–

Arabic Name Authority in the Online Environment: Options and Implications

Martha Speirs Plettner

Arabic name authority records have been handwritten using Arabic script and filed manually in book or card catalogs since the time that it was considered important to preserve this information. After the adoption of typewriters as tools in library cataloging departments, those who only had Latin script typewriters were forced into using transliteration schemes, a practice that has been criticized for compromising uniformity and accessibility.(Houissa) Later, typewriters fitted with Arabic character keys allowed authority cards to be typed in Arabic. There were also attempts to encode both Latin and Arabic scripts on cards—or in book catalogs, as the first dual-script name authorities; something that was encouraged by the catalog cards distributed between 1902 and 1997 by the Library of Congress.
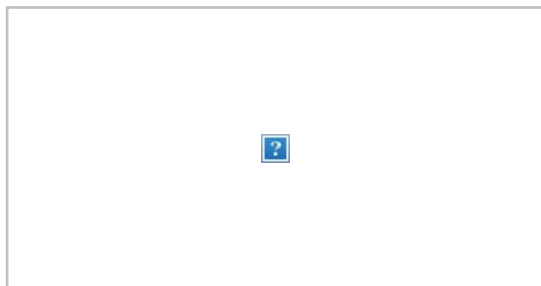
This locally composed authority card has used the ALA/Library of Congress transliteration along with the authorized vernacular version of the heading and its references.

(Figure 1) *From the AUC Library authority card file*

The methods of recording name authority information were dependent, to a large extent, on the capabilities of the available technology. The styles varied according to the formats that were defined for this cataloging. Formats and styles developed to hold this important information include those defined by AACR2, Names of Persons, Guidelines for Authority and Reference Entries (GARE), ISBDs, MARC etc. There were also many locally defined formats and style guidelines which served their purpose in an isolated bibliographic environment, but which would not work in today's world which has embraced the concept of universal bibliographic control. Shared centralized bibliographic databases will sideline those who do not follow some sort of international standardization.

The dilemma over whether the dual-script cards should be filed according to the Arabic script alphabetical sequence or the transliterated one has not ended with the passing of card catalog files. This is still a question which has yet to be solved in a way that best meets all users' access needs. (Casaubon) Should there be two different card catalogs, the traditional solution adopted in most libraries, or two different online filings of authorized name authorities--one in each script? Two separate files hamper comprehensive access as it requires two searches in order to be confident of finding all of an author's works. IFLA's UNIMARC guidelines prefer the specific vernacular of the author as the basis for authoritative headings, even though they have only been minimally incorporated in authority records and the Library of Congress currently uses only Romanized authorized headings. In the future, with the potential aid of machine intelligence, we should not have to be faced with a choice of only one type of script for retrieval purposes, however, most current online catalog structures require that there be one dominant type of authorized heading that collocates all of an author's works. This paper addresses the formats of the records within the authority file itself as opposed to focusing on the content of the fields within the authority record or the structure of the files, even though it is difficult to fully examine one part of the total authority structure in isolation.

 Since the advent of online bibliographic catalogs, name authority records have traditionally been encoded using Latin script transliteration. This is because when ASCII was the only available character set there was no choice but to use its Latin script characters in electronic records. The development of many different Romanization schemes made the confusion even worse. Realization that the use of the original script for searching would be more likely to lead to the one valid name heading rather than to many other misleading hits, along with technological advances in computing, led to work on the development of non-Roman scripts for use in the online catalog by the 1980s.

Several groups in the United States were instrumental in the implementation of non-Roman scripts in bibliographic records and have consequently established the formats and models which have been followed by the users of the large bibliographic utilities such as RLIN and OCLC. The Research Libraries Group added the capability to encode Arabic scripts in its bibliographic utility called the Research Libraries Information Network (RLIN) in 1991. Since then, over 100,000 bibliographic records containing this non-Roman script have been added to RLIN. Authority records in RLIN however continue to be Romanized. OCLC is a nonprofit membership organization serving 41,000 libraries in 82 countries and territories around the world. Their WorldCat database now has 35,311 bibliographic records which contain Arabic script, but authority records are still entered with Roman transliteration. The United States Name Authority File (USNAF) which is controlled by the Cooperative Cataloging Program at the Library of Congress uses the MUMS system which does not yet allow for the creation or display of non-Roman scripts.

Catalogers became more aware of the need to adhere to standards with the advent of online catalogs and the specification by

bibliographic utilities in their regions of formats such as MARC and styles and guidelines such as those defined by AACR2. In order to use RLIN or OCLC one is required to follow their specific guidelines for cataloging Arabic. These requirements – much like the availability of Arabic typewriters – have encouraged the adoption and spread of certain cataloging methods and styles. Arabic language catalogers located throughout the world now use the manuals on Arabic cataloging required by these large utilities. (RLIN, OCLC)

The addition of extensions to MARC formats in order to accommodate non-Roman scripts in bibliographic records in the 1980s led to the development of a bibliographic catalog record which had the capacity to include non-Roman scripts. Of course, special character sets such as ISO 11822 or Arabic MS1256 were required for Arabic script implementation. After 1993, the availability of the Unicode universal character set made inclusion of all scripts coded by the Unicode consortium possible as long as library systems were also able to manipulate this character set and the bi-directional algorithms for languages read from the right-to-left such as Arabic. (Unicode) The movement towards using vernacular scripts in MARC records has progressed since the 1980s but as yet this capacity has been almost exclusively implemented in bibliographic records.

Some research has been done on the specific issue of implementation of original Arabic headings in name authority records by Aliprand, Eilts, Vassie, Vernon and others. They have all written about problems and potential solutions in implementing dual Arabic and Latin script in the authority record, however there are not many institutions devoting staff-time to this labor-intensive practice yet, so there is little significant experience in this sort of dual-script authority cataloging. Bernard has done research into the existing Arabic language online catalogs in libraries in France and the Maghreb which concludes that, for the most part, the surveyed libraries have followed different practices with some using a mixture of original and transliterated scripts in the bibliographic record, but only a few having experimented with using vernacular script in the online authority record. (Bernard) There are many libraries in the Arab world that do all of their cataloging in Arabic without using any transliteration, but since many of these databases are not using MARC formats which are eligible for seamless sharing with other established authority databases they cannot seamlessly share records with OCLC or RLIN even though they may share with other online catalogs which have the same capabilities. There is clearly a need for sharing any experiences that can be gained from name authority file applications using non-Roman scripts in other areas of the world.

Since the advent of Unicode as an international standard in 1993, there has been discussion of how the structure of authority records could be enhanced to allow for the incorporation of original non-Roman scripts such as Arabic. There is no question that when a name is not written in its original script, a key such as a transliteration table is more necessary to understand the graphic symbols. Since less than half of the world uses Latin scripts exclusively, the inclusion of non-Roman scripts in any future plans is essential. The challenge now is to implement this encoding capability in a uniformly standardized way and to encourage its adoption by all libraries.

The following discussion focuses specifically on the MARC21 format and US groups and only touches on the theoretical advances and efforts made by other groups. Further research will benefit from a more international look at all authority formats and sharing schemes presently in use. This paper should only be regarded as a beginning.

MARBI (the American Library Association joint committee responsible for the development and maintenance of the MARC formats) has added capabilities for the addition of alternate characters in the MARC21 authority record. NACO (the National Name Authority Cooperative) has been involved in establishing Arabic names in LC authority records in the NAF for many years but they have not yet set up or implemented any specific rules for incorporating vernacular scripts in their name authority records. The decision to begin adding Arabic script name headings to the United States Name Authority File (USNAF) will be possible once the Library of Congress upgrades their system software to Unicode compliancy.

From an American perspective, the USNAF is currently the uncontested authority on name authority entries. This file is a result of cooperation from all of the NACO contributors who have compiled millions of names in a consistent manner. The fact that the British Library Name Authority File (BLNAF) has decided to merge files with the USNAF to make up the Anglo-American Authority File (AAAF) will reduce duplication and make it even more international. However, since this merged group still only represents an Anglo-American part of the world and is removed from the non-Roman script world there is need for a supra-national group. It would be ideal if there could be a global consensus on name authority--something that is much more feasible today with our global connectivity and Z39.50 protocol. IFLA could be a very important hub in the design of names donated by national bibliographic agencies. The work by Beaudiquez and Bourdon is a good beginning towards this effort. (Beaudiquez) Of course, the remaining dilemma of conflicting tensions between respecting the country of origin's preferences and meeting the needs of conventional usage in one's own country by the users of one's own catalog needs to be dealt with before any meaningful cooperation can take place.

It is imperative that the authority record formats and styles that are developed be compatible with the computer systems that are designed to handle them. This is why RLG, OCLC and many integrated library system vendors have representatives at the ALA Machine-Readable Bibliographic Information (MARBI) discussions. Any incorporation of new models has implications for the syndetic structure and indexing of the databases using them. Two models (A & B) of multi-lingual and/or multi-script authority records have been proposed

in Concise USMARC21 formats for the incorporation of original scripts in authority records. In 2001 a new discussion paper (2001-DP05) proposed a new recommended model (Model C) which uses the concept of a context marker in an adaptable record. This MARBI discussion paper, "Multilingual Authority Records in the MARC21 Authority Format" proposes solutions not yet approved for use, but offers an interesting potential solution for global sharing of authority records with the option of harvesting and loading only those fields needed for a specific context.(MARBI) The tags for the fields used here in these MARC records are 100 for authorized heading, 400 for *see from* references, 880 for alternative script data, 700 for alternate context data and 670 for notes.

```
001     n 79058331
003   DLC
008     790706n|acannaab|     |a aaa    cz n
100  10 Moses|c(Biblical leader)
400  00 Musá|c(Biblical leader)
400  10 Mosheh|c(Biblical leader)
400  00 Moise|c(Biblical leader)
880  10 |6100-02(3/r|aمـوسى|c(Biblical leader)
```

(Figure 2) *Model A—MARC21 multiscript model*

The MARC21 Concise Authority format's Model A is an example of an Arabic name authority record with the vernacular in the 880 fields. This model mirrors the handling of alternate graphic representation in MARC21 bibliographic records. However, Model A will become more and more complicated as more scripts are added to the 880 fields, and as it is not necessary to have matching Roman and non-Roman fields, the direct use of the 400 field for alternate non-Roman scripts is a simpler solution. This solution is displayed in Model B below.

```
001     n  50014331
003     DLC
005     19950123115723.1
008     800410n| acannaab|a aaa ||| cz n
010     n  50014331 |zn  89119817
040     DLC|cDLC|dDLC
053     PJ7864.A35
100  10 Husayn, Taha,|d1889-1973
400  00 Taha Hussein,|d1889-1973
400  10 Hussein, Taha,|d1889-1973
400  00 |wnna|aTaha Husayn,|d1889-1973
400  10 Husain, Taha,|d1889-1973
400  10 Huseyn, Taha,|d1889-1973
400  10 طه ,حسـين,|d1889-1973
670     His Kudama ibn Jafar, Abu al-Faraj, al-Katib
        al-Baghdadi,|bNakd an-Nathr ... 1933.
```

(Figure 3) *MARBI's proposed Model B*

MARBI's Model C or the context-marker model is the linked record model of Model B, but the entries are defined by a catalog context which is a composite entity. The primary component of the context is the body of rules under which the heading was formulated (e.g., AACR2, LCSH, RAK, RAMEAU, etc.). Additional components could be added; these might include an explicit indication of the language of catalog into which the heading fits (e.g., hun, eng, ger) or the audience for which the heading is useful (e.g. children, popular). Each heading below is appropriate for the context designated at the record level by a combination of elements which have yet to be defined.

Model C

```
001     ea55555
008/10 (Cataloging rules): c (AACR2)
008/11 (Subject system/thesaurus rules): a (LCSH)
040     $b (Language of cataloging): eng (English)
100 0# $a Cleopatra, $c Queen of Egypt, $d d. 30 B.C.
400 0# $a Cleopatre, $c Queen of Egypt, $d d. 30 B.C.
400 0# $a Kilyubatra, $c Queen of Egypt, $d d. 30 B.C.
400 0# $a Kleopatra, $c Queen of Egypt, $d d. 30 B.C.
400 0# $a Kliyubatra, $c Queen of Egypt, $d d. 30 B.C.
700 04 $a Cleopatre $b VII $c (reine d'Egypte ; $d 0069-0030 av. J.-C.)
        $? <context: French language catalog> $0 fa44444
700 04 $a Kleopatra, $c Agypten, Konigin, $b VII $d av69-v30
        $?<context: German language catalog> $0 ga33333
700 04 $a $? كليوبـترة <context: Arabic language catalog> $0 aa66666
```

(Figure 4) *Model C from MARBI paper(2001-DP05)*

This record, as it stands, is intended for an English speaking audience, accustomed to using AACR2 and the Library of Congress Subject Headings as its thesaurus. If this record were needed for an Arabic-speaking audience, the Arabic 700 heading could be flipped to become the 100 field (authorized heading) in the local catalog record. Name authority records should no longer be thought of as static sets of information, but should be seen as sources from which information can be drawn as needed for the specific context of one's online catalog. UCBIM's Working Group on Minimal Level Authority Records (MLAR) recommends the increased use of linked authority records, but considering today's available technological capabilities it seems that specific locally defined headings could be harvested from a master source record which are appropriate for the context in which they will be used. (Tillett) Danskin remarks that, "Volatility is

one of the characteristics distinguishing authority data from bibliographic records" and clearly this property of volatility adds to the need for flexibility demanded by the local context option. (Danskin) In the context of a bi-lingual audience the library may want to keep the proposed 700 fields in the harvested record for display reference only. For instance, students who are studying literature in a non-Roman script but may not feel comfortable searching using the vernacular script for searching may still want to view the script version of the author's name. It should not be necessary to have the specific language or script used in searching determine the language or script shown in the display. Flexibility should exist in the locally desired options chosen for a particular use in a local OPAC.

In order to further the implementation of these concepts, collaboration is necessary among groups like MLAR, the Foreign Marc Coalition (FMC), a group composed of members from the OCLC, RLIN and the Library of Congress which strives to provide access and distribution of foreign MARC records in the US, international groups such as IFLA, The Functional Requirements and Numbering of Authority Records Working Group (FRANAR), Linking and Exploring Authority Files (LEAF) and others in order to develop a centralized, comprehensive database of name authority records which can be used to derive locally acceptable records depending on language and other contexts of the local catalog.

Some other potential ideas for further development of international name authority sharing are:

- The implementation of an International Standard Authority Data Number (ISADN) which has been proposed as a tool in facilitating the linkage and processing of MARC records. (Willer, Bourdon2002)

- The provision for easy exchange of MARC formats. Crosswalks between formats such as UNIMARC and MARC21 should be transparent.

- Barnhart's idea of access control records where the user can select a default display form--a step in the right direction. (Barnhart)

- Sharing is essential--The shared database carries with it standards, policies and procedures which must be clear.

- Ease of use --An easy-to-use database with automated transliteration and parallel vernacular fields will further enable the effort to share standardized name authority records.

Therefore, I feel that a potential model composed of an international name authority clearinghouse which allows input of authorized headings from national agencies and customized downloading of chosen fields into context-specific databases is an obvious global solution. Meanwhile, a specific solution for Arabic name authority headings could also be developed, but with provision for its eventual incorporation into a larger global, multilingual network.

As the world becomes smaller and more virtual and technological advances make international sharing increasingly possible, a centralized clearinghouse of multi-lingual and multi-script name authority records will be the most seamless way of reducing duplication of local efforts and costs which come with this duplication. Barnhart sees the idea of a "super authority record" or an access control record as "brazenly radical" but we clearly need to look ahead to a radical shift in our present structures in order to achieve a more seamless and user-friendly result. This achievement must allow for local options in the composition of records for appropriate local contexts, while maintaining valid international exchange standards. The encoding of specific Arabic descriptive content within the appropriate fields in name authority records must be dealt with as an integral but separate issue falling within a shared global name authority file structure with the resulting consequence of comprehensive international sharing of vernacular Arabic name authority records.

A major implication of this radical shift to context-specific record structures will be the savings in time and effort earned by both the cataloger and the user. Authority work as we know it will change as the bulk of the effort will become focused on providing appropriate access and display for the context of the local audience.

End Notes

Aliprand, Joan M. *The Unicode Standard: An Overview with Emphasis on Bidirectionality*. Multi-script, Multilingual and Multi-character Issues for the Online Environment. John D. Byrum Jr. and Olivia Madison, Eds. Munchen: K.G. Saur, 1998. 95-112.

Barnhart, Linda. *Access Control Records: Prospects and Challenges*. 1996. Authority Control in the 21st Century: An Invitational Conference. OCLC. Sept. 2002. Available at: <http://www.oclc.org/oclc/man/authconf/barnhart.htm>.

Beaudiquez, Marcelle and Francoise Bourdon. Management and Use of Name Authority Files. Munchen: Saur, 1991.

Bernard, Annick. Etude sur les catalogues Arabes existants: dans le contexte plus général des projets de conversion retrospective de catalogues de bibliothèques du Maghreb. Version 3.2 Unpublished paper. 2002.

Bourdon, Francoise. *Functional Requirements and Numbering of Authority Records (FRANAR): To What Extent Can Authority Control Be Supported by Technical Means.* International Cataloguing and Bibliographic Control. 31.1 (Jan/Mar 2002): 6-9.

Bourdon, Francoise. *How Can IFLA Contribute to Solving Problems in Name Authority Control at the International Level?* IFLA Journal. 18.2 (1992): 135-137.

nd

Danskin, Alan. International Standards in Authority Data Control: Costs and Benefits. Conference Proceedings, 62   IFLA General Conference. 1996. Available at: <http://www.ifla.org/IV/ifla62/62-dana.htm>.

Eilts, John. *Non-Roman Script Materials in North American Libraries: Automation and International Exchange*. International Cataloging and Bibliographic Control. 25.3 (July/September 1996): 51-53.

Houissa, Ali. *Arabic Personal Names: Their Components and Rendering in Catalog Entries*. Cataloging and Classification Quarterly. 13.2 (1991): 3-22.

Jaudenes Casaubon, María and Nuria Torres Santo Domingo. *A Mediterranean Perspective: Spain's Biblioteca Nacional.* Automated Systems for Access to Multilingual and Multiscript Library Materials. Munchen: K.G. Saur, 1996. 45-53.

MARBI Discussion Paper, *Multilingual Authority Records in the MARC21 Authority Format*. 2001-DP05 <http://lcweb.loc.gov/marc/marbi/2001/2001-dp05.html>.

MARC21 Authorities. Sept. 2002 <http://www.loc.gov/marc/authority/ecadhome.html>.

Murtomaa, Eeva and Eugenie Greig. *Problems and Prospects of Linking Various Single-Language and/or Multi-Language Name Authority Files*. International Cataloguing and Bibliographic Control. 23.3 (July/September 1994): 55-58.

Names of Persons: National Usages for Entry in Catalogues. 3rd Ed. IFLA International Office for UBC. London: 1980.

OCLC Arabic--Quick Reference. OCLC. Sept. 2002. <http://www.oclc.org/oclc/arabic/quickreference/>.

RLIN Cataloguing Guide. RLG. Sept. 2002. <http://www.rlg.org/catguide/catguide.html>.

Tillett, Barbara B. *Authority Control at the International Level.* Library Resources and Technical Services. 44.3 (July 2000): 168-72.

Unicode Character Set. Sept. 2002 <http://www.unicode.org>.

Unimarc/Authorities. <http://www.ifla.org/VI/3/p2001/guideindex.htm>.

Willer, Mirna. *Authority Control and International Standard Authority Data Numbers: Need for International Cooperation*. 1996. Authority Control in the 21st Century: An Invitational Conference. Sept. 2002 <http://www.oclc.org/oclc/man/authconf/willer.htm#4>.

Vassie, Roderic. *Improving Access in Bilingual, Biscript Catalogues through Arabised Authority Control.* Online Information Review. 24.6 (2000): 420-428.

Vassie, Roderic. *A Reflection of Reality ---Authority Control of Muslim Personal Names.* International Cataloguing and Bibliographic Control. 19.1 (January/March 1990): 3-6.

Vernon, Elizabeth. Decision-Making for Automation: Hebrew and Arabic Script Materials in the Automated Library. Champaign-Urbana: University of Illinois, 1996.

[an error occurred while processing this directive]