set beyond that included in the MARC-8 subset. However, on a practical level, the Committee acknowledges also that character repertoire expansion will not occur in the U.S. environment until some of the technical issues that affect important constituents, such as OCLC and RLG, are resolved."

Of the issues alluded to here, that of compatibility between systems using the MARC 21 repertoire and those using an expanded UCS/Unicode repertoire appeared to be the one is greatest need of solution before the MARC 21 community could embrace a full UCS/Unicode implementation. Library of Congress announced a discussion list for this and other issues (Unicode-marc@loc.gov). Proposal 2006-04, "Technique for conversion of Unicode to MARC-8," is based on discussions at that forum, and was discussed by MARBI in January, 2006, in San Antonio.[4]

---

[4] The results of the discussion are available online at http://www.loc.gov/marc/marbi/2006/2006-04.html. A follow-up document, Proposal 2006-09, "Lossless technique for conversion of Unicode to MARC-8" was considered in June 2006 in New Orleans. This proposal is available online at http://www.loc.gov/marc/marbi/2006/2006-09.html.

MARC Proposals, Reports, and Discussion Papers Related to Character Sets and Use of Unicode in MARC 21

MARC Proposals

Updated version of the list compiled by Gary L. Smith (smithg@oclc.org) and published on the Unicode-MARC discussion list on 5 Aug 2005.

2006-09 Lossless Technique for Conversion of Unicode to MARC-8
http://www.loc.gov/marc/marbi/2006/2006-09.html
2006-04 Technique for conversion of Unicode to MARC-8
http://www.loc.gov/marc/marbi/2006/2006-04.html
2005-05 Change of Unicode mapping for the Extended Roman "alif" character
http://www.loc.gov/marc/marbi/2005/2005-05.html
2004-08 Changing the MARC-8 to UCS Mapping for the Halves of Doublewide Diacritics from the Unicode-UCS Half Diacritic Characters
http://www.loc.gov/marc/marbi/2004/2004-08.html
2002-14 Changes for UKMARC Format Alignment (Section 10: Character Set)
http://www.loc.gov/marc/marbi/2002/2002-14.html
2002-11 Repertoire Expansion in the Universal Character Set for Canadian Aboriginal Syllabics
http://www.loc.gov/marc/marbi/2002/2002-11.html
2001-09 Mapping of EACC characters to Unicode-UCS
http://www.loc.gov/marc/marbi/2001/2001-09.html
98-18 Unicode Identification and Encoding
http://www.loc.gov/marc/marbi/1998/98-18.html
98-16R Nonfiling characters in all MARC formats
http://www.loc.gov/marc/marbi/1998/98-16r.html
97-14 Addition of new characters to existing USMARC sets
http://www.loc.gov/marc/marbi/1997/97-14.html
97-10 Use of the universal code character set
http://www.loc.gov/marc/marbi/1997/97-10.html
96-10 USMARC Character Set Issues and Mapping to Unicode-UCS
http://www.loc.gov/marc/marbi/1996/96-10.html
95-3 Addition of Subfield $6 (Linkage), Field 066 (Character Sets Present) and Field 880 (Alternate Graphic Representation) to the USMARC Holdings Format
http://www.loc.gov/marc/marbi/1995/95-03.html
91-1 on   Non-roman data [Authority]
Not available as Web document
82-17 on Non-Roman alphabets [All formats]
Not available as Web document
80-6 on Non-Roman alphabets [All formats]
Not available as Web document

MARC Report

Assessment of Options for Handling Full Unicode in Character Encodings in MARC 21.
Part 1: New Scripts. A Study for the Library of Congress, by Jack Cain. (2004)
http://www.loc.gov/marc/marbi/2004/2004-report01.pdf.
Part 2: Issues. (2005)." http://www.loc.gov/marc/marbi/2005/2005-report01.pdf.

MARC Discussion Papers

2002-DP07 Changes for UKMARC Format Alignment
http://www.loc.gov/marc/marbi/2002/2002-dp07.html
Predecessor to Proposal 2002-14
2002-DP06 Repertoire Expansion in the Universal Character Set for Canadian Aboriginal
Syllabics
http://www.loc.gov/marc/marbi/2002/2002-dp06.html
Predecessor to Proposal 2002-11
2002-DP05 Guidelines for the Nonfiling Control Character Technique in the MARC 21
Formats
http://www.loc.gov/marc/marbi/2002/2002-dp05.html
See also Proposal 98-16R, DP 118 and DP 102
2001-DP05 Multilingual Authority Records in the MARC 21 Authority Format
http://www.loc.gov/marc/marbi/2001/2001-dp05.html
DP118 Nonfiling characters in MARC 21 using the control character technique
Predecessor to Proposal 98-16R
http://www.loc.gov/marc/marbi/dp/dp118.html
DP111 Alternate graphics without 880 in Bibliographic, Holdings, Authority, and
Community Information records
http://www.loc.gov/marc/marbi/dp/dp111.html
DP109 Identifying transliteration schemes in USMARC formats
http://www.loc.gov/marc/marbi/dp/dp109.html
DP102 Non-filing characters
http://www.loc.gov/marc/marbi/dp/dp102.html
DP100 Recording additional characteristics in USMARC Authority records
http://www.loc.gov/marc/marbi/dp/dp100.html
DP 73 on Unicode and USMARC character set mappings [All formats]
Not available as Web document
DP41 Non-Roman headings, Techniques for handling [Authority]
Not available as Web document
DP31 on Non-Roman data [All formats]
Not available as Web document

Compiled by Joan M. Aliprand, 6/21/06

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

CC:DA and JSC Actions Relating to Non-English Resources and Non-Roman Scripts

Adding Vernacular Names to the LC Name Authority Files (ALCTS/CCS/CC:AAM Program)

At ALA Annual Conference 1997, CC:DA agreed to support "in name only" the ALCTS/CCS/Committee on Cataloging: Asian and African Materials' program on "Adding Vernacular Names to the LC Name Authority Files" for the 1998 conference.

Additions to names entered under given name, etc. (AACR2 22.16)
[4JSC/BL/4 series]

At Midwinter 1999, CC:DA discussed [4JSC/BL/4] Additions to names entered under given name, etc. (22.16) An inconsistency between the proposed rule and rule 22.16A1 was noted, i.e., the proposed rule specifies that the title of nobility is to be given in the vernacular, while rule 22.16A1 specifies that the title be given in English if there is a "satisfactory English equivalent." As suggested in the BL proposal, CC:DA supported creating a new rule 22.16B1 and renumbering existing rules to accommodate this. The Committee disagreed with the BL on the text of the rule. To maintain consistency within the code, CC:DA suggested that the new rule read "Add, to the name of a nobleman or noblewoman entered under given name, the title of nobility (in English if there is a satisfactory English equivalent)" and that the example be revised accordingly, in keeping with the instruction given in rule 22.16A1.

[4JSC/BL/4/BL follow-up] was a reiteration of the position taken in the original [4JSC/BL/4] document. CC:DA suggested in its response that the additional proposed example should be in English, but BL did not believe the title of nobility in that example could be satisfactorily translated. What was more substantive in Brian Schottlaender's opinion was the statement in the middle of the third paragraph [4JSC/BL/4/BL follow-up] in which it says that in general it's better not to translate additions to names. Therefore, BL continued to propose 22.16B1 wording as in [4JSC/BL/4] and renumbering of following sub-rules. LC and CCC did not support this and CC:DA didn't support in first instance. It also seemed odd to Schottlaender that it's better not to translate additions to names, but BL did not go on suggest that the parenthetical in 22.16A1 be deleted, which would be consonant with that philosophy. Schottlaender, ALA Representative to the JSC asked if CC:DA wanted to restate its non-support.

CC:DA expressed support for LC position, but at the JSC Meeting, 13-15 September 2000, Additions to names [4JSC/BL/4, etc.]
proposal was withdrawn by the BL.

ANSI/NISO/ISO 3166 (Codes for the Representation of Names of Countries)

At ALA Midwinter Meeting 1995, the Report of the Task Force to Review Proposed American National Standard (`CC:DA/TF/ANSI 3166/6 and 6s`) was considered. ANSI/NISO/ISO 3166 contained many editorial inconsistencies and inconsistency in the presentation of headings (some were in direct order, some in inverted form). The headings were supposed to be politically neutral, but there were some problematic headings, such as "Falkland Islands (Malvinas)" and "Taiwan (Province of China)." Overall, however, the Task Force said that the list was good enough for the CC:DA Chair to vote "yes" on, with some suggestions for editorial changes.

Breton Initial Articles

LC submitted a proposal [`5JSC/LC/7`] related to Breton initial articles, which was available for review at the April 2006 JSC meeting, with CC:DA response due Sept. 18, 2006.

Dates (and chronological designations) not in Western-style Arabic numerals.

At the JSC Oct. 2005 meeting, it was reported that draft RDA 2.10.0.3 would provide an option for adding a date in Western-style Arabic numerals following a date not in Western-style Arabic numerals. In addition, RDA 1.5 (Language and Script of the Description) would provide an option to record any elements (including dates) in a transliterated form when that element cannot be recorded in the script used on the source. The Library of Congress was considering whether to present a separate proposal to further internationalize the recording of dates; such a proposal would likely be considered by the constituencies along with the draft of Part I.

East Asian Character Code for Bibliographic Use (`Z39.64 (proposed)`). The proposal was considered outside of CC:DA's scope, but the chair referred it to CC:AAM for input spring 1995. However, he received no response from CC:AAM before the deadline, and therefore did not vote.

German Capitalization

At the ALA Annual 2003 meeting, a rule revision proposal received from Croissant (John Hostage presenter), Revision of Appendix A, Capitalization, A.40 German [`CC:DA/Croissant/2003/1`] was considered. There had been spelling reform in Germany regarding capitalization and the proposal contained revisions to appendix A to conform to the spelling reform. CC:DA decided that the proposal looked fine, but asked for the proposal to come back to the Committee with examples in the code reviewed. At ALA Midwinter 2004, this rule revision proposal from Croissant [`CC:DA/Croissant/2003/1/Rev`] was discussed. A task force had reviewed examples in the rules to see if any needed to be revised to conform to the rule proposal, but no impacted examples in the rules were found. CC:DA approved a motion to accept the proposal and send it on to the JSC.

At the JSC meeting, 19-22 April 2004, rule proposal [4JSC/ALA/56] on Rule A.40 German orthography was agreed to with British Library and CCC amendments and with ALA to confirm non-capitalization of "acht". Rule A. 40 was included in the revision packet for 2005 resulting from [ALA/56] series on German capitalization.

Initial Articles Appendix

At ALA Midwinter Meeting 1997, CC:DA approved Initial Articles [3JSC/LC/29], although subsequently it was learned that a couple of initial articles were missing from its list. ALA endorsed the proposal with the addition of the missing articles. Afterwards, additional omissions were identified and OCLC prepared a revised proposal on CC:DA's behalf [CC:DA/Patton/1]. After the proposal was finalized, it was forwarded to JSC as [3JSC/LC/29/ALA response/2].

At ALA Midwinter Meeting 1998, the pending rule revision proposal on Initial Articles [3JSC/LC/29] was referred back to ALA for the addition of an introductory sentence apropos of the list's not attempting to be an exhaustive/comprehensive one but, rather, one including articles for the languages listed.

At the JSC fall 1999 meeting, the proposal to create a new Initial Articles Appendix [4JSC/LC/29/ALA response/3] was approved. LC agreed to finalize the document, incorporating input from [4JSC/LC/29/ALA response/3/CCC response]. LC and NLC agreed to jointly serve as the "maintenance agency" for the new Appendix. At the JSC fall 2000 meeting, the rule revision proposal was approved for inclusion in the next revision package: Initial Articles [4JSC/LC/29, etc.]

Internationalization of AACR2 and RDA

1. At the ALA Midwinter Meeting 1998, Brian Schottlaender, ALA Representative to the Joint Steering Committee of AACR, reported on the internationalization of AACR2. The JSC planned to confirm whether surveys had been conducted of non-author country use/interest in AACR2 and, if none had been done, conduct one. Language would be added to the AACR website apropos of proposals from non-author countries coming through appropriate national cataloging channels to the Chair of JSC.

At ALA Annual Conference June 1999, Schottlaender reported that in February and March 1999, Ann Huthwaite, Chair of JSC, began drafting a "White Paper" on AACR2 Internationalization. At the Committee of Principals meeting in March, it was agreed that it would be more appropriate for the CoP to consider those issues and Huthwaite continued to work on writing the paper. At the JSC fall 1999 meeting, she provided a draft document for CoP's consideration.

2. The JSC discussed internationalization both broadly and narrowly at its May 2002 meeting [4JSC/Chair/62]. Regarding the Malay names and related issues, the JSC wanted AACR to follow international standards and reflect local expertise on usage. The

Australian Committee on Cataloguing (ACOC) agreed to take the lead on follow up, comparing AACR rules for Malay names to the IFLA Names of persons and working with such interested groups as the Malay Committee on Cataloguing. ACOC expanded this work to include other names, e.g. Indonesian names, and was charged to draft rule revision proposal(s) by Aug. 9, 2002 for preliminary discussion at the September 2002 JSC meeting. In Malay Names [4JSC/ACOC/2], it was reported that ACOC had checked IFLA *Names of Persons* and turned it into rules for AACR (not what JSC had intended, but per Huthwaite's instructions), which constituents were to respond to. The JSC hoped to reduce the size of rules by just referring to the IFLA *Names of Persons*, if it worked for those languages.

Constituents responded on Malay and Indonesian names [4JSC/ACOC/2] for discussion at the April 2003 JSC meeting. LC and BL responses did not agree with the ACOC proposal; extensive JSC discussion led to the decision for ACOC to withdraw [4JSC/ACOC/2]. It was noted within ALA that CC:AAM did not provide input on the ALA position. Two changes from the LC response document entered the rules with the 2004 update package: an error in one name spelling was corrected and a footnote extended. ALA and CILIP responded by May 12, deferring to experts at LC and BL.

3. At ALA Midwinter 2005, Keiko Suzuki to read a statement sent by the CC:AAM to CC:DA: "The CC:AAM has discussed the draft of AACR3, Part 1, and is deeply concerned with its application of cataloging rules to the language of our regions. The existence of cataloging manuals for various languages is witness to the inadequacy of the current rules. CC:AAM will therefore forward to CC:DA its concerns for consideration."

Suzuki also said CC:AAM would forward, before the Feb. 11 deadline, the comments citing specific places where the current rules inadequately address cataloging situations for material of these regions with proposed solutions. CC:AAM was deeply concerned about the application of languages/scripts in AACR. Previously, a cataloger could use scripts from the item on card catalogs. The ability to use an item's own script was lost and is still unavailable for many languages/scripts when cataloging was moved to computers. Relying on systematic romanization, as had been done, was not a perfect solution for access and is a time-consuming cataloging practice for many languages. They understood there were technical issues involved, and it was not really purely about cataloging rules, and yet, AACR has a preference for English. CC:AAM would like AACR to clearly acknowledge the romanization practice for non-Roman scripts materials. Jennifer Bowen, ALA representative to the JSC, suggested that a discussion paper would be helpful.

4. At the JSC meeting, Oct. 2005, it was mentioned that draft RDA 2.10.0.3 will provide an option for adding a date in Western-style Arabic numerals following a date that is not in Western-style Arabic numerals. In addition, RDA 1.5 (Language and Script of the Description) will provide an option to record any elements (including dates) in a transliterated form when that element cannot be recorded in the script used on the source. The Library of Congress was considering whether to present a separate proposal to further internationalize the recording of dates.

At ALA Midwinter January 2006, several relevant points were discussed:

5. There was discussion of the rule concerning words or letters used repeatedly, i.e., the *Canadian BIBLIOGRAPHIES canadiennes* example. This is a result of the intent of the designer to present something eye-catching, but it's not necessarily how the resource is presenting itself. It was commented that there is a divide between describing resources that are self-describing and those that aren't – and that RDA has not decided how it wants to deal with resources that are not self-describing and another person stated that RDA should adhere to strict transcription, if RDA is resource-focused, but if user-focused then use "intelligent transcription." Some choices suggested: conform to principle of transcription but provide an alternative access point; or make an intelligent correction and provide the transcription as the alternative. It was pointed out that we have many people dealing with materials when they are not familiar with the language – and that may be facilitated with transcription from the title page is the way to go, despite the concerns.

6. The CC:AAM liaison commented on the option in RDA 1.5, which doesn't allow for a transliterated form -- CC:AAM would like to add the option that the agency can put it in a transliterated form.

7. Concerns were raised about RDA 2.3 regarding parallel and variant titles. In RDA 2.3.1.12, there will be implications for major and minor changes, if for example the French part is dropped from the *Canadian BIBLIOGRAPHIES canadiennes*. Also, the rule doesn't say that "parallel other title information" comes from the same source as "parallel title" – which generated much discussion about the source of "parallel title" and "parallel title other information."

8. For RDA 2.3.0.7, a comment was made that we should return to the practice – with respect to layout of the title page – of using the "lower" title as "title proper" and the "upper" title as "series." In broadening RDA beyond the Anglo-American cataloging community, grammatical dependence or independence may be more complicated in languages that do not use cases.

9. RDA 2.5.1.2 doesn't address multiple edition statements from multiple places in the resource, which is prevalent in Chinese publications, where the cover might state that a resource is the revised edition, but the colophon still states it is the first edition. However, in trying to simplify the rules, it was stated that we need to focus on whether the general principle is sufficiently clear for a cataloger to apply in these circumstances. This problem also occurs in Western publications, so a wording change to allow for the possibility of multiple statements was suggested by the ALA Representative to the JSC.

10. The JSC representative summarized other topics of concern, including that the rules are too Western-centric. She asked for specific examples and suggestions for how to fix this.

11. After ALA Midwinter 2006, Library of Congress presented a proposal [5JSC/LC/5] for possible changes to the text of the December 2005 part I draft, with the goal to "make *RDA* open to use by any community with a context other than English language, other

than Latin script, other than Western-style Arabic numerals, and/or other than Gregorian/Julian calendar." Rule revision proposals included rules related to: numbers expressed as words and numerals; dates in calendars not used by the agency preparing the description; recording chronological designations; transcription of date of publication; chronograms; language and script of the description; recording of edition statements; new sequences of numbering. As a result of responses from the constituencies, including CC:DA (ALA), and discussion at the spring 2006 JSC meeting, the Library of Congress is going to revise the proposal. CC:DA's comments included a need to reference transliteration schemes in RDA and the need for additional elements for RDA to record (i.e., language of the catalog record; script of the catalog record; use of transliteration and its scheme). Such data elements would make it easier to identify a need for further manipulation of records and make sharing of records across language communities "much more likely and much more fruitful."

ISO Harmonization

At 2002 Midwinter, a CC:DA Task Force on ISO Harmonization was charged with identifying and examining ISO (International Organization for Standardization) standards relevant to AACR2 resource description, focusing in particular on standards for abbreviation, symbolization, and punctuation and consulting, as needed, with the ALA representative to the National Information Standards Organization (NISO); identifying areas where ISO standards and AACR2 differ; discussing whether harmonization of ISO and AACR2 is desirable or necessary; and, if deemed appropriate, proposing rule revisions to AACR2.

The task force looked at standards for transliteration, romanization and encoding of scripts. From its final report: "The task force discovered a plethora of ISO standards for the transliteration of various languages' scripts, rendering of foreign-language diacritics, and encoding schemes such as UNICODE for the transmission and display of non-Roman scripts. Several responses from the cataloging community also expressed a preference for a given ISO standard for romanization over the romanization scheme currently adopted by that cataloger's cataloging agency."

Although the primary focus of the task force work was on abbreviations and metric symbols, it also recognized that ISO standards are relevant to bibliographic description, and may someday be an important part of the cataloging code (given the research being done by the IFLA Section on Cataloguing on the feasibility of creating international multilingual authority files and also the efforts to harmonize AACR2 with ISSN), but felt that consideration of these ISO standards at this time is not only beyond the scope of this task force's charge, but also beyond the scope of AACR2, since AACR2 did not currently prescribe any romanization scheme for any foreign script. The ISO standards relevant to description and/or access were:

- ISO 2788:1986 Guidelines for the establishment and development of monolingual thesauri

- ISO 5964:1985 Guidelines for the establishment and development of multilingual thesauri."

The task force made no proposals on romanization or foreign scripts, though a proposal on metric symbols was sent to the JSC.

Maori Initial Articles

ACOC plans to submit a proposal in August 2006, with CC:DA response due Sept. 18, 2006.

Names of persons

At Midwinter 1995, CC:DA discussed the LC suggestion that CC:DA provide input into future updates of Names of Persons. The portion on Francophone African names was forwarded to CC:AAM.

Non-Roman Scripts

January 9, 1999, James Agenbroad submitted to ALA/ALCTS/CC:DA a proposal entitled "Nonroman Headings and References" [`CC:DA/Agenbroad/1`] and, at the ALA Midwinter Meeting 1999, he provided a brief summary of his proposal to codify the practice of providing vernacular access points in addition to the romanized access points. CC:AAM sought feedback from the Committee on East Asian Libraries, the Association for Asian Studies, and the Middle East Librarians Association on the proposal.

Agenbroad proposed to add six new rules to AACR2 chapters 22 through 26 (four rules for chapters 22-25 dealing with forms of headings; two rules for Chapter 26 dealing with references) to allow for the addition of access points in non-Roman scripts when possible technically and appropriate to the item being cataloged. Agenbroad noted that rule 1.0E already allowed titles to be transcribed in the vernacular "whenever practicable." He also amended his written proposal for rules 24.1B2 and 25.2D2 to include the word "non-Roman" between the words "pertinent scripts" in the first sentence of each rule.

CC:DA agreed that it would be inappropriate to act on this proposal prior to receiving a recommendation from the Committee on Cataloging: Asian and African Materials (CC:AAM) and asked it to provide input. CC:AAM sent CC:DA the *Final report of the CC:AAM Non-Roman Access Points Subcommittee*.

At ALA Annual Conference 2000, CC:DA discussed the *Final report* and an Agenbroad alternative proposal [`CC:DA/Agenbroad/2`]. CC:DA preferred to accept the CC:AAM recommendations, i.e., to table the proposals, or vote to establish a task force.

ALCTS/CCS/CC:AAM liaison Sarah Su-erh Elman reported to CC:DA that the Non-Roman Access Points Subcommittee of CC:AAM had examined the documents listed in the report, and some other email messages that were not listed, and had discussed the issue in CC:AAM meetings at ALA Annual in 1999 and ALA Midwinter in 2000. In

principle, CC:AAM supported the idea of including additional non-Roman access points in both bibliographic and authority records to facilitate the retrieval of items in non-Roman scripts. The subcommittee agreed that Mr. Agenbroad's proposal had merit, but felt that it did not adequately address all of the broader issues surrounding the use of additional non-Roman headings and references in library catalogs. Some of these issues included the current inability of library systems to manage non-Roman-script data, and the lack of adequate provisions for integrating non-Roman headings into the current concept of authority control.

The subcommittee gave three recommendations: that CC:DA temporarily table the rule revision proposals; closely monitor the activities of the MARBI Multilingual Records Task Force; and, reopen the discussion at a later time, perhaps after ALA Annual in 2001, at which time the MARBI Multilingual Records Task Force was scheduled to have finished its report on multilingual issues related to MARC 21. At that time CC:DA could consider how the completed MARBI report related to other work in progress at MARBI, PCC, and in some groups that are part of IFLA. The subcommittee suggested that CC:DA could then work with Mr. Agenbroad and reevaluate his proposals in light of the aforementioned work.

In CC:DA discussion, several people said they were not convinced that the problems with authority record structure and MARC format coding needed to be solved before CC:DA could decide what the descriptive conventions ought to be. CC:DA could make a statement that as a matter of principle, non-Roman access points can be provided if a library wished to provide them – especially since that which was being discussed was an option, and also an important step toward providing access to users who search with non-Roman scripts.

CC:DA had an either/or choice of whether to table the Agenbroad proposals, as the report from CC:AAM recommended, or to form a task force to explore either one or both of the proposals that Agenbroad submitted. It was moved and seconded that CC:DA approve CC:DA/Agenbroad/2 revised as follows:

> 20.5 Nonroman access points. *Optionally,* ~~when technically feasible~~ libraries whose collections include and whose clientele seek items in nonroman scripts ~~should~~ <u>may</u> assign nonroman access points to records of such items. Apply the relevant rules of chapters 21-26 and the conventions of reference tools on persons, corporate bodies and titles using nonroman scripts to determine the choice and form of such access points and of cross references to them.

There was discussion on wording and whether it was clear or not that what was being proposed by Agenbroad was to allow non-Roman access points *in addition to* the roman-alphabet headings that they parallel. Problems with including non-Roman access points in bibliographic and authority records were mentioned. Barbara Tillett noted that IFLA had been addressing this for some time, and that she had served on several committees that were looking into this, as well. IFLA was taking the approach that UNIMARC and MARC formats already accommodated this kind of capability. She added another

possibility, i.e., that two authorized forms according to different languages may need to be linked within a given record, and that was another experiment that an IFLA group was doing with the Deutsche Bibliothek.

Adam Schiff restated the proposal with the changes in wording that had been proposed in this discussion since his original motion. Schiff said that the proposed addition to chapter 20 would now read:

> 20.5 Nonroman access points. *Optionally,* assign nonroman access points for resources in nonroman scripts. Apply the relevant rules of chapters 21-26 and the conventions of reference tools on persons, corporate bodies and titles using nonroman scripts to determine the choice and form of such access points and of cross references to them.

Other wording suggestions were considered and Schiff revised the proposal to:

> 20.5 Nonroman access points. *Optionally,* assign nonroman access points. Apply the relevant rules of chapters 21-26 and the conventions of reference tools on persons, corporate bodies and titles using nonroman scripts to determine the choice and form of such access points and of cross references to them.

CC:DA voted 7-0 to accept proposal and forward it to the JSC.

At the JSC meeting in September 2000, the proposal on Non-Roman Access Points [4JSC/ALA/32 series] was deferred, pending input from constituencies other than LC (the only constituency to have responded by the London meetings) – but approval looked unlikely.

All constituent responses to the document were negative, suggesting that non-Roman access should be accommodated via authority records rather than added entries. ALA was asked to consider withdrawing the document and determining whether it wished to pursue the larger authority record issue instead. At 2001 ALA Annual, the Non-Roman Access Points [4JSC/ALA/32] series was withdrawn by ALA.

Names of persons, See references (26.2A) (note: romanization mentioned in rule)
[4JSC/BL/5 series]

[4JSC/BL/5] proposed substantial additional examples, mostly by LC; ACOC, CCC, LA, and LC supported it. [4JSC/BL/5] had already been approved by all constituencies, but in approving the need for more examples became apparent and [4JSC/BL/5/BL follow-up] included these examples.

At Midwinter Meeting 1999, CC:DA supported the proposal [4JSC/BL/5] Names of persons, See references (26.2A), but noted that "Sir" now needed to be removed from the parenthetical note in the corollary example at rule 22.6B3.

The JSC approved [`4/JSC/BL/5`] at the Brisbane meeting in the fall 1999.

Turkish word 'bir'

At Annual 2003, Rule revision proposal from CC:AAM (originating from the Middle Eastern Library Association): Lin (for Deng) for deletion of Turkish word 'bir' from Appendix E, Initial articles: [`CC:DA/CC:AAM/2003/1`] was discussed. Selina Lin reported that Shi Deng had contacted the Turkish National Library regarding this proposal and suggested CC:DA delay action on the proposal until a response was received. It was also suggested that the proposal include a review of all the examples in the code. The question in the proposal about the list "Initial Definite and Indefinite Articles" in MARC 21 was answered: if the proposal is eventually approved by the JSC, the MARC list would be changed according to the changes in appendix E of AACR2.

An informal contact indicated that the National Library of Turkey would agree with the revision. This proposal was to be resubmitted to CC:DA at the Midwinter Meeting 2004 as [`CC:DA/CC:AAM/2003/1/Rev`], with the response from the Turkish National Library and revised examples as necessary.

A motion to accept the proposal, and to attach the letter from the National Library of Turkey when it arrived, was approved by CC:DA. It was also noted that some ALA filing rules would need to be changed after the revision was made and Mary Larsgaard, Chair, agreed to talk to Don Chatham, ALA Publishing about it.

At the JSC meeting, 19-22 April 2004, Ottawa, Canada, JSC agreed with [`4JSC/ALA/57`] to delete 'bir' from the list of initial articles and the revision was included in the rule revision packet for 2005.

Vietnamese names

ALCTS/CCS/CC: AAM had withdrawn its initial support for [`3JSC/LC/24`] (16 January 1996), which was an effort to eliminate the artificial comma as a filing device inserted in the middle of Vietnamese personal names. This was strongly supported by the Australians; neither the British Library nor the Library Association supported it; the Canadian Committee on Cataloguing was neutral. One of the issues that had come up was that Chinese and Hungarian names are also affected by this practice of inserting commas. Australia came to JSC indicating that what is in the BL response was no longer official Vietnamese practice, which apparently coincides with the first example in the BL response. CC:AAM discussed it at length and the committee was very split. If the proposal were expanded to include Chinese and Hungarian names there would likely be more support for the proposal. CC:AAM had not had the Australian response when they were discussing the issue in May and, if the Australian document were accurate regarding Vietnamese practice, that would change the discussion. At the July 1997 ALA Annual Conference, CC:DA voted 5-3 to reject the proposal, but after consultation with CC:AAM, said it would be willing to consider a proposal expanded to include Chinese and Hungarian names.

Family names

Library of Congress was asked by the JSC to submit a proposal [5/JSC/LC/6] for including family names as name headings; they submitted a proposal for a non-unique approach to the form of heading for family names: the same heading would be used by any family using that surname, as is done in *Describing Archives: a Content standard (DACS)*. CC:DA's response included a comment that "naming conventions vary among different cultures as well as time periods. CC:DA recommended that the proposed rules be expanded to address how to construct family names in places where surnames are not used (e.g., Iceland and much of Southeast Asia) and place in which surnames are used but family members do not necessary share the same surname (e.g., in ancient Scottish and modern American families). At its spring meeting, JSC agreed to continue to discuss what will be included in RDA in terms of access points for families.

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

Reports of activities

Actions (e.g. Proposals to JSC) and outcomes

Romanization tables for Oirat, Manchu, and Mongolian (1997 Midwinter-2001 Midwinter)
Starting at 1997 Midwinter Meeting, Hideyuki Morimoto chaired a subcommittee working with Wayne Richter to create romanization tables for Oirat and Manchu, and a revised table for Mongolian.

Non-Roman Headings and References Proposal (1998 Midwinter-2000Annual)

This proposal for adding rules to AACR for the optional addition of non-Roman access points was initially brought to CC:AAM for review by James Agenbroad at the 1998 Midwinter Meeting, and was forwarded by CC:AAM members to area studies associations for comments. It was submitted to CC:DA originally at the 1999 Midwinter Meeting. CC:DA referred it to CC:AAM which appointed a subcommittee chaired by Dawn Bastian to study it and make recommendation to CC:DA.

CC:AAM submitted its final report on Feb. 21, 2000 with three recommendations that CC:DA temporarily table the rule revision proposals; closely monitor the activities of the MARBI Multilingual Records Task Force; and, reopen the discussion at a later time.

An Investigation of Chokwe Language Report (1999 Midwinter-1999 Annual)
At the 1999 Annual Conference, Dawn Bastian reported that in our last meeting in Philadelphia we discussed an e-mail question from Autocat regarding the Chokwe language, which is one of the South African languages. She investigated the issue. From all sources she had, what she was able to determine was that the languages use the letters of roman alphabets and there were not any special characters involved in this language.

Discussion of List of Asian and African language cataloging experts (2000 Annual)
A discussion held at the 2000 Annual Conference as to whether CC:AAM should create a list of Asian and African language cataloging experts. It was mentioned that several area studies librarian associations maintained their own list of language cataloging experts. After discussion, it was decided that CC:AAM would neither create any such list nor include hyperlinks to those existing lists from within the CC:AAM web page.

Proposed Changes in Chinese Romanization Guidelines (Umlaut, Apostrophe): (2002 Midwinter-2002 Annual)
At the 2002 Midwinter Meeting, in the absence of Thompson A Yee (LC CPSO), Philip Melzer (LC RCCD) summarized the recommendation regarding proposed changes in usage of the umlaut and apostrophe in "Chinese Romanization Guidelines," which had been forwarded to CC:AAM prior to the meeting by Yee. After discussion, CC:AAM

voted to approve the recommendations. At the 2002 Annual Conference, after some discussion, the revised Guidelines were approved as presented.

Use of the Word "Oriental" In AACR (2002 Midwinter-2003 Midwinter)
In August 2001, CC:AAM received a request from John Attig, chair of a CC:DA task force to consider usage of the word "oriental" in AACR as they received several comments indicating that the term "oriental" is offensive and should not be used in AACR. At the 2002 Midwinter Meeting, the committee discussed the varying meanings of "oriental" in addition to the concern about the use of the term. A subcommittee was formed and chaired by Chalermsee Olson to write a proposed new rule on the issue, after historical review. Because no report was submitted at the 2002 Annual Conference, a new task force chaired by Reiko Yoshimura was formed. After thorough investigation, the task force submitted a report at the 2003 Midwinter Meeting which was later forward to CC:DA.

Batak Romanization table (2002 Annual)
Batak language is a Southeast Asian language. The proposal for a new Batak romanization table came from email exchanges prior to the meeting among Lucas Graves, Randall K. Barry, and Thompson A. Yee. The drafted proposal appeared in Cataloging Service Bulletin, no. 92. It was approved as proposed at the 2002 Annual Conference.

Initial Articles "Bir" In Turkish (2002 Annual-2004 Midwinter)
In February 2002, CC:AAM received a request from CC:DA to review a question regarding the Turkish article "bir." It was discussed at the 2002 Annual Conference. After receiving feedback from MELA the committee agreed to delete the initial article "bir" from AACR2 Appendix E. It was presented to CC:DA at the 2003 Annual Conference with a recommendation of getting further input from the Turkish National Library. This proposal was revised and resubmitted to CC:DA at the 2004 Midwinter Meeting. CC:DA approved it to be forwarded to JSC with supporting documentation from the Turkish National Library.

Initial articles in Tongan/Maori/Samoan (2002 Annual-2004 Midwinter)
In February 2002, CC:AAM received a request from CC:DA to review a question regarding initial articles in Tongan/Maori/Samoan. It was tabled at the 2002 Annual Conference, due to the lack of an area expert on the committee. At the 2003 Midwinter Meeting, the committee asked Michael A. Chopey, CC:DA member, to prepare a proposal. However, there are a number of variant orthographies in Tongan, and area specialists are unable to determine that there is an official one. The proposal was subsequently dropped.

Proposal for new Islamic law subject headings: **(2003 Midwinter-Ongoing)**
At the 2003 Midwinter Meeting, it was reported that the Stanford University Law Library is working on an Islamic law subject headings proposal that will provide better coverage. He proposed that a task force be formed to work with the CCS Subject Analysis

Committee (SAC). In 2005, a list of current subject headings was posted on the MELA website at http://www.mela.us/islamicsh.html.

Romanization tables for the Central Asian languages: **(2003 Midwinter-Ongoing)**
At the 2003 Midwinter meeting**,** John Eilts pointed out that the Central Asian languages that were formally using Cyrillic alphabet need new romanization tables. He suggested finding language experts and working with the ACRL Slavic East European Section.

Malay Names (2004 Midwinter)
In August 2002, the Australian Committee on Cataloguing submitted a proposal on AACR2 22.27. Malay Names. It was brought to this committee for comment. During 2004 Midwinter meeting, the committee agreed to pass CORMOSEA Southeast Asia specialists' comments to CC:DA.

Revised Kurdish Romanization table (2004 Midwinter)
At the 2004 Midwinter meeting the committee received a letter from David Hirsch, MELA past president that reported that MELA endorsed the revision at its November annual meeting. The committee approved the revision.

James Agenbroad's proposal (2004 Midwinter-2005 Annual)
In September 2003 CC:AAM received James Agenbroad's proposal to consolidate MARC-8 character repertoire expansion. The proposal was subsequently discussed by CC:AAM at its 2004 Midwinter meeting. After extended discussion the committee conceptually agreed with the proposal, which called for the character repertoire to be extended and that library systems display and access non-Roman scripts beyond JACKPHY. The committee did not however reach consensus with the technical approach Agenbroad proposed. As a consequence the proposal was tabled. Two follow-up decisions were made: 1) To study the proposal further and solicit technical guidance from system vendors; and 2) To get input from area studies specialists about prioritizing scripts in the ALA/LC Romanization Tables for expansion in the MARC character repertoire, making recommendation to MARBI through ALCTS' official liaisons.

At the 2004 Annual committee meeting members agreed that this committee doesn't have the technical expertise to review and approve this proposal. The Committee instead approved a motion urging MARBI to expand the character repertoire to include the Unicode Universal Character Set (UCS) beyond that already implemented in the MARC-8 character set. The motion was forwarded to Adam Schiff, MARBI Chair, after the 2004 Annual meeting.

The CC:AAM motion to MARBI was discussed at MARBI's Saturday meeting at the 2005 Midwinter meeting. A written MARBI response sent to CC:AAM on June 13, 2005 MARBI stated that it agreed "with the intent of this motion and supports the goal of inclusion in the character repertoire of the Unicode Universal Character Set beyond that included in the MARC-8 subset. However, on a practical level, the Committee acknowledges also that character repertoire expansion will not occur in the U.S.

environment until some of the technical issues that affect important constituents, such as OCLC and RLG, are resolved."

Concurrent with this, Shi Deng, CC:AAM Chair approached John Espley, the AVIAC (Automation Vendors Information Advisory Committee) representative to MARBI, with CC:AAM's concerns. Espley asked CC:AAM to prepare a technical issues questionnaire that he could forward to AVIAC members. At its 2005 Midwinter meeting, CC:AAM took the following actions: 1) A need-based, prioritized list of non-Roman languages and scripts would be compiled; 2) A literature review would be conducted; and 3) A questionnaire would be sent to vendors to determine the status of Unicode UCS and non-Roman languages and scripts access implementation. Based on these investigations CC:AAM hoped to determine what the appropriate next steps for advocating faster expansion of access to non-Roman languages and scripts beyond JACKPHY.

In May 2005, the CC:AAM received James Agenbroad's resolution on equal access to non-Roman resources. At its meeting in June 2005 CC:AAM considered Agenbroad's resolution and had some concerns about it. Nevertheless, notwithstanding these reservations the committee passed a motion supporting the resolution.

DDC draft proposal for Indonesia provinces (2004 Annual-Ongoing)
Giles Martin, Dewey Decimal Classification Assistant Editor, brought a draft proposal of an expansion of DDC to cover all the provinces of Indonesia. The Committee sent out the proposal to appropriate groups for comment. At 2005 Annual it was noted that nothing had been heard yet from the Indonesian National Library.

ALA/LC Chinese Pinyin Romanization Guidelines, Feb. 2004 rev. (2004 Annual-2005 Midwinter)
These were reviewed and discussed. This revision proposal recommended three changes to promote application consistency of the Guidelines. There was a concern about possible misinterpretation of one instruction in section 2B of Connection of Syllables. The committee recommended that this instruction be re-written. LC responded on August 6, 2004 that two examples were added to clarify the section's intent.

ALA/LC Ladino Romanization Table proposal (2005 Annual)
The committee members approved the romanization table without consulting experts in the field since they are the people who put this proposal together or who have been consulted.

Task Forces

RDA/AACR3 (2005 Annual-Ongoing)
CC:AAM established a task force to review RDA. The task force would be chaired by the CC:AAM liaison to CC:DA. Two to three experts from each of the five area studies regions are involved.

Education (programs, preconferences, workshops)

Program for 2000 Annual Conference: *Year of the Dragon, Mandate for Change: Challenges of Pinyin Conversion in the Online Environment*
The program focused on LC, RLG and OCLC's planned switch of the Chinese romanization scheme from Wade-Giles to Pinyin and conversion of bibliographic and authority records beginning in October 2000. There were approximately 100 participants in attendance. The audience suggested a following-up program on the project.

Program for 2002 Annual Conference: *Year of the Horse: Next Phase in Pinyin Conversion*
The program had a panel of 5 speakers, who shared their experiences with conversion clean up, explored issues, problems and successes from a variety of viewpoints and a review of best practices. There were approximately 50 participants in attendance.

Program for 2004 Annual Conference: *Library catalogs and non-Roman scripts: Development and implementation of Unicode for cataloging and public access*
This joint program with ACRL/SEES included four speakers presenting the current and future development of Unicode in library catalogs. Approximately 200 people attended, which was very well received.

Appendix E: Other ALCTS groups (in alphabetical order)

ALCTS Acquisitions Section (AS)
The Foreign Book Dealers Directories Series Subcommittee of the Publications Committee is working with the ALCTS office to develop the Foreign Book Dealers Directories, an Online Vendor Database for Africa/Middle East, Asia/Pacific, and Eastern Europe & the Former Soviet Union. It is ready for Web publication.

ALCTS Cataloging & Classification Section (CCS)
Subject Analysis Committee (SAC) sponsored a program "**SALSA de Topicos = Subjects in SALSA : Spanish And Latin-American Subject Access"** held in 2004 Annual Conference in Orlando, Florida. About 250 or more attendees

The program provided a dynamic look at the subject analysis and classification for Spanish-speaking library patrons, with an emphasis on the fluid boundary between the "domestic" and the "international". Many libraries have built valuable collections of Spanish-language materials, but catalog access is hindered by English-only subject headings. The program featured initiatives to provide culturally appropriate access, and will stimulate thought about what librarians can do to better serve their patrons.

SAC plans a publication (target date: June 2006) in both English and Spanish in the ALCTS Paper Series based on speakers' presentations.

ALCTS Collection Management & Development Section (CMDS) -- None

ALCTS Council of Regional Groups (CRG) -- None

ALCTS Preservation & Reformatting Section (PARS) -- None

ALCTS Serials Section (SS)
No action regarding non-English access or non-Roman scripts, except discussion and input on the CC:DA Agenbroad proposal mentioned in the CC:DA report.

The Serials Section Committee to Study Serials Standards has lists of several standards available for serials librarians. Relevant standards listed below:

Acquisitions Standards:

Codes for the Representation of Languages for Information Interchange (ANSI/NISO Z39.53-2001) Defines almost four hundred separate three-character alphabetic codes and adds codes for twenty-eight languages or language groups previously not represented. This edition includes changes necessary to be compatible with bibliographic language codes in ISO 639-2 (Codes for the representation of names of languages : alpha-3 code). Available in print: Bethesda, Md.: NISO Press.

Cataloging Standards:

ALA-LC Romanization Tables, 1997 ed. Transliteration schemes for non-Roman scripts. Developed and maintained by Library of Congress. Available in print: Washington, D.C.: Cataloging Distribution Service, Library of Congress. Available online: lcweb.loc.gov/catdir/cpso/roman.html

ALCTS/LITA Authority Control in the Online Environment Interest Group (ACIG)
2001 Annual Conference with theme "By What Authority? How Multiple Sources of Authority Are Changing Our Concepts of Control." Barbara Tillett discussed using the Web to link fields in authority records, to create a Virtual International Authority File, and to enable users to switch to their preferred languages of display.

2002 Annual Conference with ACIG theme "Real World Steps to Interoperability between Electronic Resources." Glenn Patton from OCLC gave a presentation on "International Efforts to Improve Interoperability." In his presentation, he mentioned MACS (Multilingual Access to Subjects) which is a OCLC Project with 4 national libraries (BL, BnF, DDB, SNL)

2003 Annual Conference: "Authority Control and Unicode -- Together at Last." Ann Della Porta and Joan Aliprand presented papers.

2005 Midwinter, speakers described various developments in the field. Ann Della Porta's LC report covered LC Unicode progress and romanization issues; Glenn Patton's OCLC report mentioned upcoming access to additional authority files, including HKCAN (Hong Kong Chinese Authority Name).

2005 Annual Conference with ACIG theme "XML and authority control." Louisa Kwok (HKUST Library) described HKUST's development of a name access control file in XML, following the model of the Virtual International Authority File (VIAF). Records representing three levels—the person, the names used by the person, and the forms of each of those names—correspond roughly to the FRBR work-expression-manifestation hierarchy. The file can link different name forms together without imposing the use of a particular preferred form across all files.

Also Joanna Yi-hang Pong (City University of Hong Kong) discussed the HKCAN project, which has constructed a centralized database of XML authorities to link between headings in different catalogs, languages, and scripts for seven Hong Kong academic libraries. HKCAN enables users to begin a search in one file and transform the search terms to match the authorized form in other files.

ALCTS/CCS Cataloging Norm Discussion Group
At 2001 Midwinter, the Discussion Group had a presentation about changing to Pinyin romanization.

Other ALCTS Interest Groups / Discussion Groups – None

Appendix F: Library of Congress

Library of Congress and Access to Non-English Resources

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.

Non-English Latin Script Resources

LC phased in MARC records for non-English Latin script resources as the MARC systems were implemented at LC in the late 1960s and early 1970s. The production of non-English Latin script cards was discontinued in the 1980s, when all card production at LC ceased.

Cataloging Policy and Processes

The Library of Congress has materials in over 530 languages and currently catalogs in about 120 languages, 64 of which are non-Latin script languages. LC is currently creating records for all Latin script resources in its ILS. It is also still creating records for JACKPHY language materials using RLG for monographs, and JACK language materials on OCLC for serials. There will be changes to these processes in the next few years as LC now has a Unicode version of the Voyager system installed that allows cataloging and retrieval in non-Latin scripts. 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.

Integrated Library System

In November 2005 with the upgrade of its integrated library management system LC completed the conversion of its MARC bibliographic, holdings and authority records to UTF-8. Users of the LC Online Catalog (catalog.loc.gov) and LC Authorities (authorities.loc.gov) can search, display, and download records in all languages and scripts represented in the LC Database. Extensive help files are available to instruct users in setting their browsers to enable search and display of these languages and scripts. In January 2005 LC's Cataloging Distribution Service began offering the option of receiving records in the MARC 8 character set or the Unicode (UTF8) character set to customers of its MARC Distribution Service.

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

Vendor Records

LC has obtained records from its book vendors around the world for many years. If they are in Latin script they are loaded into the Library's ILS and used as the basis for cataloging. Non-Latin script vendor records are loaded into RLG in several cases. With the new Unicode ILS capability LC may consider loading the non-Latin script vendor records into the LC system along with the Latin script ones.

6/16/06– Sally McCallum

Appendix G: Program for Cooperative Cataloging (PCC)

Documents from the PCC showing their activities in the area of non-Latin scripts and international cooperation.

CONSER Editing Guide Appendix E: Creating Records with Data in Non-Roman Script for Arabic Serials (2003)
http://www.loc.gov/acq/conser/appendixE.pdf

Formulario de Información para Participantes del Programa NACO
http://www.loc.gov/catdir/pcc/naco/nacoappl_spa.html

Guidelines for BIBCO Libraries Contributing Bibliographic Records for Chinese Language Materials (2000)
http://www.loc.gov/catdir/pcc/bibco/pinyinrules.html

Introduction to the Program for Cooperative Cataloging: BIBCO Core Record Standards (2005) [includes Supplementary Core for Multiple Character Sets]
http://www.loc.gov/catdir/pcc/bibco/coreintro.html

Manual para participantes del Programa Cooperativo de Autoridades de Materias (SACO)
http://www.loc.gov/catdir/pcc/saco/spmspa.pdf

Non-Roman Core Record Task Group: Final Report
http://www.loc.gov/catdir/pcc/archive/jackphy.html

Non-Latin Scripts: Additional Requirements for Core Records (1996) (Replaced Nov. 1999 with Supplementary core for multiple character sets)
http://www.loc.gov/catdir/pcc/bibco/corejack.html

PCC and Expanded Use of Non-Latin Scripts (2006)
www.loc.gov/catdir/pcc/bibco/Reser_BIBCO_06.pdf

The Program for Cooperative Cataloging: Mission, Goals, and Potential for International Cooperation (1996)
http://www.loc.gov/catdir/pcc/archive/tabbpaper.html

Supplementary Core For Multiple Character Sets (approved November 1999)
http://www.loc.gov/catdir/pcc/bibco/nonromancore.html

Appendix H: OCLC

OCLC and Access to Non-English Resources – A Brief Report

Bibliographic records for materials in languages other than English have been available in OCLC's WorldCat from the very beginning. Subject headings in languages other than English have also been included in bibliographic records from very early days because of the National Library of Canada's participation in the CONSER Program. Non-English subject headings have expanded over the years as libraries in other part of the world have begun to use OCLC services. Appropriate changes have also been made to insure that non-English headings from batchloaded records are retained. Since subject indexing of WorldCat records was added (in 1990 for reference purposes and in 1993 for technical services), all subject headings, regardless of language, have been available for searching.

Statistics: Over the past several years, the percentage of records in WorldCat that represent titles in languages other than English has risen to almost 40%. The top 10 languages other than English at the end of 2005 are: German, French, Spanish, Japanese, Chinese, Russian, Italian, Portuguese, Finnish and Latin. More statistics are available at <http://www.oclc.org/worldcat/statistics/language.htm>.

OCLC and Access to Non-Roman Scripts – A Brief Report

Technical Services Uses: OCLC first provided the capability to create bibliographic records using Chinese, Japanese and Korean scripts in 1986. In 2000, OCLC introduced a stand-alone, Windows-based application for Arabic script cataloging. Over the past two years, OCLC has introduced new cataloging interfaces and a new Unicode-based database. Part of that process included, in March 2005, moving the separate CJK and Arabic interfaces into the Connexion client so that one interface covers all scripts. In July 2005, Cyrillic, Greek and Hebrew scripts were added to the client. Searching includes full keyword searching using almost 100 keyword indexes. This completes OCLC's implementation of those non-Roman scripts that can be represented using the MARC-8 character set. Future plans call for the expansion of non-Roman script capabilities based on users' requests. In February 2006, Tamil and Thai were added and there are plans to add additional South Asian scripts during 2006. In addition to the scripts supported for record creation, Connexion client interfaces are available in Chinese (simplified or traditional), Japanese and Korean as well as English, German and Spanish.

Also in July 2005, OCLC began to provide access to the Chinese Name Authority File. The file is created and maintained by the institutions of the Joint University Librarians Advisory Committee (JULAC) of Hong Kong. Initially, the file contained about 120,000 records and grows at the rate of 8,000 to 10,000 records a year. Records are, where possible, based on records from the LC/NACO file but have been enhanced to include references in Chinese characters. The records follow the MARC 21 Authority Multiscript Records Model B.

Public Services Uses: Display of non-Roman script data was introduced for OCLC FirstSearch WorldCat in July 2002. Limited indexing was also made available at that time. As of November 2005, full indexing of non-Roman data (using the same indexes available to technical services users) was added to FirstSearch WorldCat. In addition to the scripts supported for record display, FirstSearch interfaces are available in Arabic, Chinese (simplified or traditional), Japanese and Korean as well as English, French and Spanish. Open WorldCat interfaces are available in Dutch, English, French, German and Spanish.

Statistics: In Spring 2006, there were 2.95 million records in WorldCat containing CJK characters, 86,000 containing Arabic script, 49,500 containing Hebrew script, 5,500 containing Cyrillic script, 175 containing Greek script, 22 containing Thai script and 7 containing Tamil script.

Prepared by Glenn Patton

Appendix I: RLG

RLG Development of Non-Latin Script Support

♦ RLG implemented the "CJK" enhancements to RLIN in 1983. RLG unified from existing national character sets at that time all the Chinese "Han" characters used in the People's Republic of China, Taiwan, Japan, and Korea, plus *kana* and *hangul,* into one character set: RLIN East Asian Character Code (REACC). Its three-byte structure was based on the Chinese Character Code for Information Interchange (CCCII), making it easy to normalize variant forms of the same Chinese character for indexing and retrieval. (Thus a user could search for Japanese names using PRC simplified forms, or a Chinese classical work using Japanese *joyo kanji* forms, and retrieve records regardless of the form of a character used.) REACC became the American National Standard, East Asian Character Code set for bibliographic processing (EACC) in 1987 (Z39.64), and formed the basis of the unified Han character repertoire in Unicode and the international Universal Multiple-Octet Coded Character Set (UCS) standard (ISO/IEC: 10646).

LC rejected the first RLG white paper on adding CJK scripts to bibliographic records, which mimicked the MARC format by displaying all the "alternate graphic representation" fields for the scripts in the same repeated 880 fields used for communication between systems. Instead, the RLIN application presented "parallel fields" in displays and input, with the script field immediately following the romanized equivalent. Internally each original script field was linked to its romanized equivalent and exported as an 880; a CJK-script field without a romanized equivalent (e.g., for content notes) was supported by using a "00" occurrence number in the $6 non-Roman linking subfield. The paired romanized and non-Roman script fields were used in all other scripts implemented by RLG, and also used by other system implementations. The original RLG white paper also presumed that script support would be added to authority and bibliographic formats at the same time, but this was not possible given that all NACO contributor nodes had to support scripts at the same time. Instead, users were encouraged to add parallel access points with scripts to the bibliographic record to allow users who were not familiar with ALA/LC romanization schemes the ability to retrieve authors, corporate bodies, and subjects with the scripts the material were written in.

♦ RLG implemented the Cyrillic script enhancements to RLIN in 1985, supporting Russian and other Slavic languages written in Cyrillic scripts: Belarusian, Bulgarian, Macedonian, Serbian, and Ukrainian. Most North American libraries did not take advantage of this script enhancement, since most users could identify and select the titles represented by romanization only. It's used primarily in records from Cyrillic-script resources: the Russian Academy of Science Bibliographies and the Hand Press Book database (the records from the National Library of Russia).

♦ RLG implemented Hebrew script enhancements to RLIN in 1988, supporting Hebrew, Yiddish, Ladino, and Judeo-Arabic languages. This was the first right-to-left

script implementation, and introduced the /r MARC subfield to indicate a default change in directionality. The RLG implementation preceded the Unicode bidirectionality algorithms for mixing right-to-left and left-to-right strings within a single field.

♦ RLG implemented Arabic script enhancements to RLIN in 1991, supporting Arabic, Farsi (Persian), Ottoman Turkish, and Urdu languages. The same algorithms for mixing right-to-left and left-to-right strings used for Hebrew were applied. The Arabic script implementation introduced changing the graphic used for an Arabic character depending on the letter's position within a string (initial, middle, final, or isolated) and generating the appropriate diagraphs where needed.

♦ From 1988, RLG staff were active in the Unicode Working Group, and RLG was one of the founders of the Unicode Consortium incorporated in 1991. RLG has been a Full Member and represented on the Unicode Board of Directors ever since. RLG was active in the MARBI Character Set Subcommittee to provide round-trip mappings between the MARC-8 character repertoires and Unicode.

♦ In 1991 RLG submitted a proposal to MARBI to add alternate graphic representations to the authority format (91-1). This was approved, but implementation was deferred until there was a consensus on *how* the community wished to represent non-Latin script references in the authority record.

♦ In 2000 RLG implemented "on-the-fly" conversion of the MARC-8 encodings of the Arabic, Chinese, Cyrillic, Hebrew, Japanese, and Korean scripts in bibliographic records into Unicode in its Eureka® search interface for end-users. Users could then search and display bibliographic records with these scripts with a Unicode-compliant Web browser. This marked the end of requiring special fonts and library-specific applications to provide access to the original script materials cataloged since 1983.

♦ In 2005 RLG migrated its databases to Unicode. Catalogers create and edit bibliographic records that include Arabic, Chinese, Cyrillic, Hebrew, Japanese, or Korean scripts in the RLIN21™ client with Microsoft Input Editors for character entry. Unicode formatting control codes were introduced to enter left-to-right multi-digit strings in an otherwise right-to-left field.

Current State:

The RLG Union Catalog contains over 3 million titles that include Arabic, Chinese, Cyrillic, Hebrew, Japanese, or Korean scripts. RLG currently dataloads records received in UTF-8 encoding, and will be offering UTF-8 export as an option in 2006. (For Z39.50 client users, this would require a new ExportCharacterSet parameter to retrieve records in UTF-8 rather than MARC-8 encoding.) RLG is participating in conference calls with LC, the British Library, and OCLC on the next steps needed to add non-Latin character sets into authority records. RLG plans to add support for character set repertoires outside the MARC-8 repertoire in 2006-2007. (Under
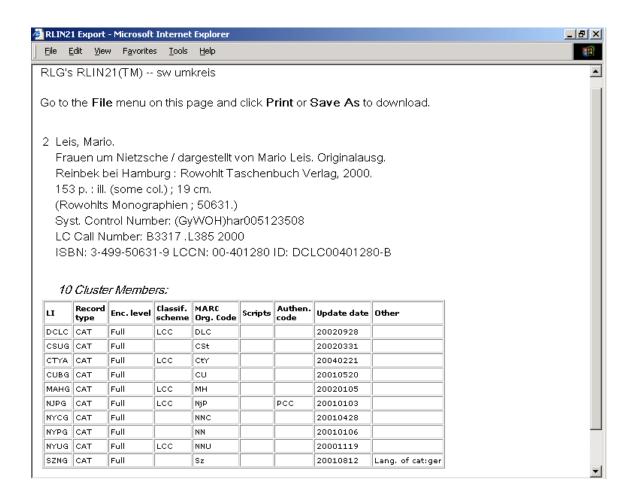
consideration: Armenian, already implemented by the British Library, and southeast Asian and south Asian scripts).

RLG Access to Non-English Resources

RLG batchloads into the RLG Union Catalog records that represent non-English cataloging, primarily from its increasingly global membership, but also book vendor records and other non-English sources. We also batchload the bilingual French-English records from the Library and Archives of Canada. The RLG Union Catalog is a "clustered" database: records representing the same title are matched based on the *description* of the item, which should be the same regardless of where the item was transcribed, even if the access points differ. Catalogers have repeatedly told us that non-English cataloging requires editing, especially the access points, and thus records that represent non-English cataloging do not appear first in record displays if there is another record for the same title with English-language cataloging.

All records in an RLG Union Catalog contribute their respective access points. Thus a search with a Dutch, French, German, Japanese, Italian, or Spanish subject word can retrieve clusters that show the Library of Congress or other English-language catalog records but also includes records from sources such as the International Institute of Social History, the Bibliothèque nationale de France, Bibliothèque municipale de Lyon, the Swiss National Library, Keio University, Toshokan Ryutsu Center, Istituto Centrale de Catalogo Unico, the Biblioteca Nacional de Espana, etc. This provides multilingual access at the same time that catalogers can choose an English-language cataloged record instead, when available.

The language of cataloging is displayed in the RLIN21 cluster grid in displays, and is exported as a 952 subfield l in response to Z39.50 client queries, represented by the value of the 040 $b in the record. An example is shown below (retrieved by a German subject word contributed by the Swiss National Library record, the one flagged with a "Lang. of cat: ger" in the "Other" column). This mechanism has served users well for twenty years. The multilingual access is available through Eureka, RLIN21, Z39.50 client queries, and RedLightGreen.

RLG's RLIN21(TM) -- sw umkreis

Go to the **File** menu on this page and click **Print** or **Save As** to download.

2 Leis, Mario.
   Frauen um Nietzsche / dargestellt von Mario Leis. Originalausg.
   Reinbek bei Hamburg : Rowohlt Taschenbuch Verlag, 2000.
   153 p. : ill. (some col.) ; 19 cm.
   (Rowohlts Monographien ; 50631.)
   Syst. Control Number: (GyWOH)har005123508
   LC Call Number: B3317 .L385 2000
   ISBN: 3-499-50631-9 LCCN: 00-401280 ID: DCLC00401280-B

   10 Cluster Members:

LI	Record type	Enc. level	Classif. scheme	MARC Org. Code	Scripts	Authen. code	Update date	Other
DCLC	CAT	Full	LCC	DLC			20020928	
CSUG	CAT	Full		CSt			20020331	
CTYA	CAT	Full	LCC	CtY			20040221	
CUBG	CAT	Full		CU			20010520	
MAHG	CAT	Full	LCC	MH			20020105	
NJPG	CAT	Full	LCC	NjP		PCC	20010103	
NYCG	CAT	Full		NNC			20010428	
NYPG	CAT	Full		NN			20010106	
NYUG	CAT	Full	LCC	NNU			20001119	
SZNG	CAT	Full		Sz			20010812	Lang. of cat:ger

Appendix J: ILS vendors/local systems

SURVEY ON NON-ENGLISH SUPPORT IN INTEGRATED LIBRARY SYSTEMS

SUMMARY OF SURVEY RESULTS

Date of compilation: March 23, 2006
Compiled by: Joan M. Aliprand

Questionnaires sent to 20 companies
7 responses received (30.5%)

Caution

Using Unicode gives the *potentiaI* to implement any of its scripts, but actual support for a script requires software and fonts for input, indexing, and presentation of the script. In the absence of detailed information about a particular ILS, it would be wise to take statements on script support such as:

      All scripts supported by Unicode.
      All of the above and any scripts that are in Unicode.
      … able to accommodate any script whose characters have codes defined by the
         Unicode Standard.
to mean:
   those scripts supported in the software and fonts used in the ILS.

RESPONSES

1.   Scripts available on ILS
1.1   Are any non-Roman scripts supported?

Yes: 5 (including one reply "All scripts supported by Unicode")
In development: 1 "full Unicode language support 2006 will be available in 2006"
No: 1

1.2   If yes, which of these [the scripts supported by individual MARC 21 character sets] are supported?

All: 4
Potentially all: 1 (system in development)
Only Cyrillic and Greek: 1

1.3   Which other scripts are supported?

All: 3, as well as 1 system under development.
One of these respondents provided details:
      "All other Unicode-based, Microsoft-supported scripts."
The others did not: (see *Caution* above):

Display only, cataloging input conforms to MARC 21 limitation to scripts also supported in MARC 21 individual character sets ("MARC-8"): 1
"Central European:" 1

[Note: Central European languages are written in either Latin script (Czech, Hungarian, etc.) or Cyrillic script (Ukrainian, Byelorussian, etc.). – jma]

2.   Textual display
2.1 Have users reported any problems with the display of text?

No/none known: 5
Yes: 2

2.2   If yes, please give details.

"Users do not always have a Unicode-compliant font on their workstations or browsers. That is the most common difficulty reported. They generally don't understand that the display depends on the capabilities of the [brand name] software, the workstation operating system itself, and the fonts installed for the application."
[This problem is not likely to be limited just to the reporting brand. – jma]

"Text display problems are connected to locally installed fonts – they are not connected to [the ILS] software."
[Fonts are not the only reason for text display problems (as reported in the preceding response). – jma]

3.   Language support at user interface
3.1   Does the user have a choice of language (as some ATMs provide) for dialog with the system?

Yes: 4
Possibly Yes: 1 "A multilingual user interface is available."
For Help: 1 "In the OPAC, the end user can toggle from English to another provided language without losing the context of the search. Help can be available in the language the user has selected."
No response: 1 [From the company that does not have support for non-Roman scripts]

3.2   If yes, what languages are available?

Reported number of languages varied from many ("up to 115") to a few.
"up to 115 supported, libraries generally select 2 or 3, most common Spanish, Vietnamese, and Korean"
34 supported (21 written with Latin script, 13 with another script)
12 "and more" (of those reported, 8 written with Latin script, 4 with another script)
5 in production (English, Spanish, Japanese, Maori, Samoan); 3 others developed.
4 (English, French, Dutch, German), "others can be added."3 (French, Spanish, Italian)

Note: Of the languages mentioned above, all except Korean and Japanese are written with Latin script. – jma

3.3   Who provides the dialog for a particular language?

Variety of methods: Customers, with company as backup; Third party
translating service; Library-designated language specialist; Our distributors; Our local
offices and distributors.
One respondent replied "The client program." No response: 1

3.4   Are there any problems with the orthography of any language?

No/none known/"Not for the supported languages:" 4
Yes: 1 (see below)
No response: 1
No usable response: 1 Respondent replied that the system "fully supports the Unicode
standard and non-Roman character sets." and went on to describe input methodology
incorrectly. "The IME (Input Method Editor) is a Microsoft Windows tool that provides
the cataloger with the ability to enter thousands of different characters used in non-
Roman languages without requiring special keyboards."

3.5   If yes, which language(s)?

"Korean display did cause difficulties in the first iteration of support, but those issues
have been resolved."

4.   Subject access in languages as well as English
4.1   Is subject access in more than one language available?

Yes: 4
Probably equivalent to Yes: 1 "No, not other than is available in the MARC records. If
the site uses Spanish language headings, for instance, those would be indexed and
available for retrieval and display, but not separately from the English-language
headings."
Yes with limitation: 1 "The end user can perform keyword, title, author, etc. searches in
any of the available languages. Subject browsing is available in English only."
No: 1

4.2   If yes, which languages?

All: 5 (including response: "This is up to the library.")
Not applicable (because of initial reply): 2

4.3   Are the subject access points controlled via an authority file?

Yes: 5
Probably equivalent to Yes: 1 "This is up to the library."
Not applicable: 1

4.4   If yes, are there language-specific authority files, or only a combined file?

Multiple subject thesaurus authority files: 1
Combined file: 4 (one respondent noted ", but you can extract only the form for one specific language if needed').
Either: 1 "This is up to the library."
Not applicable: 1

5.   Authority control
5.1   Is text in a non-Roman script allowed in an authority record?

Yes: 5
In development: 1
No: 1

5.2   If yes, is the non-Roman data restricted to particular fields?

No: 4
Yes: 1 "As MARC-compliant system, we don't allow non-Roman script in authority records other than in the defined 880 (alternate graphic) fields."
No response: 1
No usable response: 1 "Yes - full Unicode language support 2006 will be available in 2006."

LIST OF COMPANIES RESPONDING TO SURVEY AND THEIR SYSTEMS

EOS International (EOS.Web)
Ex Libris (ALEPH 500)
Geac (Vubis Smart)
Library Concepts (PC Card Catalog)
The Library Corporation (CARL•X)
Polaris Library Systems (Polaris)
VTLS Inc. (Virtua)

Appendix K: Authority control vendors

Survey on Non-English Support By Authority Control Vendors

Results of Survey

Companies that responded: Marcive; Backstage Library Works; Library Technologies, Inc. (of seven companies)

1. Does your company support non-Roman scripts in its authority control processing?

Yes – 2, with one saying it "has been processing Unicode/UTF-8 records for over a year"

No -- 1

One of the yes responses, also said,
"In the authority control process, the heading text of non-Roman headings is not converted due to the lack of standard vernacular. The links between the vernacular headings (880 fields) and Romanized heading fields are maintained during authority control processing.

If romanized heading fields are deleted or added (for example, due to a split) the occurrence number in the $6 linkage subfield is modified appropriately. Similarly, the linking tag portion of the $6 linkage subfield is modified to reflect any tag conversions in the romanized heading fields. A report is generated listing all 880 fields which do not link to a romanized heading field due to incorrect data in the $6 linkage subfield. All other 880 field data is untouched during processing.

Diacritics: Often the only difference between two CJK personal name headings is the presence or absence of a diacritic. A special CJK normalization routine is used for CJK bibliographic records. The programs assume a bibliographic record may contain CJK romanized headings if at least 1 of 3 criteria is present:

1) Specific values in a 066 Character Set Presence field that indicate use of a CJK graphic character set ($1 in a subfield $a or $) in a subfield $b)

2) An 880 Alternate Graphic representation field

3) Language code chi,jpn, or kor in the 008 bytes 35-37

When a bibliographic record is identified as a probable CJK record, the following diacritics are retained in the normalized headings; Alif, Ayn, Breve, Circumflex, Hyphen, Macron, Umlaut. The Preservation of these diacritics prevents matching a normalized CJK personal name heading to the incorrect LC authority record."

2. If yes, which of these are supported?

Arabic (used to write Arabic, Persian, Urdu, etc.)
Chinese
Japanese (including hiragana and katakana)
Korean (including hangul)
Cyrillic (used to write Russian, etc.)
Greek
Hebrew (used to write Hebrew, Yiddish and Ladino)

3 answers:
*Non-Roman scripts occurring in records for authorities processing will be
returned as supplied.
* Japanese, Korean and Chinese partially supported
* N/A.

Are you planning to support others in your processing?
    Thai
    Tamil
    Devanagari (used to write Hindi and other Indic languages)

Others (please specify)

No response from vendors.

3. Have users reported any problems with the display of authority and/or
bibliographic records in non-Roman scripts after authority control processing? If yes,
please explain.

2 – no problems

1 replied: Our reports are in HTML format and do display date in Unicode. However,
many PCs have trouble displaying CJK character. In some instances, our CJK vernacular
(880) representations are not translated into readable characters.

4. Do you perform any manual processing of authority headings?

3 – do some manual processing

If so, do you have language experts for manual processing in non-Roman
languages?

2 – no language experts

1 – yes, language experts. Most European languages, most Arabic, Hebrew, Cyrillic languages, Asian languages. All other languages we hire outside language experts.

Which language(s)?

5. Are you in compliance with the Unicode Standard (version 4.0 or high) for your authority control processing?

3 different responses

1. Company "continues with development to keep pace with libraries' changing cataloging needs."
2. Didn't really answer this question, but said "Unicode compliance will be under consideration in the re-write for our authority control processing."

3. Our process normalizes headings in order to match against national database. Normalization requires the removal of extraneous spaces, punctuation and most diacritics. Also, the downloadable Library of Congress authority database is in MARC-8 not UTF-8. However, we do have the ability to convert bibliographic records from UTF-8 to MARC-8 and visa versa.

If not, do you have plans to work on compliance?

2 -- yes

Appendix L: Reports from other groups (in alphabetical order)

Association for the Advancement of Slavic Studies (AAASS) Bibliography and Documentation Committee

"The AAASS Bibliography & Documentation (B & D) Committee, consisting of scholars and librarians in the fields of Slavic, East European, and Eurasian studies, works throughout the year on a wide range of topics and ventures that include--but are not limited to--copyright, digital projects, microfilming projects, vendor issues, collection development, and the American Bibliography of Slavic and East European Studies (ABSEES)"--Committee web site.

Oct. 1990, annual conference: roundtable "Cataloging of Slavic Materials in an Online Environment."

Nov. 1992, annual conference: panel "Databases and Electronic Communications with the CIS and Eastern Europe."

Nov. 1994, annual conference: program "Fully Searchable, Full-Text Databases in Cyrillic and Other Non-Roman Alphabets."

Nov. 2004, annual conference: panel " Slavic Digital Projects: Creation, Encoding, Delivery."

Nov. 2005, annual conference: panel "Digitization: Practical Issues for Scholars and Librarians."

ACRL Rare Books and Manuscripts Section (RBMS) Bibliographic Standards Committee

This committee's work sometimes touches on issues of non-English access. For example,

May 2005: A draft revision of DCRM(B) contains expanded guidance for materials written in non-Roman scripts. Draft: http://www.folger.edu/bsc/dcrb/DCRMBpublicreviewdraft20050531.html. The final revised version should be published by the end of the year.

ACRL Slavic and East European Section (SEES)

"Represents librarians and specialists in the fields of Slavic and East European studies and is concerned with those aspects of library service which require knowledge of Slavic and East European languages"--ALA Handbook of Organization, 2005-2006.

June 2004, ALA annual conference: Library catalogs and non-Roman scripts: development and implementation of Unicode for cataloging and public access (conference program co-sponsored by ALCTS CC:AAM and LITA)

SEES Automated Bibliographic Control Committee (ABC)

1991-1992: ABC conducted a survey of ACRL libraries on remote access to Slavic materials in their catalogs; results were distributed to survey participants.

June 1998, ALA annual conference: a report and handouts to SEES members on computer support for Cyrillic text.

2000-2002: discussion and response to LC's decision to distinguish between Serbian, Croatian, and Bosnian. ABC objected to changing the meaning of existing language codes (e.g., scr can now mean either Croatian for newer records or Serbian (roman) for legacy data) rather than creating new codes and felt that this impedes access to materials in these languages. ABC also requested guidelines to distinguish these languages. Letters were sent to LC. LC did not create new language codes, but provided some (outdated) guidelines for distinguishing the languages.

2001-2002: discussion on proposing new Cyrillic abbreviations for AACR2 App. B and a related survey of users. Though the sample of users understood the abbreviations in the context of description, they were less successful in actively producing the abbreviations, implying more abbreviations could impede access through keyword searching. The proposal was not put forward.

June 2001: ABC took over the maintenance of the Slavic Cataloging Manual website, originally authored by Jim Weinheimer of Princeton (http://www.indiana.edu/~libslav/slavcatman/).

Jan. 2003, ALA midwinter conference: discussion of concerns with ILS vendors' ability to handle diacritics (e.g., some diacritics prevent keyword searching in Innovative Web OPACs). No action taken by ABC, but individual librarians pursued the problem with their own vendors.

Jan. 2004, ALA midwinter: a proposed project with CC:AAM to normalize/improve transliteration of Central Asian languages.

Jan. 2006, ALA midwinter: a letter to LC/PCC asking for more specific guidelines for cataloging in vernacular script is being written. This is motivated by the fact that CJK and HAPI communities have different practices, and now that Cyrillic is available on OCLC, Slavic catalogers have noticed these inconsistencies and the need for more guidance across the different languages.

ACRL West European Studies Section (WESS)

WESS "represents librarians and others who specialize or are otherwise professionally involved in the acquisition, organization, and use of information sources originating in or related to Western European countries." Many of the resources WESS is dealing with are in non-English languages. Recent example activities include:

Jan. 2005, ALA midwinter: Scandinavian Discussion Group talked about Harvard's project to download of National Library of Finland MARC records to provide access to Finnish materials.

Jan. 2006, ALA midwinter: Romance Languages Discussion Group discussed vendor records for materials in Spanish, French, and Italian, exploring librarians' expectations regarding cataloging records, and how vendors may best meet users' needs.

WESS Cataloging Discussion Group: Issues of concern that have been discussed in recent years include: 1) difficulty in recruiting catalogers with good non-English language skills; 2) the rise of outsourcing, which seems to go along with less than full records and a possible degradation of subject access to non-English materials; 3) the need for non-sorting control characters in more MARC fields to handle inflected languages that use articles; and 4) concern with the gap between where we are today and full Unicode implementation. An example of the latter is the German character "esszett," which LC has not yet implemented, yet is available for use in some systems.

Association of Jewish Libraries

North American libraries have created online catalog records of Hebrew script materials that contain both Latin and Hebrew scripts (and other scripts as well) since 1988. These records are created in the national utilities (RLG Union Catalog) OCLC WorldCat*), and institutions' local systems. Records input locally are uploaded into the national utilities. Records created in the national utilities may be passed into local systems depending on the availability of non-Roman scripts in a particular system. Catalogers transcribe bibliographic data into Hebrew script as it appears on the piece in hand. This includes the title statement (245 field), edition statement (250 field), imprint data (260 field) and series statement (440 field). These are fields defined by AACR2rev. 1.0E as requiring exact transcription whenever possible. Catalogers also have the option of adding vernacular script to added entries, notes, subject headings, and series tracings. These vernacular fields appear mostly as linked fields to the romanized data that are required by current MARC conventions.

Libraries cataloging Hebraica materials in their local systems and the national utilities have varying standards for the amount of vernacular script and how much romanized data they provide. Some libraries only provide full or partial Hebrew data for the 245, 250, 250 and 4xx fields. Others also include varying amounts of Hebrew script data for main and/or added entries, notes and subject headings.

Some institutions do not provide roman script data for the statement of responsibility or beyond the title proper. In 1990, the Cataloging Sub-Committee of the Jewish and Middle East Studies Program (JAMES) of the Research Libraries Group (RLG) drafted a proposal to LC regarding this issue. LC's follow-up proposal appeared in the *Cataloging Service Bulletin* (Summer 1990), p.45. The proposal was never implemented.

In the summer of 1993, the Association of Jewish Libraries presented a panel of four speakers on current practices and standards of cataloging Hebraica in RLIN. The resultant papers appeared in v. 8 of *Judaica Librarianship*.

Current practices for adding Hebrew script access points to bibliographic records were looked at in a brief survey in Spring 2005.** There do not currently exist any standards

or guidelines for authority control for Hebrew and Yiddish names in Hebrew script appearing in MARC 21 bibliographic records. Names in Hebrew script may appear in many different forms as access points in bibliographic records in the national utilities and local catalogs. Most libraries that do create bibliographic records with Hebrew script provide some sort of main and added entry in Hebrew script, in addition to the romanized form. Some, but not all libraries do perform some sort of authority work for the Hebrew script forms of the heading but a great deal of inconsistency exists. At least one North American library does provide Hebrew script cross-references in its local authority file. The LC/NACO Authority File at the present time is restricted to Latin Script.

*OCLC, in Summer 2005, made Hebrew script available in WorldCat.
** The results of this survey were given at the 2005 Annual Conference of the Association Jewish Libraries in a presentation on anticipating the use of Hebrew script in the LC/NACO Authority File. The paper that grew out of that presentation will appear in LRTS (50:2).

German-North American Resources Partnership (GNARP) Bibliographic Control WG

This is a partnership of German and North American Libraries associated with ARL, which is reporting to WESS on its activities. "The Bibliographic Control Working Group has been charged to assess and evaluate the current state of German-language cataloging in North America, to develop strategies to maximize German-language cataloging expertise on a national level through knowledge and resource-sharing, and to establish closer ties to the cataloging community in Germany. The goal of the latter is to work toward bringing our respective cataloging cultures closer together, and to harmonize our various rules and practices to allow for more effective sharing of bibliographic data."

GNARP will be conducting a survey in spring 2006 of monographic catalogers at large libraries (including ARL libraries) to determine how libraries make use of European records, concentrating on vendor records and records for German materials. The results will be published when completed.

Middle East Librarians Association
Activities in Cataloging and Cataloging Policy Formulation

ABOUT MELA: The Middle East Librarians Association (MELA) was established November 5, 1972 by 17 librarians attending the sixth annual Middle East Studies Association (MESA) meeting in Binghamton, New York. "*It is the purpose of the Middle East Librarians' Association to facilitate communication among members through meetings and publications; to improve the quality of area librarianship through the development of standards for the profession and education of Middle East library specialists; to compile and disseminate information concerning Middle East libraries and collections and to represent the judgment of the members in matters affecting them; to encourage cooperation among members and Middle East libraries, especially in the acquisition of materials and the development of bibliographic control; to cooperate with other library and area organizations in projects of mutual concern and benefit; to*

promote research in and development of indexing and automated techniques as applied to Middle East materials."

Members: 102 as of November 2005
Publication: *MELA Notes* (ISSN 0364-2410)
Website: http://www.mela.us

Since its establishment, MELA has worked with ALA and has had at least one liaison reporting on activities of interest to its membership. MELA members have represented organization formally to review *AACR*, and have represented Middle East cataloging interests on the ALCTS committee, Cataloging and Classification: Asian and African Materials (CC:AAM).

What follows is a general overview of some of the activities, events and discussions that have taken place within MELA regarding cataloging and cataloging policy issues over the years. MELA members continue to play an important consulting role in the standards and practices for cataloging of Middle East materials.

1975: The workshop "Cooperation Among Middle East Libraries of North America," held in Ann Arbor, MI from May 26-31, 1975, and includes a task force on Cataloging/Automation. Recommendation is made for a National Union catalog of Near Eastern materials.

1976: Discussion of draft rules for LC romanization tables for Ottoman Turkish.

1981: Cataloging and Technical Processing Committee established to work with MELA's ALA liaison.

1982: MELA Panel on COMRAD (Committee of Machine Readable Arabic Data) is held; members include the head of the Near East Section of Library of Congress and eight librarians from major research libraries. An earlier working paper by the Committee written in 1978 surveyed then-existing practices for handling non-Roman scripts. Contextual information is given for practices concerning South and East Asian scripts.

1985: Collection of variant vocalization of names, publishers and phrases from MELA libraries; Library of Congress starts their *Preferred Vocalization* lists, and also provides a list of resources used to determine the preferred choice.
Involvement in CC:AAM activities for name changes for political jurisdictions involving territorial changes (i.e. "Ottoman Empire")

1987: "Romanizing Arabic script catalog records" published in *MELA Notes*. This outlines Ohio State's project to transfer Arabic and Hebrew cards to romanized script records in OCLC, whose records were then transferred to Ohio's local OPAC.

1988: Demonstration of Arabic-script searching on a Mac-SE is done by a MELA member at the annual meeting.
Publication of the first volume of the *Near East National Union List*.

1990: Survey of existing software for Arabic and Hebrew-script cataloging published in *MELA Notes*.
Analysis of Middle Eastern materials in RLIN is published in *MELA Notes*.

1991: "Arabic personal names headings" is published in *MELA Notes*. These are the guidelines used at Princeton University that will eventually form the core guidelines for the Arabic NACO Funnel project.

1995: A cataloging committee is established.

1997: Arabic NACO Funnel formally established and based at Princeton. A NACO workshop on Arabic headings is held during MELA's annual meeting.

2000: LC solicits requests for assistance on KBP and revision of the BP classification schedules.
Prototype of Arabic Cataloging manual set up at Cornell. A major topic of focus already is the policy for the inclusion and standardization of non-Roman fields in bibliographic records. The problems introduced by non-standard romanization of Arabic and Persian words continue unresolved.
Harvard begins discussions with the leading suppliers of Arabic materials from Cairo and Beirut to provide acquisition records that can be uploaded in MARC format.

2003: The problem of the current lack of a useable Kurdish romanization scheme is discussed; the CC:AAM representative reports that LC has a draft table under consideration.
The problem of the Turkish word "bir" and its role as a definite or indefinite article is discussed. The CC:AAM representative has submitted a proposal included in the CC:AAM report.

2004: The materials created for the Arabic cataloging manual is now on the WWW at http://infoshare1.princeton.edu/katmandu/acm/acmmain.html.
MELA reviews the Kurdish romanization table; sends letter of support and recommendation to CC:AAM.

2005: Kurdish romanization table approved by ALA/LC.
Committee on Cataloging is formally re-activated [charge and accomplishments/agenda to date are attached below].

Committee on Cataloging: 2005 Report

MELA has a newly-reconstituted Committee on Cataloging, whose charge is to identify and consider issues, make recommendations and work with other appropriate bodies regarding issues related to the cataloging and classification of Middle East materials in libraries. Its members include:

- Joyce Bell (Princeton)
- Muhammad al-Faruque (University of Illinois, Urbana-Champaign)
- Meryle Gaston (UC-Santa Barbara) (Chair)
- William Kopycki (University of Pennsylvania)
- Fereshteh Molavi (Yale)
- Sarah Ozturk (Library of Congress)

The committee's first task during 2005 was to draft, distribute and tabulate the results of a survey of cataloging practices among MELA members and their respective institutions. Information on this has been posted on the Committee's website at:
http://library.princeton.edu/departments/tsd/acm/ConC/conc.html

The Committee met November 19 to discuss its agenda that included:

- Persian romanization issues. The committee is discussing the creation of a Persian Cataloging Issues Subcommittee; this subcommittee will address issues concerning improved bibliographic access to these materials.
- Islamic law classification: LC is soliciting suggestions for expanding the schedule for KBP (Islamic Law); LC also reports that it is considering the development of a BPI schedule for Islam, expanding its current place in BP.
- SACO funnel: A possible pilot project was discussed to have a SACO funnel that would complement the existing NACO Arabic Funnel project in PCC. One of the important areas that such a funnel could play a role in is the contribution of subject headings specific for Islamic law. A memo on this topic was sent to MELA membership to gauge interest.
- Non-roman script fields: The Committee intends to compile "best practices" for libraries creating bibliographic records using Arabic-script fields. Institutions with Unicode-compliant OPACs will be queried to determine public-service issues and practices in this regard.
- A Middle East "Cataloger's Desktop/Reference Shelf": One of the by-products of the Cataloging Survey previously mentioned was the compilation of the bibliography of authoritative reference sources that respondents found useful to aid in cataloging materials in the languages of the Middle East and surrounding region. These sources, grouped by language, have been posted at: **http://library.princeton.edu/departments/tsd/acm/ConC/conc.html**.

Public Library Association Cataloging Needs for Public Libraries Committee
by Vivian Bordeaux, Chair

The Cataloging Needs for Public Libraries Committee as a group has not addressed the complex issues surrounding non-English access. I am sure that the Committee members are faced with this problem and remain very frustrated with purchasing and cataloging these items.

I will briefly discuss how two Public Libraries, Bridgeport Public Library, Bridgeport Connecticut and King County, Seattle, Washington handle non-English materials.

In communities with heavy immigrant populations it is a challenge to provide non-English materials for this diverse population. We have found that many vendors have difficulties in obtaining foreign language materials and libraries are forced to look to other sources. Kings County (Seattle, Washington) collects world language materials in all formats and has a collection of at least 19 foreign languages.
Once the materials are acquired the catalogers have the job of finding records in the databases like the Library of Congress, OCLC and their local area. Some databases lag behind in providing English subtitles for books and audiovisual materials. While the majority of the records are found, the ones that are not found present problems. If there is no one on staff who speaks the language the cataloger creates the best record possible with the information on hand. I have noticed at the Bridgeport Public Library (Bridgeport, Connecticut) that we tend to get materials in Spanish from only Mexico and Spain, and that French Creole materials are unavailable.

Slavic Librarians Summer Workshops
(located at Univ. of Illinois, organized by various Slavic librarians)

May 1989, Conference on Access to Slavic Materials in North America: Retrospective conversion of card catalogs (a presentation and discussion of difficulties in converting cards printed in Cyrillic and other non-Roman scripts and the need to transliterate that which was not transliterated before).

June 2002, Cataloging Workshop: a presentation on diacritics and transliteration of non-Roman scripts.

July 2004, Slavic Digital Text Workshop: Creating Cyrillic texts and OCR, and Delivering Cyrillic texts and XSLT (workshop sessions); Tips for Slavic/Cyrillic computing (web page created, http://www.wisc.edu/creeca/reeca/online/UICU_presentation.html).

July 2005, Slavic Digital Text Workshop: several sessions on encoding and delivering texts in Slavic languages (presentations at http://www.library.uiuc.edu/spx/slw-dtw/Presentations05.htm).

Bibliography of Unicode™ in the Library Environment
Compiled by David Nelson and Joan Aliprand
May 2006
(All URLs accessed on June 16, 2006)

PRIMARY SOURCES

The Unicode Standard : http://www.unicode.org/standard/standard.html.
   Includes links to the specification of the current version (4.1) of the Unicode Standard, as well as to the code charts and the PDF version of the published book (Version 4.0).

The Unicode® Consortium. http://www.unicode.org

LC CATALOGING POLICY STATEMENT

Library of Congress. Cataloging Policy and Support Office. "Unicode Implementation at the Library of Congress: Cataloging Policy Position." Latest revision: March 16, 2006.
http://www.loc.gov/catdir/cpso/unicode.pdf
        Note that LC states "This document will be revised as the situation warrants."
        The document ends with a list of issues still to be addressed.

MARC

Library of Congress. Network Development and MARC Standards Office. "MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media."
http://www.loc.gov/marc/specifications/

"Assessment of Options for Handling Full Unicode in Character Encodings in MARC 21."
Part 1: New Scripts. A Study for the Library of Congress, by Jack Cain. (2004)
http://www.loc.gov/marc/marbi/2004/2004-report01.pdf.
Part 2: Issues. (2005)." http://www.loc.gov/marc/marbi/2005/2005-report01.pdf.

Part 1: MARBI Discussion; Comments from OCLC and RLG

MARBI Meeting Minutes (June 2004):
http://www.loc.gov/marc/marbi/minutes/an-04.html

Comments from OCLC and RLG in MARC Forum archives, June 2004
http://listserv.loc.gov/listarch/marc.html

        11. OCLC Comments on MARBI
        Papers, from: Richard Greene, Date: Tue, 22 Jun 2004

        20. RLG Comments on MARBI Unicode
        Report, from: Joe Altimus , Date: Tue, 22 Jun 2004.

Part 2: MARBI Discussion
Representatives of OCLC and RLG were directly involved in the development of Part 2.

MARBI Meeting Minutes (June 2005):
http://www.loc.gov/marc/marbi/minutes/an-05.html

Archives of UNICODE-MARC Discussion List. http://listserv.loc.gov/listarch/unicode-marc.html

OVERVIEWS OF THE UNICODE STANDARD

Desgraupes, Bernard. *Passeport pour Unicode*. Paris: Vuibert informatique, 2005. ISBN 2-7117-4827-8.

Gillam, Rich. *Unicode Demystified*. Boston, MA: Addison-Wesley, 2002. ISBN 0-2017-0052-2

Graham, Tony. *Unicode: a Primer*. Foster City, CA: M&T Books, 2002. ISBN 0-7645-4625-2.

Korpela, Jukka. *Unicode Explained*. O'Reilly Media, [forthcoming]. ISBN 0-596-10121-X.
> Projected publication date: June 2006.

RELATED TOPICS—TERMINOLOGY, CHARACTER SETS, ISSUES

Aliprand, Joan. "Issues to Consider in Multilingual, Multiscript World." Presented 4 October 2002. Slides used at presentation, and video of presentation: http://hkcan.ln.edu.hk/opening/index.htm

Aliprand, Joan. "The Structure and Content of MARC 21 Records in the Unicode Environment." *Information Technology and Libraries* 24, no. 4 (Dec, 2005): 170-179.

Bausenbach, Ardie. "Character Sets and Character Encoding: A Brief Introduction." *RLG DigiNews* 8, (April 15, 2004). http://www.rlg.org/en/page.php?Page_ID=17068#article2.

Breeding, Marshall. "Integrated Library Software: A Guide to Multiuser, Multifunction Systems." *Library Technology Reports* 40, iss. 1 (Jan.-Feb. 2004).

Coyle, Karen. "Unicode: The Universal Character Set." *Journal of Academic Librarianship* 31, no. 6 (Nov, 2005): 590-592.

Gippert, Jost. "Language-Specific Encoding in Multilingual Corpora: Requirements." (1999). http://titus.fkidg1.uni-frankfurt.de/personal/jg/pdf/jg1999e.pdf

Jilovsky, Cathie and Andrew Cunningham. "Multilingual Cataloguing : Practical and Technical Issues." Slides used at presentation: www.ifla-stockholm2005.se/pdf/postconfdoc%20IFLA-swe05%20pdf/Stockholm%20presentation.pdf

Korpela, Jukka. "A Tutorial on Character Code Issues." http://www.cs.tut.fi/~jkorpela/chars.html

Mates, Barbara T. "Computer Technologies to Aid Special Audiences." *Library Technology Reports* 40, iss. 3 (May-Jun. 2004).
    "Special-audiences" include patrons for whom English is a second language.

Script Encoding Initiative. http://www.linguistics.berkeley.edu/~dwanders/

Starr, Joan. "Design Considerations for Multilingual Web Sites." *Information Technology and Libraries* 24, no. 3 (Sep, 2005): 107-116.

Wells, J. C. "Orthographic Diacritics and Multilingual Computing." *Language Problems and Language Planning* 24, no. 3 (2000): 249-272.

Wen, Danyang, Tetsuo Sakaguchi, Shigeo Sugimoto, and Koichi Tabata. "Multilingual Access to Dublin Core Metadata of the ULIS Library." *Journal of Digital Information* 2, no. 2 (Nov 2001, 2001).

Wood, Alan. "Unicode and Multilingual Support in HTML, Fonts, Web Browsers and Other Applications" http://www.alanwood.net/unicode/

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

Arsenault, Clément. "Retrieval Issues for Chinese-Language Library Resources." *International Information & Library Review* 37, no. 2 (Jun, 2005): 107-116.

Chetana, M., and P. Sumangala Devi. "Information Retrieval Strategies in Indian Languages: A Guideline." *Information Studies* 11, no. 4 (Oct 2005, 2005): 251-264.

Chou, Charlene. "The Forest and the Trees: HKCAN beyond CJK Cataloging." Presented 4 October 2002. Slides used at presentation, and video of presentation: http://hkcan.ln.edu.hk/opening/index.htm

Gee, Quintin. "Review of Script Displays of African Languages by Current Software." *New Review of Hypermedia & Multimedia* 11, no. 2 (2005): 247-255.

Ibrahim, Ahmed Elhafiz. "Displays of Arabic Script on Web-Based OPACs in GCC Institutions." *Electronic Library* 23, no. 4 (Aug 2005, 2005): 419-432.

Jacobs, Jane W., Ed Summers, and Elizabeth Ankersen. "Cyril: Expanding the Horizons of MARC 21." *Library Hi Tech* 22, no. 1 (2004): 8-17.

Jacobs, Jane W. and Malabika Das. "Making the Cyrillic OPAC a Reality." *Slavic & East European Information Resources* 6, no. 2/3 (2005): 135-149.

Khurshid, Zahiruddin. "A Survey of the Arabian Gulf Library Automation Marketplace." *Electronic Library and Information Systems* 37, no. 4 (2003): 226-233.

Lunde, Ken. *CJKV Information Processing*. Beijing, Sebastopol, CA: O'Reilly, 1999. ISBN 1-56592-224-7.

OCLC. "International Cataloging: Use Non-Latin Scripts." http://www.oclc.org/support/documentation/connexion/client/international/#/

PAN Localization Project. *Survey of Language Computing in Asia 2005*, by Sarmad Hussain, Nadir Durrani, Sana Gul. Lahore: Center for Research in Urdu Language Processing; Ottawa: International Development Research Center, 2005. ISBN 969-8961-00-3.
      Available as a book and also in PDF from: http://www.idrc.ca/uploads/user-S/11446781751Survey.pdf

"Seminar on Using Unicode UTF-8 for Online Catalogue on Innopac System" [held] July 11, 2003 (Updated August 4, 2003). Slides used at seminar, some papers, and video of presentations: http://www.lib.cuhk.edu.hk/seminar/unicode/
      Includes "Overview of Chinese Character Encoding," by Ki Tat LAM.

TITUS (Thesaurus Indogermanischer Text- und Sprachmaterialien) http://titus.uni-frankfurt.de/indexe.htm.

Urs, Shalini R., N. S. Harinarayana, and Mallinath Kumbar. "A Multilingual Multi-Script Database of Indian Theses: Implementation of Unicode at Vidyanidhi." *In* International Conference on Asian Digital Libraries (5th : 2002 : Singapore). *Digital Libraries: People, Knowledge and Technology, Proceedings.* Berlin; New York: Springer, 2002. (*Lecture notes in computer science*). ISBN 3540002618.

Yu, Abraham J. "Technical Services: Cataloging and Authority Control of Chinese Language Materials." Slides used at presentation: http://sun3.lib.uci.edu/~ajyu/LucePitt/Luce-summer-institute-04-yu.ppt

UNICODE AND LIBRARIES

ALCTS Cataloging and Classification Section's (CCS) Committee on Cataloging: Asian and African Materials (CCAAM), ACRL Slavic and East European Section (SEES) and LITA. "Library Catalogs and Non-Roman Scripts: Development and Implementation of Unicode for Cataloging and Public Access," ALA Annual Conference, June 2004. http://www.ala.org/ala/alcts/alctsconted/presentations/presentations.htm
      Presentations (Slides used by speakers):

True Scripts in Library Catalogs - The Way Forward
Joan Aliprand, RLG

Unicode Encoding and Online Data Access
Dr. Jost Gippert & Ralf Gehrke, Universitat Frankfurt, Vergleichende
Sprachwissenschaft

Libraries, Patrons, and the Web: A Transformation of Knowledge Management,
Discovery and Delivery
Gary Houk, Corporate Information Technology, OCLC

Library Catalogs and Non-Roman Scripts
Michael Kaplan, Ex Libris Liaison

Multi-script Decisions for Cataloging Policy at the Library of Congress
Dr. Barbara Tillett, Cataloging Policy & Support Office, Library of Congress

Barry, Randall. "Challenges of the Web to Libraries with Multilingual and Multiscript
Collections." http://www.w3.org/2002/02/01-i18n-workshop/Barry.html
　　　Position paper presented at W3C Internationalization Workshop (2002 : Washington,
　　　DC).

Jilovsky, Cathie. "Unicode: A Tool for System Interoperability and Human
Communication." http://www.vala.org.au/vala2004/2004pdfs/45Jilov.PDF

"The Multicultural Library: Staff Competence for Success." Satellite conference
organized by: IFLA Section on Library Services to Multicultural Populations, Swedish
Library Association, [and] The International Library, Stockholm.
http://www.ifla-stockholm2005.se/
　　　Slides from 14 presentations on this site

Shaker, Ali Kamal. "Bibliographic Access to Non-Roman Scripts in Library OPACs: A
Study of Selected ARL Academic Libraries in the United States." Ph.D., University of
Pittsburgh, 2002. Abstract: http://etd.library.pitt.edu/ETD/available/etd-12122002-
015433

Tull, Laura. "Library Systems and Unicode: A Review of the Current State of
Development." *Information Technology and Libraries* 21, no. 4 (Dec, 2002): 181-185.

MULTILINGUAL/MULTISCRIPT AUTHORITY FILES

Aliprand, Joan. "Scripts, Languages, and Authority Control." *Library Resources &
Technical Services* 49, no. 4 (Oct, 2005): 243-249.

International Conference on Authority Control: Definition and International Experiences.
Papers. (Florence, Italy: 2003). http://www.sba.unifi.it/ac/en/intro_eng.htm

Papers presented at this conference are also available through E-LIS: E-prints in Library and Information Science http://eprints.rclis.org/
Topics include multilingual authority control and East Asian names.

Khairy, Iman. "Authority Control of Arabic Personal Names from the Classical Period at the Bibliotheca Alexandrina." http://web.uni-bamberg.de/unibib/melcom/khairy_eng.pdf
Paper presented at the MELCOM International Conference (27th : 2005 : Alexandria, Egypt).

Lerner, Heidi. "Anticipating the Use of Hebrew Script in the LC/NACO Authority File." *Library Resources & Technical Services* 50, no. 3 (July, 2006) [in press].

Tillett, Barbara. "A Virtual International Authority File." Slides used at presentation: http://www.iccu.sbn.it/upload/documenti/TillettAF.ppt
Presented at Catalogazione e controllo di autorità: Giornate di studio. (2002 : Istituto della Enciclopedia Italiana : Roma). All presentations available at http://www.iccu.sbn.it/genera.jsp?id=96

Tillett, Barbara. "A Virtual International Authority File." http://www.nii.ac.jp/publications/CJK-WS3/cjk3-08a.pdf

Vassie, Roderic. "Improving Access in Bilingual, Biscript Catalogues through Arabised Authority Control." *Online Information Review* 24, (2000): 420-428.

Weinberg, Bella Hass, and Joan Aliprand. "Closing the Circle: Automated Authority Control and the Multiscript YIVO Catalog." *International Cataloguing and Bibliographic Control* 31, (2002): 44-48.