**Adding Arabic Script to Voyager Records for Materials in the History of Medicine Division (HMD) collections.**

**Project leader:**

Salima M’seffar, Associate Fellow

August 2011

**Project sponsors:**

Laura Hartman, the History of Medicine Division

Ginny Roth, the History of Medicine Division

**Table of Contents**

Abstract …………………………………………………………………………………….…......3

Introduction4

Procedures4

Results6

Discussion7

Recommendations8

References……………………………………………………………………………………………………………………………….10

**Abstract**

OBJECTIVE:

To add vernacular Arabic script to Voyager bibliographic records for approximately 58 items in the HMD collections: 28 Arabic manuscripts and 30 modern Arabic posters. The objective of this project was also to use Transliterator software, test the outcome on Voyager bibliographic catalog and make recommendations to improve Transliterator software.

METHODS:

The project was accomplished in three phases:

* Becoming familiar with *ALA-LC Romanization Tables* [1]by correcting the 28 historical Arabic manuscripts records data that was entered decades ago.
* Create the required Romanized data for the Images at the History of Medicine, especially for 30 modern Arabic posters.
* Correct some Arabic posters’ MeSH terms.

RESULTS

For the 28 historical Arabic manuscripts, all the records needed corrections in order to be transliterated correctly. The ALA tables were used for that but this wasn’t enough since for some words there was no matching Romanized character within ALA tables. The records fields that were subject to corrections are the title, author and notes.

For the 30 Modern Public Health Posters, some of them were wrongly described. MeSH terms were added according to the image interpretation by the cataloger.

CONCLUSIONS

After correcting the existing records and editing the new Romanized Arabic data I concluded that in overall the ALA tables work well with Transliterator software except for AL Hamza that represents an issue when it comes on the line at the beginning, middle or the end of a word. To enter the word in the format that would be converted correctly with Transliterator will not be understood by Western users because vowels are omitted.

**Introduction**

The main objective of this project is to improve public access to and knowledge of historical Arabic language materials in the NLM collections to the Arabic speaking users, by making these items easy to locate on the web through the use of an Arabic keyboard.

The NLM collections targeted are the historical Arabic manuscripts and modern Arabic public health posters. The historical Arabic manuscripts, called also *Islamic Medical Manuscripts* available from <http://www.nlm.nih.gov/hmd/arabic/arabichome.html>. Most of these rare books were published between the 11th and 19th centuries. One of them is the oldest manuscript of the NLM: *al-juz al-thālith min kitāb al-Ḥāwī fī al-ṭibb* written by Abū Bakr Muḥammad ibn Zakarīyā al-Rāzī that in 30 November 1094.

The second collection is the Arabic public health posters. These posters are published by Middle Eastern public health institutions to communicate a message about a specific topic to the local populations. Some of them are related to Tobacco Control, others to family planning, hands hygiene, nutrition or breast feeding. The posters are also searchable on LocatorPlus. As it is for the historical Arabic manuscripts audience, these posters will be accessible to Arabic speaking users in the world. In addition to their bibliographic records, the posters themselves are digitized and available online for free in the database: *Images from the History of Medicine* from: <http://www.nlm.nih.gov/hmd/ihm/>. They can be also browsed by country or publisher.

Working on this project required:

* Learning the American Library Association (ALA) standards for transliterating Arabic script into Romanized data, capable of being searched by Western keyboards.
* Developing familiarity with using ExLibris Voyager cataloging software, MARC21 Unicode data, and Transliterator software for converting transliterated Arabic data into Arabic script. The availability of data in Arabic script opens the collection to a broader community in the Arab world. The data is able to be searched with both Western and Arabic keyboards.
* Searching the following NLM databases: LocatorPlus, NLMCatalog, and Images from the History of Medicine (IHM).

**Procedures**

First, I began with records for historical manuscripts because their Romanized Arabic data was entered decades ago and may not be up to the current ALA standards, especially for diacritics. The use of the Transliterator program reveals whether or not the Romanized data is correct. Correcting the data should provide a useful introduction to the ALA Romanization Tables.

These tables were created in order to help librarians as well as users in describing and locating Arabic literature at Western libraries, using either western or Arabic keyboards. Also these tables help in matching the Arabic alphabet to the Roman alphabet. At the same time I had to develop familiarity with using Voyager, cataloguing software, combined with Transliterator software. A request to access these systems on Citrix from my profile was made for me.

Laura Hartman, Rare Books cataloger at the History of Medicine Division, provided the training in the use of Transliterator software and gave a basic overview of inputting diacritical marks into Voyager and MARC Bibiographic, referring me to the reference work on the subject put out by the Library of Congress called “*Understanding MARC Bibliographic*” [2]. Since Laura is familiar with using the ALA Tables and Transliterator software for Cyrillic items, the first phase of working on historical Arabic manuscripts was accomplished in the HMD offices. That way HMD staff were nearby to answer my questions, enabled me to come back to the original manuscripts and read the hand writing of the authors.

Checking the data entered decades ago and comparing it to the original manuscripts revealed some misspelling mistakes of author names or titles. This gave my project a new perspective: it wasn’t just about entering the correct Romanized Arabic data but also correcting cataloguing mistakes. These errors would have given for a long time a wrong interpretation of the history of these manuscripts and their affiliation.

After completing the required Arabic manuscripts records and once familiar and comfortable with using the ALA Romanization Tables, it was easier to create the Romanized data required for the Images from the History of Medicine (IHM) records. Since IHM has never been tested for the proper display of vernacular Arabic data, working on IHM records needed to be conducted on the test server.

Ginny Roth, from the Images and Archives Section at The History of Medicine Division gave me a basic overview of image cataloging. She also printed the posters’ cataloguing records that needed edits and highlighted for me the fields to be added, edited or modified. These fields are:

245: Poster’s title to be entered in vernacular Arabic script

260: Poster’s author or institutional authority

500: General notes referring to where the title was taken from, and the meaning of it

710: The corporate name

995: last records update date

To work on the posters I preferred to do it at my desk since the posters were accessible online in digital format and it was easy to adjust the size of the images in order to read the texts. For that, I needed to access to the following software via Citrix:

* Voyager Cataloging Module (Group 2 security level)
* Test Voyager Cataloging Module
* MacroExpress
* Transliterator Software
* Arabic IME must also be installed on pc using Citrix in order for Transliterator to work properly

For both HMD collections, the outcome was tested in productions versions of LocatorPlus and NLMCatalog

**Results**

58 Voyager bibliographic records are upgraded with vernacular Arabic script, by adding or correcting existing Romanized data. At the beginning of the project, while I was getting familiar to ALA standards, upgrading one record took me about 3 to 4 hours. After being familiar, each record was done in less than 30 minutes, if there were no complications. Otherwise, it took 1 to 2 hours of testing and reediting the record on Transliterator software and Voyager.

These complications are related to Voyager not coordinating between some ALA vernacular Arabic data. So when the record is imported into Transliterator system, the Arabic script wasn’t correct: words’ letters were split into parts. Then it took longer than expected to find the tip or solution that would give coordination between the data entered in Voyager and its transliterated version through Tansliterator Software.

As an example of problems encountered while using Transliterator software:

For the record number 1455480: Romanizing the word  **حياة** wasn’t easy with the instructions given in ALA- LC Romanization tables for Arabic scripts concerning Tā’ marbūṭa

 **“ة”** as they come in page 13, section 7.

Spelling the word as recommended: “ḥayāh” and using Transliterator to Arabic, gave me حياه which is wrong and has a different meaning in Arabic. Transliterator doesn’t convert in that case the “ah” as Tā’ marbūṭa, but as “**He”** (letter).

In that case I just kept the Romanized spelling as “ḥayāh” and looked online for the Arabic translation of that word which means Life. I copied it and pasted it in the **880 Tag ‡6 245-01** (the title in Arabic) on Voyager exported record. This was the only way to deliver the data in a format that both Western and Arabic speaking users would discover the information.



**Discussion**

The results obtained were not really expected since LocatorPlus, NLM Catalog have tested successfully for the display of Arabic script. These errors in the display are related to the fact they were entered wrongly to Voyager records decades ago for the historical Arabic materials. The IHM posters have never been edited in vernacular Arabic, which eliminates misspelling errors but doesn’t mean that they are perfectly described. Some posters didn’t have the right MeSH terms, in some other cases, the title given by the cataloger doesn’t meet the meaning of the poster’s title. The cataloger attributes the title according to one’s interpretation of the picture. The picture can be interpreted in different ways. As the cataloger is not an Arabic speaker, she/ he can’t read the Arabic title or messages given in the poster.

 As an example: in the poster number 1450388:

The translated title was not given entirely as it came, and a part of it was missing. In addition to: “AIDS can happen to you”, I added the missing part: “protect yourself”.

Other issues came out of entering the Romanized script in capital letters, or small letters. This has different impact on the Arabic transliterated script. Example: **al-Īdz** transliterated as **الإيدز** doesn’t give the same result when the “Ī” comes in small letter.

In the record 1437875 and other similar record, there is publisher’s institutional authority hierarchy to respect while entering the field 260. Writing subtitles in Romanized Arabic script wasn’t obvious. In order to recover the order of the publishers: subdivision of the publisher in Transliterator, I had to start from “sub-publisher” followed by publisher because the Arabic read from right to left and Romanized script writes from left to right. Example: in order to obtain the correct order of publisher’s hierarchy:

**الوكالة المساعدة للطب الوقائي: وزارة الصحة: المملكة العربية السعودية**

**al-mamlakah al-ٰarabyah al-saٰūdyah: Wizārat al-Ṣiḥḥah: al-wikālah al-musāٰidah lilṭib al-wiqā’ī**

Then I noticed that the hierarchical order of publishers is reversed in Transliterator software imported record. But then when I exported to Voyager, the order as it comes in Transliterator, is reversed in voyager. Finally the initial order how it was entered in Voyager is recovered.

**Recommendations**

According to the results obtained and their discussion, I would highly recommend first to hire an Arabic speaking cataloger that would work on the NLM’s Arabic materials. If this is not possible, then teleworking or outsourcing the NLM’s Arabic materials records might be an option too. This is important since the Historical Arabic manuscripts records were wrongly described because only an Arabic speaker would know the difference between some letters that seem to be similar when they come Romanized but they are not in reality in the original language. Taking these letters for similar break the meaning of the titles and give a totally different interpretation of what the material is about.

Also having an Arabic speaking cataloger would help in describing correctly the Arabic public health posters for the IHM.

According also to my experience with these posters, it will be highly recommended to enter their Vernacular Arabic data by country of origin. In most cases, the posters coming from one country are published by the same institution or public health authority. Entering in vernacular Arabic script the name of this institution from scratch for each record is time consuming because of testing them in Transliterator software and the efforts of remembering successful tips and spelling used in first records. Working on the posters by country will allow the cataloger, once the correct Arabic script data found for the publishing institution, to copy paste this data in all the records belonging to the same authority and country. The cataloger is then sure that importing the data to Transliterator will give the correct conversion.

Finally, in order to resolve the different interpretation in Transliterator of capital and small letters, I would recommend that the NLM, the Library of congress and other libraries using ALA tables combined with Voyager software work together in editing additional standards for the impact of using capital / small letters on the Arabic script. Another alternative would be to enhance the capability of Voyager software in recognizing the letters no matter their format.

**References**

[1]: Barry R. K. ALA-LC Romanization Tables page [Internet]. Library of Congress; 1997. Available from: http://www.loc.gov/catdir/cpso/roman.html

[2]: Furrie B. **Understanding MARC Bibliographic: Machine-Readable Cataloging [Internet]. Washington, DC: Library of Congress; 2009. Available from: http://www.loc.gov/marc/umb/**