

مكتبة قطر الوطنية

QATAR NATIONAL LIBRARY

عضو في مؤسسة قطر Member of Qatar Foundation

# Multilingual Access to Digital Resources Using Linked-Data

## Overview of Tools and Standards

Iman Khairy Weheba  
Senior Cataloging Librarian  
Qatar National Library  
[iweheba@qf.org.qa](mailto:iweheba@qf.org.qa)

# TABLE OF CONTENT

---

01

## **Multilingual Access**

The User's  
Perspectives

02

## **Linked-Data**

Tools and  
Standards

03

## **Multilingual Portals**

Comparative  
Analysis

# Introduction

## Multilingual Access and Digital Libraries

With the growing significance of digital libraries and the Internet, more and more digital resources become accessible to a wide and geographically disperse public. The development of technologies that enable access to digital resources regardless of geographic or language barriers is a key factor for truly global sharing of knowledge. In an international milieu, digital libraries are composed of varied types of institutions from two or more countries with differences in the language, script and systems of metadata. The advantages of multilingual linked-data are:

- ✓ No institution needs to revise its rules and records to create a single, merged file of metadata to be used in the digital library.
- ✓ No need for human or automatic translation for translating the metadata.



# Multilingual Access Needs

## The User's Perspectives

### Multilingual Access Needs

- ✓ Foreign language readers to judge the relevance of Arabic resources to their research needs
- ✓ Middle Eastern languages readers to search in Arabic and retrieve all the resources in Arabic and foreign languages

### User's Perspectives:

#### ■ Foreign Users

- ✓ I am looking for materials in digital form about the social history of Copts in Egypt
- ✓ I am looking for an Arab classical author called al-Thanuki i.e al-Tanukhi a variant name for Abu al Ala al-Maarri

#### ■ Middle Eastern Users

- ✓ I am looking for materials about “al-Khitab al-dini al-Islami = Islamic religious discourse” in English language.



# Multilingual Access Needs

## The User's Perspectives

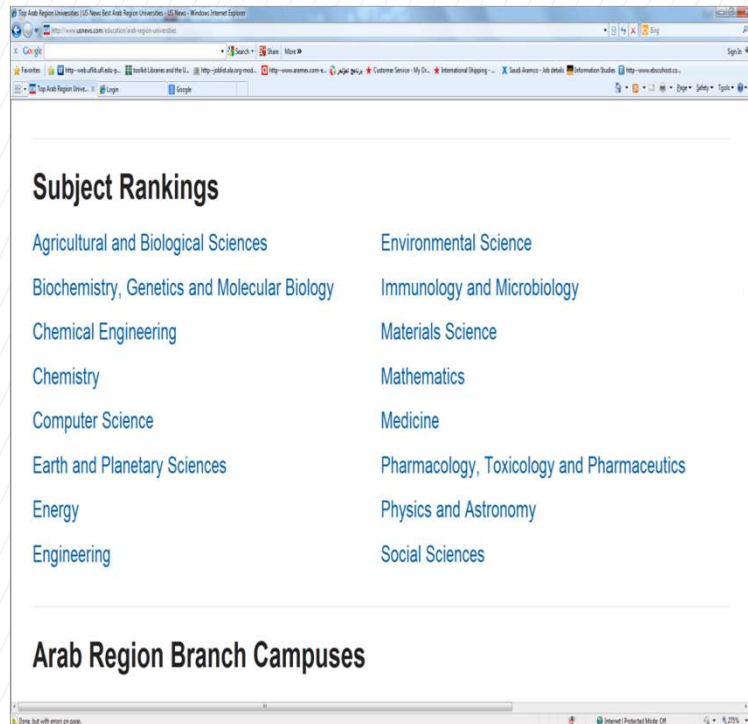
There is a need for more focus on users and identifying their expectations and goals from cross-lingual retrieval of digital and electronic resources, for example:

- ✓ Identifying how the users interact with discovery systems of multilingual digital libraries.
- ✓ What specifically are the disparities in needs and expectations when users specialize in social sciences and humanities vs. science and technology?
- ✓ What specifically are the disparities in needs and expectations when users speak different languages?



# Multilingual Access Needs

## Arabic Digital Scholarship



### Science vs. Humanities

- ✓ The US News, known for its rankings of the best colleges in the US, announced its first Arab region list in a session at the World Innovation Summit for Education at Doha, Qatar. The report is available at: <http://www.usnews.com/education/arab-region-universities>.
- ✓ It was noted that highest subject rankings are for scientific fields with social sciences coming as the last subject in the rankings. Is there a language factor in these rankings? More research is needed to answer this question



# Multilingual Access Needs

## Arabic Digital Scholarship

### ✓ Science vs. Humanities:

Al-Aufi, Ali. 2010. "An investigation of digital scholarship and disciplinary culture in Oman". *Library Hi Tech*. 28 (3): 414-432.

In the above cited article, Al-Aufi conducted a research using a survey at Sultan Qaboos University (SQU) in Oman. The primary objectives of the research are to investigate the use of networked electronic resources and its impact on patterns of research and scholarly communication in a developing Arab country, using SQU as an example; and identify the disciplinary differences reflected in the use of networked information for research and scholarly communication.



# Multilingual Access Needs

## Arabic Digital Scholarship

### ✓ Science vs. Humanities:

Al-Aufi, Ali. 2010. "An investigation of digital scholarship and disciplinary culture in Oman". *Library Hi Tech*. 28 (3): 414-432.

Al-Aufi's survey highlighted the following two important linguistic research factors:

- ✓ The answers regarding all scholarly communication activities indicated that respondents from science disciplines have more positive responses than respondents from humanities and social science disciplines.
- ✓ The answers to questions about the respondents' perception of Arabic as a scholarly language revealed that respondents in the humanities and social science disciplines have more positive responses than respondents from the science disciplines.





# Multilingual Access Needs

## Arabic Digital Scholarship

To conclude about the above, Middle Eastern libraries and database vendors in the Arab world and internationally need to develop strategies for the following:

- ✓ Provide scholars—particularly those working in the humanities—with multilingual access system that support them in compiling and accessing digital scholarly publications
- ✓ Assist in using networking technologies to build and sustain research communities for these same disciplines in a multilingual environment
- ✓ Develop multilingual access system to fully indexed Arabic scholarly publications along with complete Arabic citation index.



# Linked-Data

---

## Tools and Standards



# Linked-Data

## Overview

Linked Data is a methodology for providing **relationships** between things (**data, concepts and documents**) anywhere on the web, using: **URI's** for identifying, **RDF/XML** for describing, **HTTP** for publishing, these things and relationships in a way that they can be interpreted and used by humans and software.



# Linked-Metadata Overview



## URIs (LC Permalink)

- ✓ The “linked-metadata” can also use a referenced URIs in the metadata records as the linking mechanism for linking metadata elements (names, subject or bibliographic) within or across multiple systems.
- ✓ For example LC Permalink can be used for linking to LC bibliographic and authority records



# Linked-Metadata

## Opportunities and Challenges

### Opportunities:

- ✓ Building virtual collections
- ✓ Standards independent
- ✓ Richer user experience

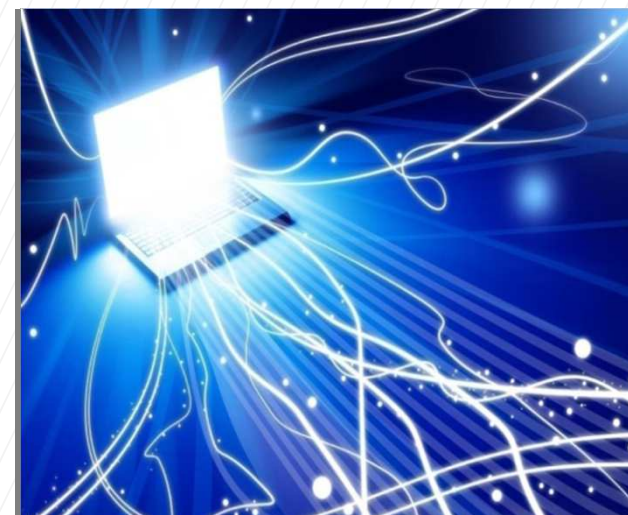
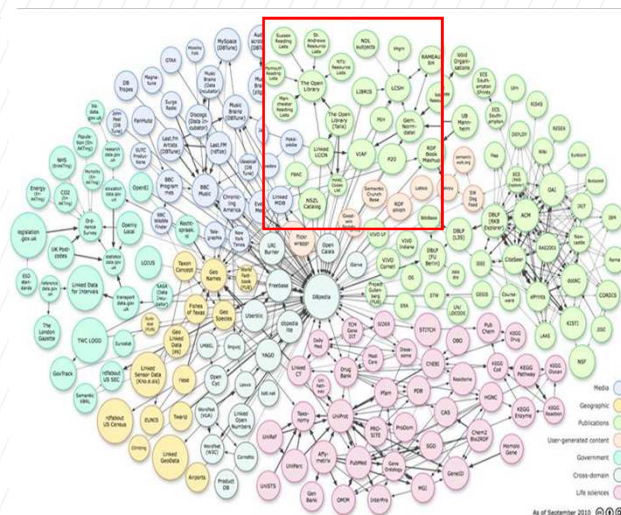
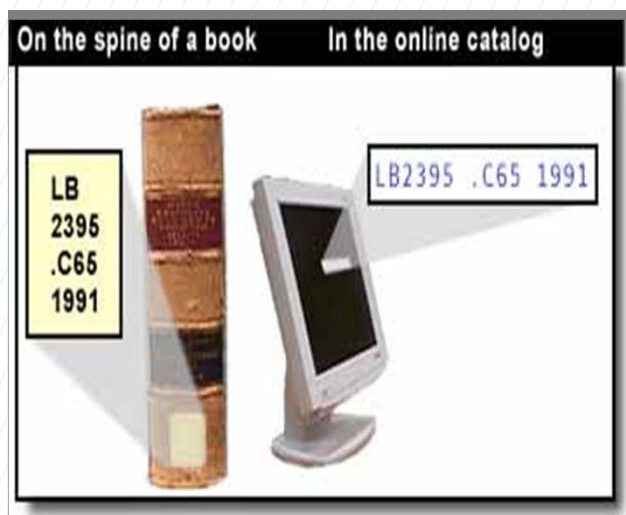
### Challenges:

- ✓ Provenance, governance and sustainability issues
- ✓ Quality of interlinked data
- ✓ Developing a workable framework, systems and standards
- ✓ Needs robust infrastructure not available to every library



# New Ways of Cataloging

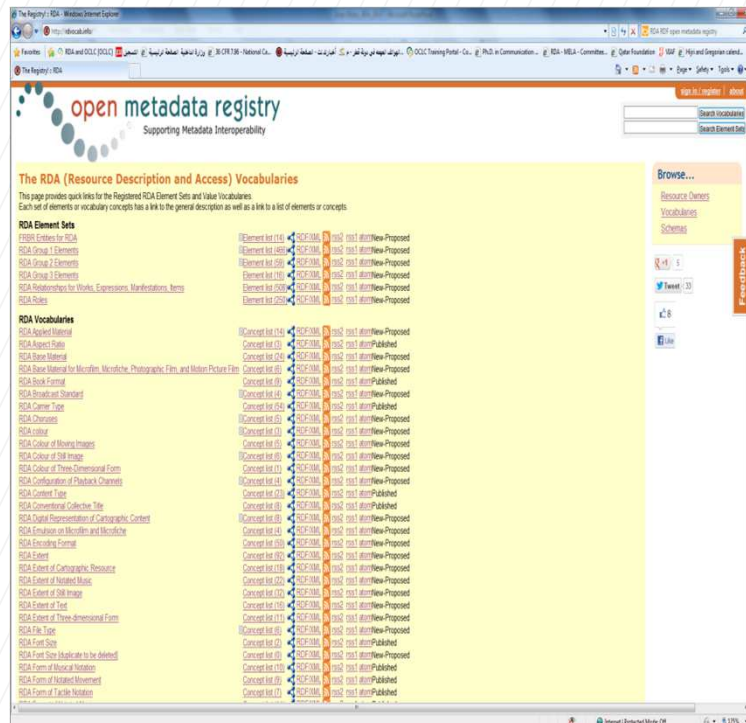
## RDA and BIBFRAME Transition



RDA/BIBFRAME goals is to move catalog data to be part of the Semantic Web, not only on the Web and support for linked-data standards

# AACR2 to RDA Transition

## Web Linked Metadata



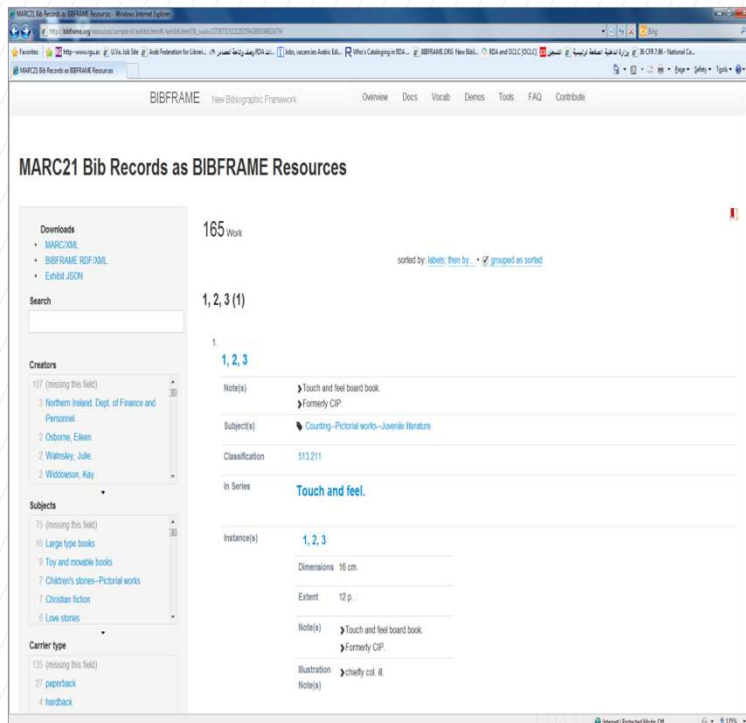
## RDA as RDF/XML Vocabulary:

- ✓ Data on the Semantic Web is modeled on the Resource Description Framework (RDF) that breaks data down into sets of triples, with each triple consisting of a subject, predicate, and object.
- ✓ RDA entities are now defined as RDF vocabularies on the Open Metadata Registry, which brings library data a giant step closer to the Semantic Web.



# MARC 21 to BIBFRAME Transition

## Web Linked Metadata



## BIBFRAME as RDF/XML:

- ✓ BIBFRAME is an undertaking by the Library of Congress and Zephiera Company for MARC 21 transition to a more semantic web enabled linked data standard.





# International Linked-Data Projects

## Worldcat.org Linked Bibliographic Data



### Project Overview:

- ✓ WorldCat.org now offers the largest set of linked bibliographic data on the Web. With the addition of Schema.org mark-up to all book, journal and other bibliographic resources in WorldCat.org.
- ✓ WorldCat.org is now available for use by intelligent Web crawlers, like Google, that can make use of this metadata in search indexes and other applications.



# International Linked-Data Projects

## VIAF Linked-Authorities

VIAF  
Virtual International Authority File

Search

Select Field: Personal Names    Select Index: Egypt    Search Terms: ديكنز تشارلز    Search

1 heading found for *ديكنز، تشارلز (in egaxa)*

Heading	Type	Sample Title
1 <a href="#">Dickens, Charles, 1812-1870</a>	Personal	Oliver Twist
<a href="#">Dickens, Charles (Charles John Huffam), 1812-1870</a>		Oliver Twist
<a href="#">Диккенс, Чарлз, 1812-1870</a>		Oliver Twist
<a href="#">ديكنز، تشارلز، 1812-1870</a>		Oliver Twist
<a href="#">Dickens, Charles English novelist and draftsman, 1812-1870</a>		The Pickwick papers.
<a href="#">Dickens, Charles</a>		David Copperfield, 2002:
<a href="#">Dickens, Charles John Huffam, 1812-1870</a>		

## Project Overview:

- ✓ Virtual International Authority File (VIAF) is a single view of merged authority records from the participating institutions in VIAF.
- ✓ OCLC Research developed an automatic matching algorithm based on data in authority and bibliographic records from the participating institutions



# How to approach Linked Data

## Background Materials

### ✓ Books:

Harper, Corey A. 2012. *Linked data for libraries, archives and museums*. [Gaithersburg, Md.]: National Information Standards Organization.

Hooland, Seth van, and Ruben Verborgh. 2014. *Linked data for libraries, archives and museums: how to clean, link and publish your metadata*. London : Facet Publishing.



# How to approach Linked Data

## Background Materials

### ✓ Webinars:

Webinars on Linked Data are offered at the FAO of the UN at: <http://aims.fao.org/linked-open-data-webinars-at-aims> There is one in Arabic language. There is a metadata recommendation at FAO, called LODE-BD (How to select appropriate encoding strategies for producing Linked Open Data (LOD)-enabled bibliographic data). The recommendation is available at: <http://aims.fao.org/lode/bd> and another Webinar on LODE-BD is also available at: <http://www.youtube.com/watch?v=6yNPcJXJj0U>



# How to approach Linked Data

## Background Materials

### ✓ Websites:

European libraries use of Linked Open Data (LOD):

<http://www.theeuropeanlibrary.org/tel4/newsitem/5802>

USA Project LD4L (Linked-Data for Libraries):

<https://wiki.duraspace.org/display/ld4l/LD4L+Use+Cases>

USA Bibframe testing and implementation projects:

<http://hangingtogether.org/?p=4487>

Authority files and linked data (different approaches)

<http://www.uebertext.org/2014/07/name-authority-files-linked-data.html>



# How to approach Linked Data

## Background Materials

### ✓ Websites:

A relevant multilingual access project from Japan that allows users to cross-lingual search. They use name authorities to connect other linked-data resources. Full paper is available at the Dublin Core Metadata Initiative (DCMI) Website:

<http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/150>

and the PPT at:

<http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/150/129>

Other useful presentations from the DCMI Conference:

<http://dcevents.dublincore.org/index.php/IntConf/dc-2013/schedConf/presentations>



# How to approach Linked Data

## Background Materials

### ✓ Websites:

Multilingual LCSH Linked Data Project:

<http://id.loc.gov>

Getty Vocabularies Translations and Linked Open Data:

<http://www.getty.edu/research/tools/vocabularies/training.html>



# How to approach Linked Data

## 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.





# How to approach Linked Data

## 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 BIBALEX online catalogs. Possibility to be linked to bibliographic records described in other languages.



# How to approach Linked Data

## Categories of Linking

### ✓ **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 mapping:**

Developing standards for automatic matching. **Needs further research and investigations.**



# How to approach Linked Data

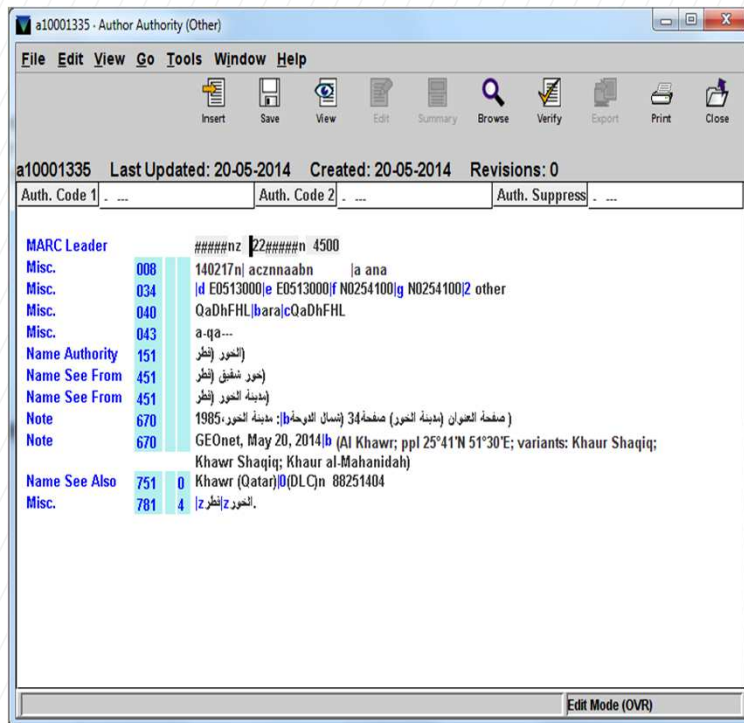
## Opportunities and Challenges

- ✓ **The metadata records do not yet exist** investigating linking to available multilingual bibliographic records from (WorldCat.org) which are available as linked-data and also incorporate multilingual authority data from the Virtual International Authority File (VIAF.org) which are also available as MARC 21 data in XML.
- ✓ **The Metadata Records only represent a half standard (description standard only such as DC, EAD, and mainly all other non-MARC metadata schemas)** developing an XML-based schema or document type definition (DTD) for mapping data from available MARC 21 authority records from the participating institutions local systems or provide links to VIAF.



# How to approach Linked Data

## Developing QNL Bilingual Linked-Data



## Project Overview

✓ We can start by the following simple things:

1) Adding the equivalent name in English in the linking fields 7XX in the authority records along with the LCCN number in subfield 0.

2) Making QNL name and subject authority published as linked data, like LC

<http://id.loc.gov/>



# Multilingual Metadata

## Metadata Authority Description Standard (MADS)

```
<?xml version="1.0" encoding="UTF-8" ?>
- <mads:mads xmlns:mads="http://www.loc.gov/mads/v2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
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="indirect">
  <mads:topic authority="bibalex">القانون الإداري (فقه إسلامي)</mads:topic>
  </mads:authority>
- <mads:related type="broader">
  <mads:topic>الفقه الإسلامي</mads:topic>
  </mads:related>
- <mads:related type="equivalent">
  <mads:topic authority="lcsb">Administrative law (Islamic law)</mads:topic>
  </mads:related>
- <mads:related type="equivalent">
  <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>
```



# Arabic Translation Tools

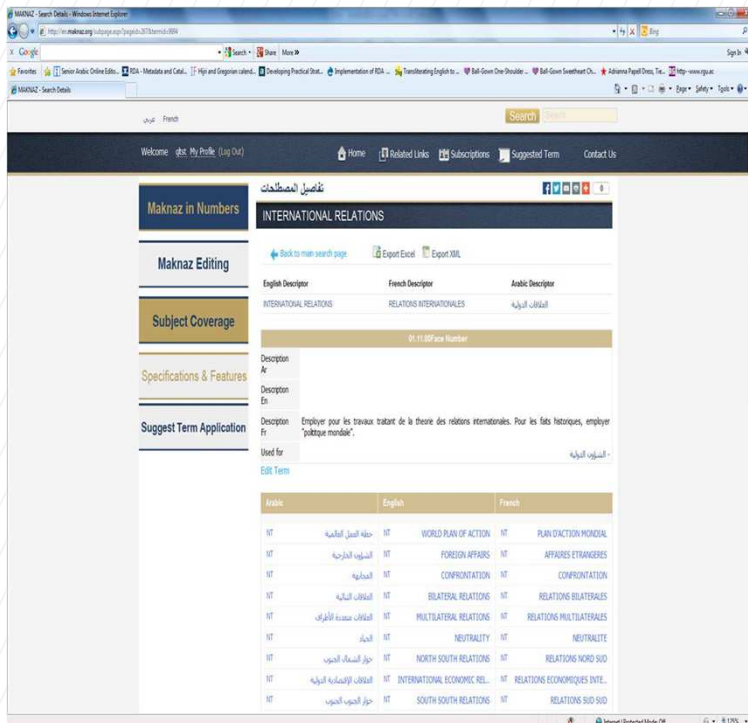
## Authorities

- ✓ **Bibliotheca Alexandrina Authority File (BIBALEX)**  
[http://www.bibalex.org/bacatalog/catalog\\_en.aspx?skin=default&lng=en](http://www.bibalex.org/bacatalog/catalog_en.aspx?skin=default&lng=en)
- ✓ **Virtual International Authority File**  
[www.viaf.org](http://www.viaf.org)
- ✓ **GEONet Names** (Arabic script geographic names are also available)  
<http://earth-info.nga.mil/gns/html/>



# Arabic Translation Tools

## Subject Heading



## Expanded Thesaurus:

- ✓ The Expanded Thesaurus ([www.maknaz.org](http://www.maknaz.org)) produced by Abdul Hameed Shoman Foundation, Dubai Municipality, Juma Al-Majed Centre for Culture & Heritage.
- ✓ It is a trilingual thesaurus with subject terms in Arabic, English and French

# Arabic Translation Tools

## Subject Headings Subdivisions

- ✓ **Arabic Translation of the LCSH Free-Floating Subdivisions**

[http://www.aruc.org/fckeditor/editor/filemanager/connectors/aspx/userfiles/file/501\\_Free-Floating%20Subdivisions.pdf](http://www.aruc.org/fckeditor/editor/filemanager/connectors/aspx/userfiles/file/501_Free-Floating%20Subdivisions.pdf)

- ✓ **Arabic Translation of the LCSH Chronological Subdivisions for Middle East Countries**

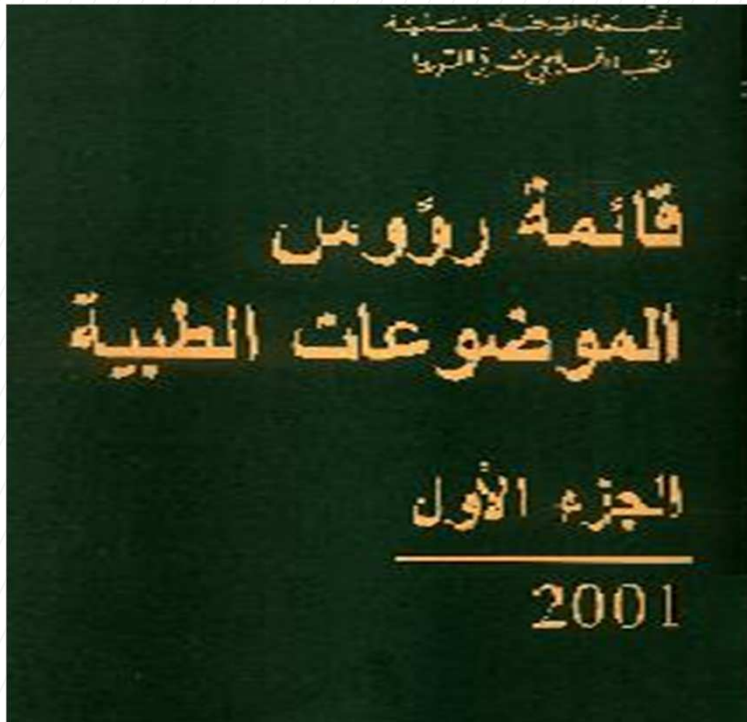
[http://www.aruc.org/fckeditor/editor/filemanager/connectors/aspx/userfiles/file/501\\_Free-Floating%20Subdivisions.pdf](http://www.aruc.org/fckeditor/editor/filemanager/connectors/aspx/userfiles/file/501_Free-Floating%20Subdivisions.pdf)





# Arabic Translation Tools

## Subject Heading



### Arabic MeSH:

- ✓ Qaimat ruus al-mawduat al-tibbiyah = Medical subject headings (MeSH), National Library of Medicine, translated into Arabic by the World Health Organization, Regional Office for the Eastern Mediterranean.
- ✓ It is very useful for translating medical and scientific terms such as DNA, mad cow disease, etc.

# System Tools

## Web-Based, Open Source Systems

- ✓ General digital repository management software that supports the description of PDFs, websites, audio, video, etc. Supports main metadata schemas such as MARCXML, MODS, and DC.

<http://www.dspace.org/>

- ✓ Archives information management software. Supports the Encoded Archival Description EAD finding aids metadata for describing archival materials, manuscripts and digital objects such as photo albums, etc.

<http://archivesspace.org/>

- ✓ Museums collection management software. Support template based metadata records that covers museum subjects from Art to Zoology, also includes authority control feature for persons, places, concepts, etc.

<http://www.collectionspace.org/>



# Multilingual Portals

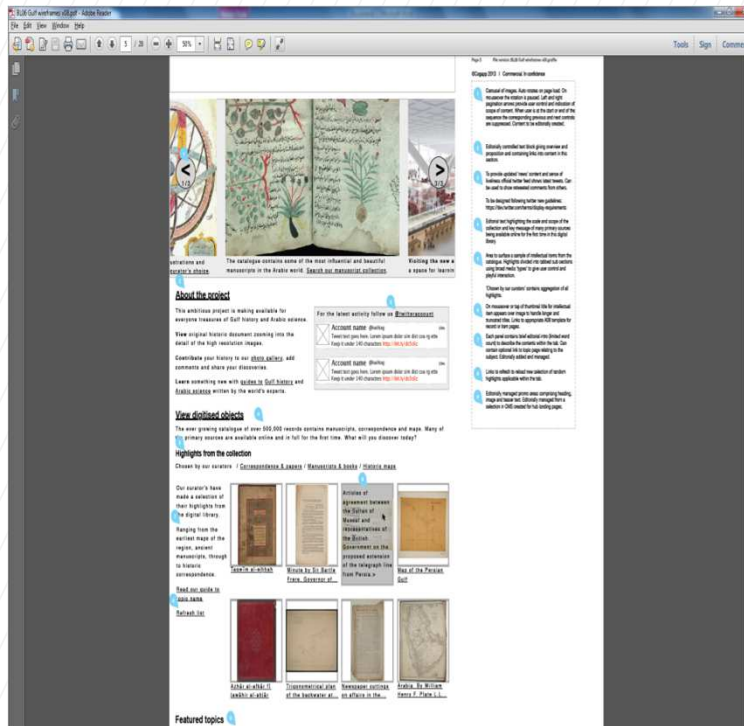
---



# Multilingual Portals

## Qatar Digital Library

### Gulf History and Arabic Science Portal



## Project Overview

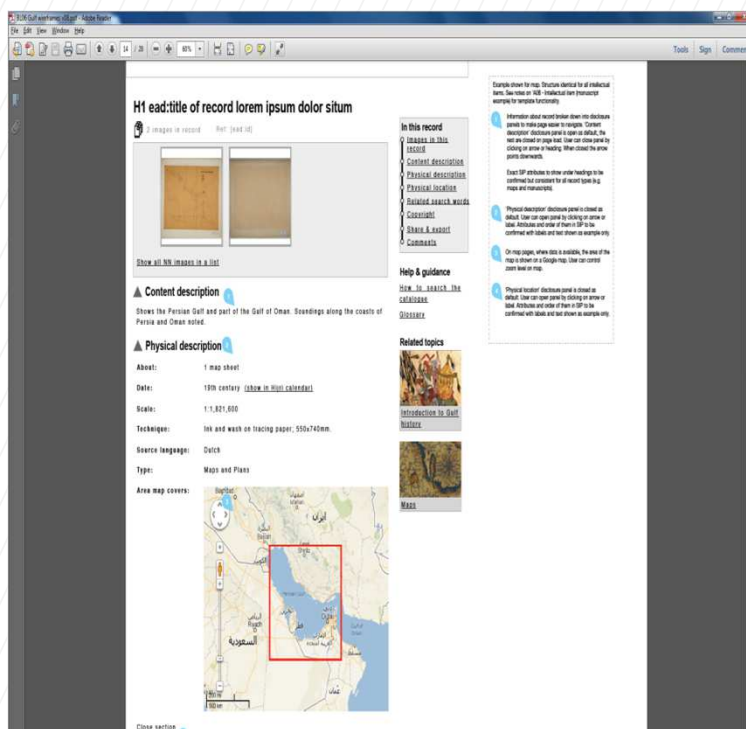
- ✓ This portal ([www.qdl.qa](http://www.qdl.qa)) provides bilingual (Arabic and English) metadata and contextual documents for all the digitized resources (500000 digital materials).
- ✓ It mainly depends on human translations of the metadata and contextual documents with some use of machine translations and terms mappings whenever applicable.



# Multilingual Portals

## Qatar Digital Library

### Gulf History and Arabic Science Portal

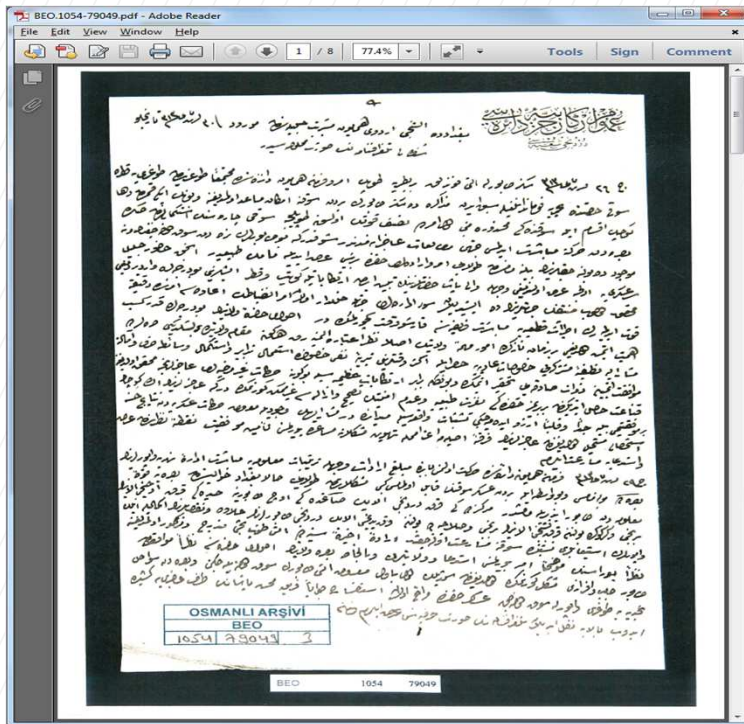


## Project Overview

- ✓ The next phases of the project intend to explore the use of linked-data as a third method of bilingual access, in addition to human and machine translations.
- ✓ The project will be paralleled by the development of QNL bilingual authority files for all Arabic name and subject headings.

# Multilingual Portals

## Ottoman Archives



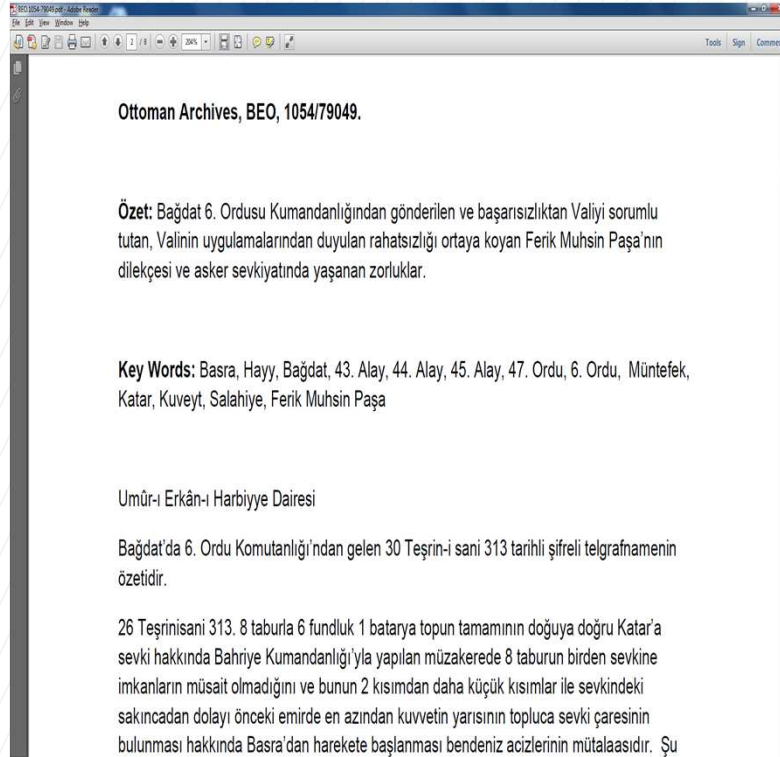
## Project Overview

- ✓ The Ottoman Archives are a collection of historical sources related to the history of the Arab region in general and the Gulf region in particular.
- ✓ It is intended to include digitized images of the original Ottoman document along with trilingual descriptions in Turkish, Arabic, and English.



# Multilingual Portals

## Ottoman Archive



## Turkish Descriptions:

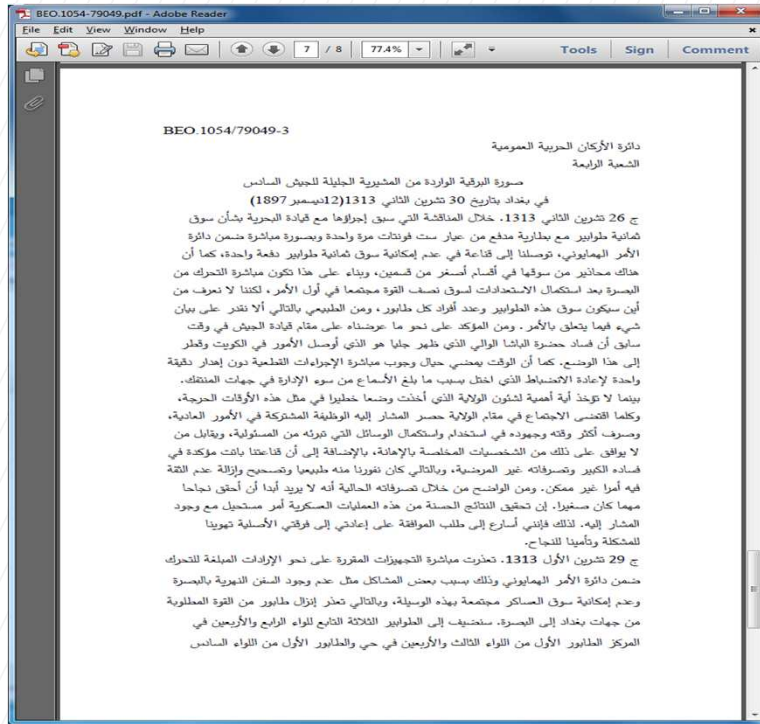
- ✓ The image on the left is the description of the original Ottoman document in modern Turkish.

# Multilingual Portals

## Ottoman Archives

### Arabic Descriptions:

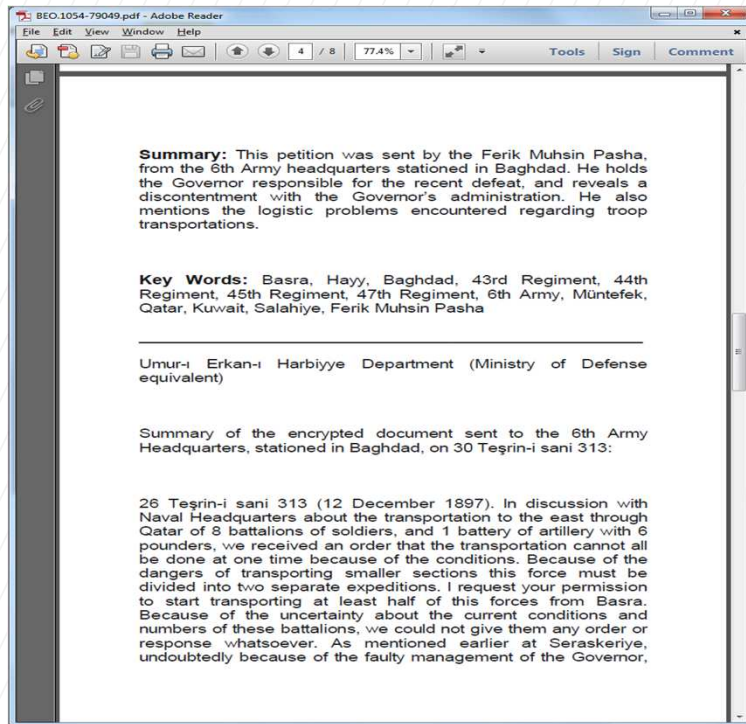
- ✓ The image on the left is the description of the original Ottoman document in Arabic.





# Multilingual Portals

## Ottoman Archives



### English Descriptions:

- ✓ The image on the left is the description of the original Ottoman document in English.

# Multilingual Portals

## Ottoman Archives

	A	B	C
1	Title (Turkish)	Bağdat' da 6. Ordu Komutanlığı'ndan gelen 30 Teğrin-i sani 1313 tarihli şifreli telgrafnamenin özetidir.	245
2	Translated Title (English)	Summary of the encrypted document sent to the 6th Army Headquarters, stationed in Baghdad, on 30 Teğrin-i sani 1313.	242
3	Translated Title (Arabic)	242صورة الوثيقة الواردة من المشورة الخليفة تقي الدين في بغداد بتاريخ 30 تشرين الثاني 1213 ديسمبر 1897.	
4	Personal Name (Repeatable)	N/A	100
5	Corporate Name (Repeatable)	Umür-i Erkan-ı Harbiyye Dairesi	110
6	Family Name (Repeatable)	N/A	100
7	Document number	BEO, 1054/79049.	9xx
8	Dates	12 December 1897	264 subfield c
9	Summary (Turkish)	Bağdat 6. Ordusu Kumandanlığı'ndan gönderilen ve başarısızlıktan Valiyi sorumlu tutan, Valinin uygulamalarından duyulan rahatsızlığı ortaya koyan Ferik Muhsin Paşa'nın dilekçesi ve asker sevkياتında yaşanan zorluklar.	520
10	Summary (English)	This petition was sent by the Ferik Muhsin Pasha, from the 6th Army headquarters stationed in Baghdad. He holds the Governor responsible for the recent defeat, and reveals a discontentment with the Governor's administration. He also mentions the logistic problems encountered regarding troop transportations.	520
11	Keywords (Turkish)	Basra, Havy, Bağdat, 43. Alay, 44. Alay, 45. Alay, 47. Ordu, 6. Ordu, Müntefek, Katar, Kuveyt, Salahiye, Ferik Muhsin Paşa	653
12	Keywords (English)	Basra, Havy, Baghdad, 43rd Regiment, 44th Regiment, 45th Regiment, 47th Regiment, 6th Army, Müntefek, Qatar, Kuwait, Salahiye, Ferik Muhsin Pasha	653
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			

## Metadata Mappings:

- ✓ The image on the left is the metadata mapping of the trilingual (Turkish, Arabic and English) metadata to the relevant MARC 21 fields.



# Recommendations

- ✓ Developing national Arabic authority files by each national library in the Arab world and link them to VIAF (Bibliotheca Alexandrina and the National Library of Lebanon are examples)
- ✓ The initiation of international committee that can bring together researchers from across a broad range of disciplines to consider different aspects of developing a multilingual access model relevant to Arabic.



# Conclusion

While current digital library systems fully support the storage and management of multilingual metadata, they are not capable of querying across different languages. The sought solution was to develop a linked-data model that support Cross-Lingual Information Retrieval (CLIR) of documents written and cataloged in different languages, particularly cross-lingual retrieval of names and subject headings.

This presentation briefly addressed this lacuna and highlighted its technical particularities with reference to some relevant projects and background materials. Finally, it suggested collaboration between interested libraries for this major aspect of digital humanities that will help researchers find all linked digital resources.



# THANK YOU

---

Website: [www.qnl.qa](http://www.qnl.qa)  
Phone: +974 4454 6039 | PO Box:5825, Doha – Qatar

**QATAR NATIONAL LIBRARY**