UCSF

UC San Francisco Previously Published Works

Title

Authority Control of Arabic Personal Names: RDA and Beyond

Permalink

https://escholarship.org/uc/item/54j4f70d

Journal

Cataloging & Classification Quarterly, 59(2-3)

ISSN

0163-9374

Authors

Dagher, Iman

Soufi, Denise

Publication Date

2021-04-03

DOI

10.1080/01639374.2020.1845896

Peer reviewed

Authority Control of Arabic Personal Names: RDA and Beyond

Iman Dagher[a]* and Denise Soufi[b]

*[a]Library Resource Acquisitions & Metadata Services, University of California, Los Angeles,
California, USA. ORCID ID 0000-0002-1793-8578*

*[b]Resource Description & Management, University of North Carolina at Chapel Hill, Chapel
Hill, North Carolina, USA. ORCID ID 0000-0002-6394-0476*

Iman Dagher, idagher@library.ucla.edu, UCLA, Los Angeles, CA 90230, USA

Iman Dagher has been the Arabic and Islamic Studies Catalog Librarian at the University of
California, Los Angeles, since June 2008. She participates in the NACO, BIBCO and SACO
programs of the Program for Cooperative Cataloging (PCC), and served as a member of the PCC
Policy Committee from 2018 to 2020. Since 2013 she has been the Arabic NACO Funnel co-
coordinator. Prior to coming to UCLA, Iman worked for ProQuest, served as head librarian for
the Lebanese National Library Rehabilitation Project, and was a librarian at the American
University of Beirut.

Denise Soufi began her career in 2009 at Princeton University as the Islamic Manuscript
Cataloger under a grant-funded project. She is currently the Middle Eastern Cataloger at UNC-
Chapel Hill, where she catalogs materials in Persian, Arabic, Hebrew, and other Middle Eastern
languages. She also works in Special Collections, where she is working on a project to catalog
UNC's collection of Renaissance and early modern European manuscripts.

Authority Control of Arabic Personal Names: RDA and Beyond

This paper discusses the basics of creating name authority records for Arabic personal names in accordance with RDA instructions and PCC guidelines. A background into the use of romanization for non-Latin scripts in bibliographic and authority records is provided to establish context. Issues with romanization that are particular to Arabic are addressed. Separate sections on modern and classical names provide an overview of the major challenges, and strategies to enhance discovery are outlined. The paper concludes with an examination of the possible benefits of identity management and other changes in the authority control landscape for names in non-Latin script.

Introduction

As one of the main gateways to facilitating access to library resources, personal names have been given special consideration by librarians for a long time. The majority of bibliographic records have personal names as access points, whether as a creator or contributor (e.g. author, editor, illustrator, actor, artist) or as a subject (e.g. biography, views on a topic, contributions in a specific field). As Abdel Fattah and Ren state, "Proper nouns may be considered the most important query words in information retrieval."[1]

Authority control for personal names remains one of the main library functions that maintains and safeguards the flow of information between library data and users. Traditional authority control is costly in that it requires strong cataloging skills, periodic training, and continuous data maintenance. The process of selecting and maintaining unique and consistent authorized access points (AAPs) for personal names according to specific instructions and best practices, along with providing links to possible alternative forms of names (cross references or variants), permits the user to search the catalog under known names. This process also redirects the search if necessary to the authorized access point, and as a result allows for the collocation of related data under that unique name string.

However, this process is not always straightforward. Names can be complex and difficult. One person can use a variety of names (real name, literary name, pseudonym), change names over a period of time (maiden and married names), or write using different forms (full name versus initial), languages (Arabic, Persian, Chinese, Russian), or scripts (Latin, non-Latin). Different people may even share identical names. Libraries have been collectively and collaboratively working hard on sorting this complexity to ensure that the flow between information and its users is accurate and efficacious, and to maintain the integrity, validity, and reliability of their catalogs. Through the Name Authority Cooperative (NACO), founded in 1976 and now part of the Program for Cooperative Cataloging (PCC), several hundred institutions – following specific standards and established guidelines – have maintained a globally shared, reliable database of over 10 million authority records of persons, corporate bodies, family names, titles, and jurisdictions.[2] Despite the movement toward a linked data environment that will rely heavily on identity management systems, the NACO program remains active and effective and is still adding more participants, both within and outside of the United States.[3]

In order to add some flexibility to its program and to allow the inclusion of more communities with special interests into the NACO program, nineteen NACO funnel projects have been established based on a shared interest that is subject-, geographic-, and/or language-based, as well as predicated on consortial membership.[4] A number of these funnels are dedicated to various non-Latin script languages (i.e. languages that do not use the Latin alphabet in writing systems, such as Cyrillic, Arabic, Hebrew, Chinese, Japanese, and Korean), since each language has its own specific features and challenges necessitating different practices in creating authority data. Examples of language-based NACO funnels include the CJK (Chinese, Japanese, and Korean), Hebraica, Cyrillic, and Arabic funnels.

The Arabic NACO Funnel specializes in overseeing the creation of name authority records (NARs) for Arabic personal names, corporate bodies, works and expressions, and geographic names. The funnel began in 1997 under the leadership of Joyce Bell, at the time the Arabic Cataloger at Princeton University, with the goal of improving the discovery of Arabic materials by overseeing the creation of standardized, consistent Arabic name authority records to be added to the Name Authority File (NAF). It provides institutions with Arabic collections and language expertise staff the opportunity to participate collectively by following best practices in compliance with NACO guidelines.

The main purpose of this article is to provide a foundation for authority control of Arabic personal names and to address some of the challenges that catalogers may encounter during this process; it will then conclude with an examination of future trends such as identity management and their effect on authority work in non-Latin script languages.

Non-Latin Script Materials: Path to Discovery, Romanization versus Script

To meet the needs of teaching and research, materials in various non-Latin scripts have been widely collected by academic libraries in the US and the Western world for over a century. Processing and granting access to these materials so that they coexist with the Latin script materials in libraries' catalogs has been a challenge mainly due to system-related issues, technical problems, and limitations in language expertise among library staff.

Romanization is one of the important paths taken by libraries to incorporate bibliographic and authority data describing non-Latin script materials into their catalogs. ISO standards define romanization as the "script conversion from non-Roman to Roman script by means of transliteration, transcription or both."[5] Historically, North American libraries transcribed the original script onto cards and added a romanized title at the bottom. With the advent of electronic catalogs, initially only Latin (i.e. Roman) script could be used, so all descriptive and access fields had to be romanized.[6] Romanization is still applied and required despite the introduction

of vernacular scripts into library systems.[7] The romanization of non-Latin script data in NARs was required by the *Anglo-American Cataloguing Rules, 2nd Edition* (*AACR2*).[8] The current descriptive cataloging standard, *Resource Description and Access* (*RDA*), instructs catalogers to record metadata "in the language and script in which they appear on the sources from which they are taken," but most libraries in the US and Canada follow the optional addition to this instruction, which is to record data "in a transliterated form in addition to the form recorded in the script used on the source."[9]

Many voices have questioned the cost effectiveness, intensive labor, and margin of errors in creating romanized data, as well as the usefulness, necessity, and practicality of maintaining romanization alongside the presence of vernacular script. For example, Aissing states that for readers familiar with the original script, "the transliteration could be a serious obstacle resulting in partial or even total loss of information."[10] And Kunts points out that "for opponents of transliteration, transliteration is unreliable and serves neither librarians, bibliographers, nor users of bibliographic systems. The time spent transliterating text … is wasteful and unproductive."[11]

There is no doubt that the presence of the original script in shared bibliographic files can tremendously improve the discovery of these materials. When searching in the original script, the need to learn the rules of romanization tables and the agony of predicting how a name was romanized and entered are not necessary. This is especially true for languages that normally omit diacritics which indicate vocalization, such as Arabic, Persian, and Hebrew. Furthermore, metadata in the original script is more legible and less confusing to users.

Nevertheless, catalogers and library staff still support romanizing at least some library data. Lack of staff language expertise is the primary reason cited for maintaining romanized data.[12] Thanks to romanization, records for non-Latin script materials can be integrated into one library processing and handling system, allowing staff without language expertise to more easily complete many procedures that would otherwise be difficult, such as acquisition, serials check-

in, and interlibrary loans. This can be critical for materials that do not have an ISSN, ISBN, or call number. Furthermore, romanization facilitates discovery for users who are not proficient in the language or script, and permits browsing and sorting within one alphabetical list of results in the library catalog.[13] Another notable advantage of romanization is the ability to support integration with legacy library catalogs from the pre-Unicode era and which do not include scripts. A recent project has been initiated by OCLC and the UCLA Library to retrospectively add scripts to records in OCLC for Cyrillic and Armenian;[14] however, such a project may be more challenging for certain languages, such as Arabic and Hebrew. A recent survey on romanization found "that the absolute majority of respondents consider romanization an important aid in many library operations," with name and title elements considered to be the most important for romanization.[15] It is unlikely, therefore, that romanization will disappear from bibliographic and authority records in the Western world in the foreseeable future.

Having provided this background on the history and use of romanization in cataloging, we can now introduce some of the basic precepts and main challenges of traditional authority work for Arabic personal names, with the caveat that the cataloger should always consult the latest RDA instructions, the sources mentioned in this article, and other sources specified by the NACO program for the most complete and up-to-date information.

Romanization and Script in Arabic Name Authority Records

According to current PCC practice, all AAPs in the NAF must be established in Latin script; names in non-Latin script must be romanized according to the appropriate ALA-LC romanization table[16] and entered in MARC authority records following Model A (vernacular and transliteration).[17] The script forms of the names are added as variants, but they are not paired with the corresponding romanized fields, as is the case in bibliographic records. They are also added with romanized cited names in the 670 MARC field connected by an equals sign.

The ALA-LC romanization table for the Arabic language includes 26 rules with examples demonstrating the specificity of each rule and has been revised several times with enhancements and corrections (last revision 2014).[18] Complementing this table, the other main sources to consult on romanizing specific words are first Hans Wehr's *A Dictionary of Modern Written Arabic*,"[19] followed by *al-Munjid*.[20] Although these sources are language dictionaries and do not include listings of proper names, consulting them is still helpful since many personal names (given or surnames) have meanings as ordinary adjectives (Jamīlah [beautiful]), nouns (Jamāl [beauty]), or descriptive/attributive names (al-Kāmil [the Perfect]).

While romanization can be simple and straightforward for some languages, romanizing the Arabic language is much more complicated and challenging, with some additional difficulties pertaining to the nature of Arabic personal names in particular. The first challenge is represented by *tashkīl*, or vocalization/vowelization, which can be defined as the process of adding the appropriate diacritical marks (*ḥarakāt*) to Arabic letters. The diacritical marks indicate short vowels as follows (shown on the letter د): دَ = *fatḥah* (short a), دِ = *kasrah* (short i), دُ = *ḍammah* (short u), دّ = *shaddah* (gemination or consonant doubling), and دْ = *sukūn* (absence of any vowels). Arabic language materials usually appear without any diacritics on published resources, and it is up to the reader to fill in these short vowels based on their familiarity with the language. When a personal name is not vocalized on the resource, the process of *tashkīl* can be problematic as more factors are introduced, such as unfamiliarity with the specific pronunciation of the name for both catalogers and users, or differences in pronunciation for the same name between countries and regions in the Arab world. Consensus may be difficult to reach on all forms of romanized names; however, following the rules and best practices can help ensure some level of consistency and better access.[21] Some specific challenges in vocalization due to an absence of diacritics are as follows:

- Lack of *shaddah* in the displayed form can result in romanizing the name incorrectly

كريم = Karīm or Kurayyim

- Lack of short vowels can cause ambiguity in the pronunciation of the name and possibly the gendering of the name:

كميل = Kamīl or Kumayl

علوان = ʿAlwān or ʿIlwān or ʿUlwān

حسن = Ḥasan, male name, or Ḥusn, female name

While all these examples look identical in the Arabic script without diacritics, they can be pronounced very differently, resulting in a variety of romanized forms. Romanizing is especially difficult when there is no obvious or familiar meaning or language root of the name to rely upon. Familiarity with the language and culture helps the cataloger to properly vocalize the name.[22]

However, in certain cases catalogers may need to go beyond the source in hand to verify how a name is pronounced in order to romanize it properly. For example, consulting the usage in the NAF, as well as bibliographic records, and checking the most commonly used romanized form can permit certain consistency in the NAF. Also, occasionally some non-Latin script sources may provide a "non-standardized" romanized form of the name in the Latin script parallel information. When present, such data can assist in determining the proper romanization. For example, the romanized form of the Arabic name found on resources in the case of "Intisar Ahmed Hassan Hameed" provides a reasonable indication that the name حميد is pronounced Ḥamīd (حَميد), not Ḥumayyid (حُمَيّد) or Ḥumayd (حُمَيْد). Furthermore, searching other outside resources and reference books such as biographical dictionaries can provide additional helpful information, especially if found vocalized or romanized, such as *al-Munjid* for Arabic names. With more contemporary names, some web resources may display the vocalized form or the Latin script form of the name, such as on LinkedIn or Facebook. In certain puzzling and rare situations, consulting YouTube for interviews that mention the author's name can be helpful in determining how a name is pronounced. Consulting native speakers or contacting the person

(when feasible) to obtain the proper pronunciation of the name, or for other information needed

to break a conflict, might end up being the ultimate approach.

When citing the resource cataloged in the MARC authority record, the Library of

Congress's *Descriptive Cataloging Manual* (*DCMZ1*) instructs the cataloger to indicate when the

name is vocalized by adding "[voc.]" or partially vocalized by adding "[part. voc]." If the name

is found already romanized, it is cited with the addition "[rom.]."[23] Providing this information

helps with justifying the forms of the names used in the AAP or the variant(s).

There are also some guidelines pertaining to pronunciation that can help catalogers

maintain consistency when creating the AAP. For example, some Arabic names may include an

element called *nisbah*, an attribution indicating origin, birthplace, homeland, or residence that

usually ends with "ī." Although *nisbahs* may be pronounced differently in various regions, it is

recommended to maintain consistency in order to facilitate access. Therefore, when romanizing

such names, it is important to follow the standardized romanized form of the place as the root.

For instance:

مصري = Miṣrī, not Maṣrī; موصلي = Mawṣillī, not Mūṣallī; سعودي = Saʿūdī not Suʿūdī

The Arabic letter pronounced like the English letter J (ج) is pronounced as the English hard G in

Egyptian colloquial. The instruction in the Arabic romanization table is to romanize the form

displayed on the source according to the language and not the dialect;[24] hence:

ماجدة = Mājidah, not Māgdah; جمال = Jamāl, not Gamāl

Some foreign names may be used by Arabs or may occur in an Arabic context written with

Arabic letters. Romanizing can be especially challenging when the actual sound is not

represented in the Arabic alphabet. The instruction is to romanize according to the rules for

Arabic romanization.[25] For example:

فيليب (Philip) = Fīlīb, not Fīlīp; بول (Paul) = Būl, not Pūl

This also occurs when adding Arabic variants to Latin script names in the NAF.

Modern Arabic Personal Names: Some Basic Issues

The Arabic NACO Funnel follows RDA guidelines for creating NARs. For Arabic personal names, the funnel recognizes two types: classical names and modern names. Guidelines for creating NARs vary for each type and include instructions for choosing the preferred form of the name, suggested variants, sources consulted, and citing guidelines. Authors who have been active since the beginning of the 20th century are considered to be modern.[26]

Choosing the preferred form of a name is the first step in creating the AAP for a person. According to RDA, this is the "name by which the person is commonly known. The name chosen can be the person's real name, pseudonym, title of nobility, nickname, initials, or other appellation."[27] The main source of information is the manifestations associated with the person.

The basic instruction in RDA is that modern authors are assumed to have surnames[28] and should be entered under the surname, followed by a comma, then the rest of the name in the order it is displayed on the source, starting with the given name (*ism*).[29] It should be noted that the concept and the adoption of surnames among Arabs has not been fully and consistently employed among all communities in Arab countries; for some the assumed surnames are actually the fathers' or grandfathers' given names.

In some instances, the surname is not obvious, and the form of the name displayed on the resource consists of several elements. The general instruction to follow is to consider the last element as the surname and choose it as the entry element.[30] For example:

On the resource: Muḥammad Sāmī ʿAbd al-Salām Ḥasānayn

Form in AAP: Ḥasānayn, Muḥammad Sāmī ʿAbd al-Salām

There is no need to question whether "Sāmī" or "ʿAbd al-Salām" is part of the surname.

The form of the name itself may vary in fullness in printed Arabic resources (title page, cover, CIP data, spine, etc.). In RDA, the title page is considered to be the main source of information,[31] and the form displayed on the title page is the basis for choosing the preferred

form. Nevertheless, citing other forms of the name appearing on the resource in the authority record can be helpful for identification and/or differentiation purposes. While we are instructed to take the form on the title page as the basis for the preferred name and consider it as "commonly known," it should be noted that in some instances, other evidence may suggest a different form of the name than the one displayed on the title page, such as common and consistent usage or the person's self-preference.

Ibn or bin (son of) and bint (daughter of) are the patronymic elements known as nasab and are part of a type of compound name. Although less common in modern names, some countries and regions still use these elements in their names. The nasab seems to be displayed inconsistently in currently published resources; for example:

Name on title page: Hānī ʿAlī Ḥawwās

CIP data: Hānī ibn ʿAlī Muḥammad ʿAlī Ḥawwās

This happens because people tend to use their fathers' or grandfathers' names as part of their names and then inconsistently add or drop the word indicating the nasab (i.e, ibn or bint).

Arabic personal names are often accompanied by academic titles or terms of honor. The guidelines in RDA are to omit these terms when choosing the preferred name unless needed to distinguish one AAP from another.[32] Some examples of Arabic titles include: al-Ustādh (honorific title), al-Duktūr (title indicating a person holding a doctorate degree or doctor of medicine), al-Muhandis (engineer), Āyat Allāh (Shiite religious leader), Ḥājj (traditional title of respect for Muslim pilgrims), Shaykh (title of social respect, or for chiefs of tribes or religious leaders or scholars), and al-Sayyid.

Table 1

Arabic form on book	Romanized form	Authorized form
الاستاذ السيد عفيفي	al-Ustādh al-Sayyid ʿAfīfī	ʿAfīfī, al-Sayyid
الدكتور رضوان السيد	al-Duktūr Riḍwān al-Sayyid	Sayyid, Riḍwān
السيد محمد عبد الحق	al-Sayyid Muḥammad ʿAbd al-Ḥaqq	ʿAbd al-Ḥaqq, Muḥammad

Al-Sayyid, or Sayyid, in particular needs to be treated with attention since it may vary in nature. It can be a given name, a surname, or a term of honor. It can also be a title for Shiite religious leaders or scholars who are descendants of the family of Muḥammad, the prophet of Islam. Table 1 shows the use of al-Sayyid as a given name, as a surname, and as an honorific. In the last example, al-Sayyid should be dropped from the entry because it is functioning as a term of honor. To determine the nature of this element, catalogers may need to check the display of the name cited in other places of the publication or even examine external sources.

Normalization

In order to enhance discovery in online catalogs, names are converted to normalized forms in which all diacritics and most punctuation are removed, all letters are converted to uppercase, and all modified letters are converted to their unmodified equivalents.[33] When romanized, different Arabic names may normalize to the same form; for example:

Sālim (سالم) and Salīm (سليم) normalize to Salim

Amal (أمل) and Amāl (أمال) normalize to Amal

The presence of the script variant is very helpful in such cases.

Compound names

Compound names are common among Arabic personal names and should always be written separately when romanized. Names beginning with ʿAbd (servant of) are usually combined with one of the names of Allāh (God):

عبد الله = ʿAbd Allāh (not ʿAbdulllāh); عبد الكريم = ʿAbd al-Karīm

Other instances of compound names consist of Muḥammad with another name, such as Muḥammad ʿAlī and Muḥammad Ḥusayn. These can be tricky to determine, so observation of the occurrence of the name in the resources can help.

Arabic names may begin with the initial definite article *al-* which is the equivalent of "the" in English. The instruction in RDA is to omit it when that part of the name is entered as the first element;[34] however, it is retained and entered in lower case in the remaining part of the name. For example:

Form on resource: al-Ḥasan ibn Mūsá al-ʿAnānī al-Jazāʾirī

Form in AAP: Jazāʾirī, al-Ḥasan ibn Mūsá al-ʿAnānī.

On the other hand, the word *āl* (آل, meaning "family" or "clan") should be retained as a separate element of a compound name. It is always capitalized and followed by the name of the family:[35]

آل ثاني، حمد بن علي بن جبر = Āl Thānī, Ḥamad ibn ʿAlī ibn Jabr

Variants in the Name Authority File

According to RDA instructions, adding variants is not required but rather left to the cataloger's judgment; however, the alternative linguistic form of a name may be added as a variant.[36] The Arabic NACO Funnel highly recommends adding different possible forms of romanization as variants, as well as non-standard romanizations found on the resource. [37] These variants provide the user with the ability to retrieve the AAP for the name regardless of the different possible ways of pronouncing the name. Occasionally it is impossible for the cataloger to romanize a name correctly with certainty, making the need for variants all the more crucial. More importantly, adding the script variant is the ultimate key to discovery since it will grant access to the exact typographical form of the name.

Attributes, Disambiguation, and Maintenance

With RDA, authority records can be more resourceful and dynamic with greater potential to link to related authority lists and databases in anticipation of better serving current and future bibliographic needs. It has allowed authority records to move "beyond what is required for an access point and toward a record for the person."[38] The NAR can be seen as a snapshot of a

person's identity. RDA instructions permit adding several attributes and relationships to describe a person using the different MARC fields.[39]

Although adding these attributes is left to the cataloger's judgment, instances of ambiguity resulting in wrong attributions in the shared file demonstrate the potential value of such attributes in fulfilling the user task of finding and identifying agents,[40] especially when we consider that the name authority data will be used for generations to come. Conversely, in adding such data, catalogers should exercise caution and respect for the person's privacy and include useful information that is only publicly sharable by them and that may serve in the identification and disambiguation process.[41]

Some of the many brief name authority records in the NAF that have very little data recorded about the person fail to identify and distinguish them from others with similar names and may also link works to the wrong name. Similarly to other languages, this can be very important for Arabic names that can be communal and recycled within families and across generations. It can be challenging for catalogers to use the NAR with confidence and link it to the proper bibliographic resources. Recording attributes is one method for solving the issue of false attribution and for preventing duplication in the NAF. This is one of the main maintenance functions that imposes time and intellectual effort on both NACO catalogers and Library of Congress staff.[42] It cannot be ignored as it will jeopardize the primary function of the authority file to serve as a critical link in the chain of connecting agents and resources. Certain guidelines are essential to avoid duplications among AAPs and to prevent such broken links, such as searching the authority file thoroughly and efficiently for existing NARs. With non-Latin script names such as Arabic, it is necessary to search in different possible romanized forms, and to verify the common romanized form used by other catalogers. For example, for the Arabic name حسني, it is necessary to search under both Ḥusnī and Ḥasanī. Also, keyword searching on

different name elements can retrieve records of existing NARs of the person that have been created based on different usage/displayed forms such as:

Ṣabbāgh, Hānī, 1934- / Sabbāgh, Hānī Rājī

Salwī, Muṣṭafá / Salwī, Muṣṭafá bin al-ʿArabī.

Searching names in the original script can be the best solution, on the condition that a script variant is available in the existing NAR, especially for unfamiliar names with dubious pronunciation. One such example is the Arabic name بلحولة، المنصف which had the duplicates: Balḥawlah, al-Munṣif / Bilḥūlah, al-Munṣif. Likewise, searching under the Latin script form of the name if available is also necessary, since some authors write in both Latin and non-Latin script languages. Example of resolved duplicates:

Khūrī, Jurays Naʿīm / Khoury, Jeries, 1972-

Examining the citations in the authority record for occurrences of other forms that are not reflected in the existing variants can also be helpful, especially with cases in which the fuller form of the name is cited in the authority record but is not part of the preferred name or the existing variants. This is very common in some Arabic publications when the fuller form of the name is displayed in the CIP, title page verso, etc. Example of resolved duplicates:

Amīn, ʿAbd al-Bāsiṭ Muḥammad

ʿAbd al-Hādī, ʿAbd al-Bāsiṭ Muḥammad Amīn Sulaymān

In case of conflict, there are attributes that can be used in the AAP in order to distinguish one access point from another, such as the title of the person (a term indicative of rank, honor, or office), fuller form of name, profession or occupation, and period of activity.[43]

Dates

Dates in modern Arabic resources may be provided in the Hijrī (Islamic lunar) calendar. Catalogers are expected to convert the date to the Gregorian calendar when adding the dates in

the AAP and variants or in the MARC 046 field.[44] Since the lunar Hijrī calendar usually does not correspond exactly to the solar Gregorian calendar, the converted date is often a choice between two Gregorian years. For example, a birth date of 1389 Hijrī converts to 1969 or 1970, with subfield d in the AAP coded as follows:

$d 1969 or 1970-

But having the day and month in the Hijrī calendar on resources allows one to identify a precise Gregorian date; for example, a birth date of the 6th of the Hijrī month Muḥarram, 1389, corresponds to March 25, 1969, allowing for an exact birth year in subfield d of the AAP. In the citation in the authority record, dates are transcribed as found on the resource, and if both Hijrī and Gregorian dates are given, they are both recorded.[45]

In Arabic resources published in Iran, dates are usually provided using the Iranian Hijrī-Shamsī calendar,[46] which is a solar calendar that began in the same year as the Islamic lunar calendar. Because the Iranian year begins on the spring solstice, conversion from a Hijrī-Shamsī year[47] always results in a choice between two Gregorian years unless the date is more specific.

Special Challenges with Modern Arabic Names

Modern North African Names

In modern North African names, if the words *ibn* and *abū* do not contain the initial alif, they are romanized as *bin* and *bū*, respectively.[48] References using *ibn* and *abū* should be made to ensure adequate access, particularly if the term is at the beginning of a name. Additionally, a compound name beginning with *bin* may be displayed in a combined form in the Arabic script. Such names are romanized with a hyphen between *bin* and the remainder of the name.[49] A reference is provided for the romanized compound form as well as a reference using *ibn*. An example is the name عبد الله بنسعيد, which has the following AAP and romanized variants:

100 1 Bin-Saʿīd, ʿAbd Allāh

400 1 Binsaʿīd, ʿAbd Allāh

400 1 Ibn Saʿīd, ʿAbd Allāh

All of these variants enhance discovery by providing a variety of ways that such a name could be romanized.

North African names may also be presented in inverted order on the resource. If it is not clear as to whether the name is in inverted order, the preferred name is established in direct order.[50] For those unfamiliar with North African names, such a name may pose romanization challenges if it is not of Arabic origin and does not contain vowel marks to aid in pronunciation. Special resources are available for the pronunciation of North African names.[51]

Modern Iranian Authors Writing in Arabic

It is not uncommon for an author whose name is clearly of Iranian origin to write in Arabic. In such a case, the name is established in Arabic and romanized according to the Arabic romanization table, but the choice of entry is based on the conventions for Persian. Modern Persian names frequently have a compound surname which is used as the entry element; references are made from all other surnames present[52] in order to aid users who may not be familiar with the construction of Persian names. An example is the author Muḥammad ʿAlī al-Ḥusaynī al-Abṭaḥī al-Iṣfahānī. While his name is romanized according to Arabic romanization conventions, the choice of entry is not the last part of his name, but rather the three-part surname. References are then made from the other surnames, retaining the order of the parts as presented on the piece in hand.

100 1 Ḥusaynī al-Abṭaḥī al-Iṣfahānī, Muḥammad ʿAlī, $d -1950 or 1951

400 1 Abṭaḥī al-Iṣfahānī, Muḥammad ʿAlī al-Ḥusaynī, $d -1950 or 1951

400 1 Iṣfahānī, Muḥammad ʿAlī al-Ḥusaynī al- Abṭaḥī, $d -1950 or 1951

Should the author also write in Persian, additional references can be created based on the guidelines for establishing a modern Persian author.[53]

Romanized Name on Resource

Sometimes the resource, while in Arabic, also includes a romanized form of the author's name. This romanized form should be entered as a variant if it does not normalize to the AAP, as it provides an important access point for users not well-acquainted with the language or the romanization rules. However, if the romanized surname on the resource begins with the equivalent of the definite article "al-" (ال, also romanized "el") and is capitalized and separated from the rest of the surname with a hyphen or space, two references are made: one from the surname as a whole, and one from the surname with the "Al-" equivalent dropped and moved to the end of the name[54] (as is done for Arabic names in their romanized form). An example is Yāsir ibn ʿAbd al-ʿAzīz al-Ghaslān, whose name is romanized Yasser Al-Ghaslan on the resource. References from the found romanization are created both with "Al-" as part of the surname and without:

> 400 1 Al-Ghaslan, Yasser
>
> 400 1 Ghaslan, Yasser Al-

If the "al-" equivalent is not capitalized, it is simply dropped from the reference.

Classical Arabic Names

Pre-modern names in any language have special instructions due to the wide variety of naming conventions that have existed across time and cultures, and reference sources usually need to be consulted in order to determine the preferred form of the name. In the case of Arabic names, the *Arabic NACO Manual* defines a classical Arabic name as belonging to a person who was active prior to the 20th century.[55] Such names are generally assumed to not contain surnames according to RDA, which supplies instructions for names that are "originally written in the Arabic alphabet" and that "do not contain a surname or a name performing the function of a surname."[56]

Parts of a Classical Arabic Name

The parts of a classical Arabic name have been explained in a number of articles[57] and in RDA.[58] The latter lists seven parts to consider when formulating the preferred name in Arabic script:

(1) *Khiṭāb*: honorific compound name, the last part of which can be al-Dīn, al-Islām, etc. Examples: Naṣīr al-Dīn, Zayn al-Islām

(2) *Kunyah*: honorific compound name beginning with *abū* (father of; *abī* in genitive) or *umm* (mother of). Examples: Abū Bakr, Umm Zayd

(3) *Ism*: given name, single or compound

(4) Patronymic (*nasab*): compound name indicating descent (previously explained)

(5) *Laqab*: descriptive epithet describing a personal quality. Examples: al-Jāḥiẓ (the bug-eyed), al-Rashīd (the rightly guided), al-Qāḍī al-Fāḍil (the prominent judge)

(6) *Nisbah*: attributive name typically ending in ī, indicating geographical origin or residence, tribe or descent, sect (*madhhab*), or profession. Examples: al-Ṭabarī (from Tabaristan), al-Ayyūbī (of the Ayyub family line), al-Iskāfī (the cobbler)

(7) *Takhalluṣ*: pen name; rarely a component of a classical Arabic name but used in Persian and Urdu names

Constructing a Classical Arabic Name

Choosing the first element of a classical Arabic name can be the most difficult step in the construction of the preferred name due to the lack of a surname. According to RDA, the first element consists of the part or parts of the name by which the person is best known.[59] This is determined by consulting at least one reference source aside from the piece being cataloged. The *Arabic NACO Manual* recommends the following reference sources along with instructions for their use:[60]

Khayr al-Dīn al-Ziriklī, *al-Aʿlām* [The Luminaries][61]

Carl Brockelmann. *Geschichte der arabischen Litteratur* [History of Arabic Literature][62]

ʿUmar Riḍā Kaḥḥālah, *Muʿjam al-Muʾallifīn* [Index of Authors][63]

Other reference sources may be consulted if the person cannot be found in the above list; the *Encyclopaedia of Islam*, for example, may be helpful for well-known persons who are not authors.[64] If the person cannot be found in any reference sources, guidelines for choosing the first element are provided in the *Arabic NACO Manual*[65] and RDA.[66]

Having determined the first element of the name, a decision must be made as to which parts will be included in the remainder of the name, as well as how the parts should be ordered.[67] Applying the instructions often results in the following construction for the preferred form: [first element], [ism + patronymic]. Note that the first element is basically functioning as a surname. Examples of this construction:

Bakrī, Muḥammad ibn Aḥmad   Māridīnī, ʿAbd Allāh ibn Khalīl

Variant Access Points

References recommended by the *Arabic NACO Manual* include those prescribed for modern names as well as several which are particular to classical names.[68] For example, if the resource includes a non-standard romanization, including this form as a variant enhances discovery for users who are not familiar with the language or the romanization rules. Making references from unused "surnames" also improves access since users are not likely to be familiar with the guidelines for choosing the first element of the name. *Nisbahs* and *laqabs* are commonly used to create variants. For example, for the name Muḥammad ibn Aḥmad al-Bakrī al-Ṣiddīqī, Bakrī is determined to be the entry element and a variant is constructed using the *nisbah* al-Ṣiddīqī:

100 1 Bakrī, Muḥammad ibn Aḥmad

400 1 Ṣiddīqī, Muḥammad ibn Aḥmad

Lastly, a direct order reference beginning with the given name and containing all the elements in the authorized access point can be helpful for users since the first element is not actually a surname, but merely functioning as such:

> 100 1 Bakrī, Muḥammad ibn Aḥmad
>
> 400 0 Muḥammad ibn Aḥmad al-Bakrī

Challenges

Most of the challenges encountered in creating NARs for classical Arabic names are similar to the challenges in establishing pre-modern names in general, such as conflicts between reference sources or uncertainty about dates, as well as the problems encountered in establishing Arabic names in particular, such as difficulties with vocalization. Searching the NAF can be particularly problematic, as the name presented on the title page may not correspond to the established AAP or its variants, or it may be provided in a truncated form that, when searched, returns too many results. To find the authority record for a pre-modern author in such a case, one may have to search the bibliographic file for other editions of the work to see if they have been cataloged with the correct NAR, or check the resource to find out if the editor has provided any helpful information about the author's name, such as variants or a fuller form. Reference sources may also have to be consulted; for classical Arabic names Brockelmann's *Geschichte der arabischen Litteratur* is a good starting point since it contains an index of titles. If the correct NAR is identified, it is important to add a reference from the form of the name found in the resource in case it is used again in other resources. Otherwise, a new NAR should be created if thorough searching does not lead to any results.

Beyond Traditional Authority Control

As library data has become more global and shareable, the Virtual International Authority File (VIAF)[69] has come to play a significant role in linking data from national and regional libraries

around the world into one "super authority record" bearing a unique VIAF ID. Only three Arab libraries are currently represented in VIAF: Bibliotheca Alexandrina (in Egypt), the Lebanese National Library, and the National Library of Morocco.

In preparing for a linked data environment, VIAF fits well in the efforts of libraries to move away from traditional authority control, in which unique strings are tied to a set of standards and secured in a monitored record, towards unique and persistent identifiers that present web-resolvable entities. Identity management systems such as ISNI, ORCID, and Wikidata differentiate between entities by using identifiers and do not require the use of a uniquely constructed string; they are also capable of accommodating most languages and scripts (including Arabic), a big plus for the many scripts that cannot be used in a MARC authority record. ISNI and Wikidata allow collaborative and global contribution in addition to individuals' input.

As part of its strategic plan, the PCC is working to "accelerate the movement toward ubiquitous identifier creation and identity management at the network level."[70] To that end, it has initiated several projects in preparation to accommodate a *NACO Lite* program that would encompass "streamlined procedures for contributing authority data and … broadened participation in that activity."[71] One such project is the PCC URIs in MARC Pilot, in which participants add Uniform Resource Identifiers (URIs) to bibliographic and authority records.[72] The success of such pilots may lead to *NACO Lite* authority records that require only a name string and a URI from a source such as Wikidata, as recently described by Frank.[73] The URI serves as a stable identifier for the entity and obviates the need for the time-consuming and costly tasks of creating and maintaining complicated authority records, as well as repeatedly romanizing names in non-Latin scripts. For Arabic personal names, such an approach can be beneficial when local experts and Arabic scholars can enrich identity management systems with aliases, credentials, and related works, or when data from Arab libraries are linked to these

identities; moreover, the quality data already contained in the NACO file can be linked to such systems using the authority control number. A lower barrier to participation in NACO record creation will hopefully expand the number of identified Arabic names in the NAF in particular, and non-Latin script names in general.

Conclusion

Name authority control for non-Latin script languages poses serious challenges. While some of these challenges are shared across languages, each language also has its own features and issues. Due to the nature of the Arabic language and script, name authority control for personal names presents certain complexities which we have attempted to elucidate. The romanization requirement in particular produces a number of challenges that require special guidelines in order to ensure robust discovery and access. Identity-based management in the linked data environment has the potential to ameliorate some of the difficulties associated with authority control for names in non-Latin scripts.

Notes

[1] Mohamed Abdel Fattah and Fuji Ren, "English-Arabic Proper-Noun Transliteration-Pairs Creation," *Journal of the American Society for Information Science & Technology* 59, no. 10 (2008): 1675.

[2] Judith P. Cannan, Paul Frank, and Les Hawkins, "LC/NACO Authority File in the Library of Congress BIBFRAME Pilots," *Journal of Library Metadata* 19, nos. 1-2 (2019): 39.

[3] Ibid., 40.

[4] John D. Byrum Jr., "NACO: A Cooperative Model for Building and Maintaining a Shared Name Authority Database," *Cataloging & Classification Quarterly* 38, nos. 3-4 (2004): 244.

[5] ISO 3.1.6.14 ("ISO 5127:2017(en) Information and documentation — Foundation and vocabulary," International Organization for Standardization, accessed May 25, 2020, https://www.iso.org/obp/ui/#iso:std:iso:5127:ed-2:v1:en:sec:3.1.6.14).

[6] ALCTS Non-English Access Working Group on Romanization, *Report*, Dec. 15, 2009, 3. http://www.ala.org/alcts/sites/ala.org.alcts/files/content/ianda/nonenglish/apd15a.pdf

[7] Ali Kamel Shaker, "Bibliographic Access to Non-Roman Scripts in Library OPACs: A Study of Selected ARL Academic Libraries in the United States" (PhD diss., University of Pittsburgh, 2002), 23-24; OCLC, *International Cataloging: Use Non-Latin Scripts* (Dublin: OCLC Online Computer Library, 2016), 3.

[8] Rule 22.3C2, *Anglo-American Cataloguing Rules*, 2nd ed. (Chicago: American Library Association, 2005 update). https://www.rdatoolkit.org

[9] RDA 1.4 (Language and Script), *Resource Description and Access*, RDA Toolkit, American Library Association, Canadian Federation of Library Associations, and CILIP, last modified April 2017, accessed May 18, 2020, https://www.rdatoolkit.org. (This paper was prepared before the implementation of new RDA, expected release in December 2020)

[10] Alena L. Aissing, "Cyrillic Transliteration and its Users," *College & Research Libraries* 56, no. 3 (1995): 5.

[11] Blair Kuntz, "Is the Organization for Standardization (ISO)'s Arabic Transliteration Scheme an Improvement over Library of Congress?" *MELA Notes* 78 (2005): 55.

[12] ALCTS Non-English Access, 13.

[13] Ibid., 6-12.

[14] Merrilee Proffitt, "Кириллица в WorldCat" [Cyrillic in WorldCat], Hanging Together: the OCLC Research Blog, OCLC, April 15 2020, retrieved October 12, 2020. https://hangingtogether.org/?p=7868

[15] Linked Data for Production: Pathway to Implementation (LD4P2), Non-Latin Script Materials Affinity Group, *Survey on Romanization: Report*, March 4, 2020, 16-17. https://docs.google.com/document/d/1HVmtQxXSE0aphipdACNTTck6rzT3gPEoqjzKynYgL78/edit

[16] RDA 9.2.2.5.3 (Preferred Name for Person: Name Found in a Non-preferred Script) and corresponding LC-PCC PS; the ALA-LC Romanization Tables are available at https://www.loc.gov/catdir/cpso/roman.html.

[17] "Appendix C - Multiscript Records," MARC 21 Format for Authority Data, Library of Congress, last modified May 27, 2020, accessed June 19, 2020, https://www.loc.gov/marc/authority/ecadmulti.html.

[18] Library of Congress, *Arabic*. https://www.loc.gov/catdir/cpso/romanization/arabic.pdf. For other systems of romanizing Arabic, see David R. Lawson, "An Assessment of Arabic Transliteration Systems," *Technical Services Quarterly* 27, no. 2, 164-177.

[19] Hans Wehr, *A Dictionary of Modern Written Arabic*, 4th ed. (Ithaca, New York: Spoken Language Services, 1994).

[20] Luwīs Maʿlūf, *al-Munjid* [The Helper] (Bayrūt: al-Maṭbaʿah al-Kāthūlīkīyah lil-Ābāʾ al-Yasūʿīyīn, 1908). Available in numerous reprints.

[21] Kuntz, "Is the Organization for Standardization (ISO)'s Arabic Transliteration Scheme …," 63.

[22] Ali Houissa, "Arabic Personal Names: Their Components and Rendering in Catalog Entries," *Cataloging & Classification Quarterly* 13, no. 2 (1991): 3.

[23] Policy, Training, and Cooperative Programs Division, Library of Congress, *Descriptive Cataloging Manual Z1: Name and Series Authority Records*, last revised April 2020, 670 Source Data Found, 3. https://www.loc.gov/catdir/cpso/dcmz1.pdf

[24] Library of Congress, *Arabic*, 1.

[25] Ibid., 9, rule 22.

[26] "Modern Authors," Arabic NACO Manual, Princeton University Library, last modified July 14, 2014, accessed May 18, 2020, https://library.princeton.edu/departments/tsd/katmandu/cp20/aranatoc.html.

[27] RDA 9.2.2.3 (Choosing Preferred Name for Person).

[28] RDA Appendix F.1.1.1 (Names in the Arabic Alphabet: Application).

[29] RDA 9.2.2.9.2 (General Guidelines on Recording Names Containing a Surname: Part of the Name Treated as a Surname).

[30] "Modern Authors," Arabic NACO Manual.

[31] RDA 2.2.2.2 (Preferred Source of Information: Manifestations Consisting of One or More Pages, Leaves, Sheets, or Cards…).

[32] RDA 9.19.1.7 (Authorized Access Point Representing a Person: Other Term of Rank, Honour, or Office).

³³ *Descriptive Cataloging Manual Z1*, Introduction, 6.

³⁴ RDA Appendix C (Initial Articles).

³⁵ Heather Hedden, "Arabic Names," *The Indexer* 25, no. 3 (2007): C12.

³⁶ RDA 9.2.3.9 (Variant Name for Person: Alternative Linguistic Form of Name).

³⁷ "References," Arabic NACO Manual.

³⁸ Chris Oliver, *Introducing RDA: A Guide to the Basics* (Chicago: American Library Association, 2010), 61.

³⁹ RDA Section 9 (Identifying Persons).

⁴⁰ RDA 0.4.2.1 (Responsiveness to User Needs).

⁴¹ Heather Lea Moulaison, Felicity Dykas, and John M. Budd, "Foucault, the Author, and Intellectual Debt: Capturing the Author-Function Through Attributes, Relationships, and Events in Knowledge Organization Systems," *Knowledge Organization* 41, no. 1 (2014): 40-41.

⁴² On NACO bibliographic file maintenance, see https://www.loc.gov/aba/pcc/naco/bfmguide.html.

⁴³ RDA 9.19.1 (Authorized Access Point Representing a Person).

⁴⁴ Online calendar converters can be used for this purpose, such as the one available at Islamic Philosophy Online, http://www.muslimphilosophy.com/ip/hijri.htm.

⁴⁵ "Dates," Arabic NACO Manual.

⁴⁶ An exception is dates for pre-modern names, which are usually provided in the Hijrī calendar.

⁴⁷ An online calendar converter providing conversion between the Iranian, Gregorian, and Hijrī calendars is available at Iran Chamber Society, http://www.iranchamber.com/calendar/converter/iranian_calendar_converter.php.

⁴⁸ "Modern Authors," Arabic NACO Manual.

⁴⁹ Library of Congress, *Arabic*, 10, rule 25.

⁵⁰ "Modern Authors," Arabic NACO Manual.

⁵¹ Such as ʿAbd al-Wahhāb Bin Manṣūr, *Kashshāf asmāʾ al-usar al-Maghribīyah* [Discloser of the Names of Moroccan Families] (al-Rabāṭ: al-Maṭbaʿah al-Malikīyah, 1972).

52 "Modern Authors," Arabic NACO Manual. For examples of Persian surnames, see "Types of Surnames," Persian NACO Manual: Personal Names, Princeton University, accessed May 20, 2020, https://library.princeton.edu/departments/tsd/katmandu/persian2/pernatoc.html.

53 Persian NACO Manual: Personal Names.

54 Rule 22.4, *Library of Congress Rule Interpretations*, 2nd ed. (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 2002). https://www.itsmarc.com/crs/mergedProjects/lcri/lcri/22_4__lcri.htm. These instructions are currently still followed by the Arabic NACO Funnel.

55 "Classical Names," Arabic NACO Manual.

56 RDA Appendix F.1 (Names in the Arabic Alphabet).

57 Houissa, "Arabic Personal Names," 10-12; Magda El-Sherbini, "Bibliotheca Alexandrina's Model for Arabic Name Authority Control," *Library Resources & Technical Services* 57, no. 1 (2013): 7; Iman Khairy, "Authority Control of Arabic Personal Names from the Classical Period at the Bibliotheca Alexandrina," revised version, January 2006 (paper presented at MELCOM International 27th Annual Conference, Alexandria, Egypt, May 23-25, 2005), 1-2. Both El-Sherbini and Khairy leave out *takhalluṣ*, and El-Sherbini also leaves out *khiṭāb*.

58 RDA Appendix F.1.1.4 (Names in the Arabic Alphabet: Order of Parts).

59 RDA Appendix F.1.1.2 (Names in the Arabic Alphabet: First Element).

60 "Classical Names," Arabic NACO Manual.

61 First published: Cairo, Egypt: al-Maṭbaʿah al-ʿArabīyah, 1927-1928. Available in numerous reprints. The 1954 printing is available at Arabic Collections Online, http://dlib.nyu.edu/aco/.

62 Second edition published: Leiden: Brill, 1932-1949. Reprinted in 1996. Also available online by subscription through Brill, as well as in English translation: *History of the Arabic Written Tradition* (Leiden: Brill, 2016-2019).

63 First published: Damascus, Syria: al-Maktabah al-ʿArabīyah, 1957-1961. Available in several reprints. The 1957 printing is available at Arabic Collections Online.

64 For other possible reference sources, see Khairy, "Authority Control"; El-Sherbini, "Bibliotheca Alexandrina's Model," 10; RDA F.1.1.1 (Names in the Arabic Alphabet: Application).

[65] "Classical Names," Arabic NACO Manual.

[66] RDA Appendix F.1.1.2 (Names in the Arabic Alphabet: First Element).

[67] RDA Appendix F.1.1.3 (Names in the Arabic Alphabet: Essential Parts of the Name) and F.1.1.4 (Names in the Arabic Alphabet: Order of Parts).

[68] "References," Arabic NACO Manual.

[69] See http://viaf.org/.

[70] Program for Cooperative Cataloging, *PCC (Program for Cooperative Cataloging) Strategic Directions: January 2018-December 2021*, revised January 24, 2019, 5. https://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2021.pdf

[71] Erin Stalberg, John Riemer, Andrew MacEwan, Jennifer A. Liss, Violeta Ilik, Stephen Hearn, Jean Godby, Paul Frank, Michelle Durocher, and Amber Billey, "Exploring Models for Shared Identity Management at a Global Scale: The Work of the PCC Task Group on Identity Management in NACO," *Cataloging & Classification Quarterly* 58, nos. 3-4, 430.

[72] Ibid., 432-434.

[73] Judith P. Cannan and Paul Frank, "Identity Management Principles to Rethink Traditional Authority Control" (presentation at 2020 LD4 Conference on Linked Data in Libraries, online, July 6-31, 2020).