Multilingual Access to Arabic Digital Resources Using Linked Data Mechanisms and Challenges

Iman Weheba
Arabic Cataloging Librarian
Qatar National Library
iweheba@qf.org.qa

Middle East Librarians Association Annual Meeting
November 15-17, 2012
Denver, Colorado
This presentation aims at exploring three key issues:

1) The issues introduced by multilingual information access needs for print and digital resources in Arabic from a user perspective and the Bibliotheca Alexandrina’s experience in this regard.

2) Examining LC cataloging/metadata standards and their utility for providing multilingual access to Arabic print and digital resources.

3) Developing a multilingual linked-data model for linking bibliographic, authority and metadata records within and across different systems.
Linked Data Definition

Data which contains URI’s as identifiers for concepts described in the data and URIs to identify the relationships between those concepts

A richer linking mechanism for the web that takes us from hypertext links (document to document) to hyperdata links (across things that documents are about)...  

A term coined by Tim Berners-Lee
The Multilingual linked data can also use a cross-referenced URIs in the metadata records as the linking mechanisms for linking multilingual metadata across multiple library systems.

As an example, the library of Congress uses Permalink as a persistent identifier that can be used for linking to bibliographic and authority records from the Library of Congress catalog.

The next slide contains a snapshot of LC authority record with Permalink. The same model can be used to link from and to the multilingual metadata records in library catalogs. This will allow cross-lingual retrieval across multiple catalogs and digital library systems.
- Multilingual information access needs
- The User perspectives
- Bibliotheca Alexandrina’s Experience
Multilingual Access Needs

- Foreign language readers to judge the relevance of Arabic resources to their research needs
- Arab readers to search in Arabic and retrieve all the resources in Arabic and Latin script languages
Observations from Users’ queries:

Foreign Users
- I am looking for an author called “Thanuki” = Tanukhi = Abu al-Ala al-Maarri
- I am looking for materials in electronic form about the social history of Copts in Egypt

Arab Users
- I am looking for materials about “al-Khitab al-dini al-Islami” in English language
Bibliotheca Alexandrina’s Experience

Developing the BA Bi-script/Multilingual Name and Subject authority file (BIBALEX)

BIBALEX and VIAF

www.viaf.org

BIBALEX and the Multilingual LCSH Linked-Data Project (Authorities and Vocabularies)

http://id.loc.gov
Examining LC cataloging/metadata standards
The next screenshot for BIBALEX authority record which uses MARC 21’s “Model B” for multi-script records, where the author name is established in Arabic script using linking fields (7XX) to link the name to its equivalent authorized from in Latin script in the LC/NACO authority file.
<table>
<thead>
<tr>
<th>008</th>
<th>110614n</th>
<th>acnnnaabn</th>
<th>a aaa  d</th>
</tr>
</thead>
<tbody>
<tr>
<td>040</td>
<td>\a EG-AIBA \b ara \c EG-AIBA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>a محمد حمدا، محمد حمدان، 1936-2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>a محمد حمدا، محمد حمدان، 1936-2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>670</td>
<td>a محمد حمدا، محمد حمدان، 1936-2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>700</td>
<td>a Mihrāti, Muḥammad Ḥamdān, d 1936-2004 \v (DLC)no2011088941</td>
<td></td>
<td></td>
</tr>
<tr>
<td>999</td>
<td>a VIRTUA x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>060904i anznmbabn a ana d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>040</td>
<td>\a EG-AIBA \b ara \c EG-AIBA \f bibalex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>القانون الإداري (فقه إسلامي)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>الفقه الإسلامي</td>
<td></td>
<td></td>
</tr>
<tr>
<td>750</td>
<td>0 \a Administrative law (Islamic law) \0 (DLC)sh 88006404</td>
<td></td>
<td></td>
</tr>
<tr>
<td>750</td>
<td>7 \a Droit administratif (droit islamique) \2 ram</td>
<td></td>
<td></td>
</tr>
<tr>
<td>999</td>
<td>\a VIRTUA y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>001</td>
<td>ocn694829436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>003</td>
<td>OCoLC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>005</td>
<td>20110608012448.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>101208e2009 ku b 000 0 ara d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>040</td>
<td>\a LEILA \c LEILA \d EGAYA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>035</td>
<td>\a (OCoLC)694829436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>050</td>
<td>4 \a P17866.T33 \b ZS4 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>082</td>
<td>4 \a 892.716 \2 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>049</td>
<td>\a MAIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1 \6 880-01 \a Badi, Salim Nuaym Muhammed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>1 0 \6 880-02 \a Ta’ammulat fi shir’r D. Manir Safid al’Ubaynah fi al-fatrah min 1956 M-2001 M \c talif Salim Nuaym Muhammed al-Badi.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>\6 880-03 \a al-Tab’ah 1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>\6 880-04 \a Hawalli, Dawlat al-Kuwayt ; \a Madinat Nasr, al-Qahirah ; \b Maktabat al-Falah lil-Nashr wa-al-Tawzi’, \c 2009.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>\6 391 p. ; \c 25 cm.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>504</td>
<td>\a Includes bibliographical references (p. 391).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>600</td>
<td>1 0 \6 880-05 \a ‘Ubaynah, Manir Safid, \d 1946- \c Criticism and interpretation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>880</td>
<td>1 \6 02-245/8/6\a 2 ناملات في شعر د. مؤلف محمد الشيخ العنبية في الفترة من 1937-2002 م \c تأليف سالم محمد محمد الناعي.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>880</td>
<td>1 0 \6 03-250/8/6\a 2 الطبعة (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>880</td>
<td>\6 04-260/8/6\a 2 حولي، دوله الكويت ؛ \a مدينة مصر، القاهرة ؛ \a مكتبة الفلاح للنشر والتوزيع، \c 2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>880</td>
<td>1 7 bibalex</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
  <titleInfo altRepGroup="02" script="Latn">
    <title>Ta'ammulat fi shi'r D. Mání Sa'id al-Utaybah fl al-fatrah min 1956 M-2001 M</title>
  </titleInfo>
  <name type="personal" usage="primary" altRepGroup="01" script="Latn">
    <namePart>Ba'dli, Sālim Nujaym Muḥammad</namePart>
  </name>
  <typeOfResource>text</typeOfResource>
  <genre authority="marcgt">bibliography</genre>
  <originInfo script="Latn" altRepGroup="0304">
    <place><placeTerm type="code" authority="marccountry">ku</placeTerm></place>
    <place><placeTerm type="text">Hawalli, Dawlat al-Kuwayt</placeTerm></place>
    <place><placeTerm type="text">Madinat Naṣr, al-Qahirah</placeTerm></place>
    <publisher>Maktabat al-Falah lil-Nashr wa-al-Tawzi'</publisher>
    <dateIssued>2009</dateIssued>
    <edition>1⃣<edition>
    <issuance>monographic</issuance>
    <language><languageTerm authority="iso639-2b" type="code">ara</languageTerm></language>
    <physicalDescription><form authority="marcform">print</form>
    <extent>391 p. ; 25 cm.</extent></physicalDescription>
    <note type="statement of responsibility" altRepGroup="00" script="Latn">ta'līf Sālim Nujaym Muḥammad al-Ba'dli.</note>
</originInfo>
</mods>
<note type="bibliography">Includes bibliographical references (p. 391).</note>

<subject altRepGroup="05" script="Latn" authority="lcsh">
  <name type="personal">
    <namePart>ʻUtaybah, Māni’ Sa‘īd</namePart>
    <namePart>1946-</namePart>
  </name>
</subject>

<titleInfo>
  <title />
</titleInfo>

<topic>Criticism and interpretation</topic>

<classification authority="lcc">PJ7866.T33 Z54 2009</classification>
<classification authority="ddc" edition="22">892.716</classification>

<name type="personal" usage="primary" altRepGroup="01" script=""/>
  <namePart>بياد، سالم نجم محمد</namePart>
</name>

<titleInfo altRepGroup="02" script=""/>
  <title>تتوليدات في شعر د. سالم نجم الحديثة في الفترة من 1956 م. 1001 م</title>
</titleInfo>

<note type="statement of responsibility" altRepGroup="00" script=""/>
  <name type="personal"/>
  <namePart>беид، سالم نجم محمد</namePart>
  <namePart>1946-</namePart>
</name>

<titleInfo script="" />
  <title />
</titleInfo>

<titleInfo script="" />تتوليدات في شعر د. سالم نجم الحديثة في الفترة من 1956 م. 1001 م</titleInfo>

<identifier type="oclc">694829436</identifier>

<recordInfo>
  <descriptionStandard>aacr</descriptionStandard>
  <recordContentSource authority="marcorg">LEILA</recordContentSource>
  <recordCreationDate encoding="marc">101208</recordCreationDate>
  <recordChangeDate encoding="iso8601">20110608T1448.000Z</recordChangeDate>
  <recordIdentifier source="OCoLC">ocn694829436</recordIdentifier>
  <recordOrigin>Converted from MARCXML to MODS version 3.4 using MARC21slim2MODS3-4.xsl (Revision 1.70)</recordOrigin>
</recordInfo>

</mods>
<?xml version="1.0" encoding="UTF-8" ?>
xmlns:xlink="http://www.w3.org/1999/xlink" version="2.0" xsi:schemaLocation="http://www.loc.gov/mads/v2 http://www.loc.gov/standards/mads/mads-2-0.xsd">
  <mads:authority geographicSubdivision="not applicable">
    <mads:name type="personal">
      ﻣﺤﻤﺪان، ﻣﺤﻤﺪاﻟﻤﺼﺮاﺗﻲ،<mads:namePart>1936-2004</mads:namePart>
    </mads:name>
  </mads:authority>
  <mads:related type="equivalent">
    <mads:name type="personal" authority="naf">
      ﻣﺤﻤﺪان، ﻣﺤﻤﺪاﻟﻤﺼﺮاﺗﻲ،<mads:namePart>1936-2004</mads:namePart>
    </mads:name>
  </mads:related>
  <mads:variant type="other">
    <mads:name type="personal">
      ﻣﺤﻤﺪان، ﻣﺤﻤﺪاﻟﻤﺼﺮاﺗﻲ،<mads:namePart>1936-2004</mads:namePart>
    </mads:name>
  </mads:variant><mads:note type="source">اﻟﻤﺠﺪ، ﻣﺪارﺟ2010 : ص.140</mads:note>
</mads:mads>
<xml version="1.0" encoding="UTF-8" >
    <mads:authority geographicSubdivision="indirect">
      <mads:topic authority="bibalex">القانون الإداري (فقه إسلامي) </mads:topic>
    </mads:authority>
    <mads:related type="broader">
      <mads:related type="equivalent">
        <mads:topic authority="lcsh">Administrative law (Islamic law)</mads:topic>
      </mads:related>
      <mads:topic authority="ram">Droit administratif (droit islamique)</mads:topic>
    </mads:related>
    <mads:recordInfo>
      <mads:recordOrigin>Converted from MARCXML to MADS version 2.0 using MARC21slim2MODS-2-0.xsl (Revision 2.04)</mads:recordOrigin>
      <mads:recordContentSource authority="marcorg">EG-AIBA</mads:recordContentSource>
      <mads:languageOfCataloging>
        <mads:languageTerm authority="iso639-2b" type="code">ara</mads:languageTerm>
      </mads:languageOfCataloging>
    </mads:recordInfo>
  </mads:mads>
Figure 1. Conceptual equivalence illustrates some possible conceptual mismatches in the process of semantic mapping between two subject headings in different languages.

For example in Arabic we have sufficient literary warrants for:

- Koran--Seven Successive Readings
- Koran-- Three Completing Readings
- Koran-- Ten Readings
- Koran-- Fourteen Readings
- Koran-- Irregular Readings
- Koran-- Forged Readings
- Koran-- Interpolated Readings

The equivalent LCSH for all the above subjects is: Koran—Readings which is the broader term of the above subject headings.

**Solution suggested by PSD:**

130 00 Quran--Qirā’āt al-sab‘
730 00 Koran—Readings
730 04 Koran--Seven Successive Readings
The following subject heading illustrates the challenges faced in conveying the original concept in the process of mapping from the Arabic concept of a word to LCSH:

150  al-Ahkam al-sultaniyah
750 0 Public law (Islamic law)
The Arabic term “Rabb al-usrah” can be literally translated as “Family custodian”, which represents Fathers in LCSH. While the Arabic and English terms are not identical, they precisely represent the same concept.

150  Rabb al-usrah
750 0 Fathers

The LCSH term “Househusband” doesn’t have equivalent term in Arabic.
Developing Multilingual linked-data models
Multilingual linked data models

Generally developing **Multilingual linked data models** needs further research and investigations. However, before making a decision regarding the best cross-lingual data models, it is important to review the differences in concept and purpose between the following three categories of cataloging policies and standards available to libraries:
Multilingual Cataloging Policies and Standards

- **Monolingual, multi-script catalogues** like the LC's online catalog.

- **Twin monolingual, mono-script catalogues** are apparently still the dominant set-up in the Arab world.

- **Multilingual, bi-script catalogues** like the American University in Beirut (AUB) and the BA online catalogs.
Developing Multilingual Linked Data Models

- **Monolingual, multi-script catalogues** like the LC's online catalog. Possibility to be linked to multilingual subject headings and bibliographic records described in other languages.

- **Twin monolingual, mono-script catalogues** are apparently still the dominant set-up in the Arab world. Possibility to be linked to bibliographic and authority records in other scripts and languages.

- **Multilingual, bi-script catalogues** like the American University in Beirut (AUB) and the BA online catalogs. Possibility to be linked to bibliographic records described in other languages.
Some Proposed Models

**Manual Matching**
- **Related Description:**
  Linking bibliographic records described in different languages, example Arabic book described by BA and BNF. Needs to be developed.
- **Related Authorities:**
  Linking authority records in different languages and scripts. Already developed using 7XX tags in authority records.

**Automatic matching:**
Recommendations & Future Work

- Further research to be done through different MELA Committees according to their respective specializations.

- This research should continue and benefit from the results of the work of the Non-English Access Committee.

- There is a significant need to have a substantial contribution by system developers and reference librarians.
Collaboration and partnerships between libraries, especially collaboration with the organizations responsible for developing internationally-accepted standards such as the Library of Congress and OCLC.

There is a need for more focus on users and identifying their expectations and goals from cross-lingual retrieval of Arabic resources.

The initiation of an annual gathering within MELA to bring together researchers from across a broad range of disciplines to consider different aspects of multilingual access models and systems.
Conclusions

**Advantages of linked data**
- Building virtual collections
- Standards independent
- Richer user experience

**Challenges of linked data**
- Provenance, governance and sustainability issues
- Quality of interlinked data
- Developing a workable framework, systems and standards
- Needs robust infrastructure not available to every library
Questions?
Thank you