

# **The Significance of Transliteration in Enhancing Accessibility and Exchange of Arabic Research Data**

**Mohamed El-Zalabany M.A.**

Cataloging Librarian - Library of Congress – Cairo Office  
Ph.D Candidate, Information Studies- Cairo University

[mh\\_zalabany@hotmail.com](mailto:mh_zalabany@hotmail.com)

[www.linkedin.com/in/mohamedelzalabany](http://www.linkedin.com/in/mohamedelzalabany)

# Research Data on Arabic Content

Research data is: *“the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.”*

Research data on Arabic content, heritage, and culture is essential for global research, offering unique insights that enrich our understanding of societal dynamics, historical evolution, and cultural diversity, thereby promoting international collaboration and understanding.





# Towards FAIR Arabic Research Data

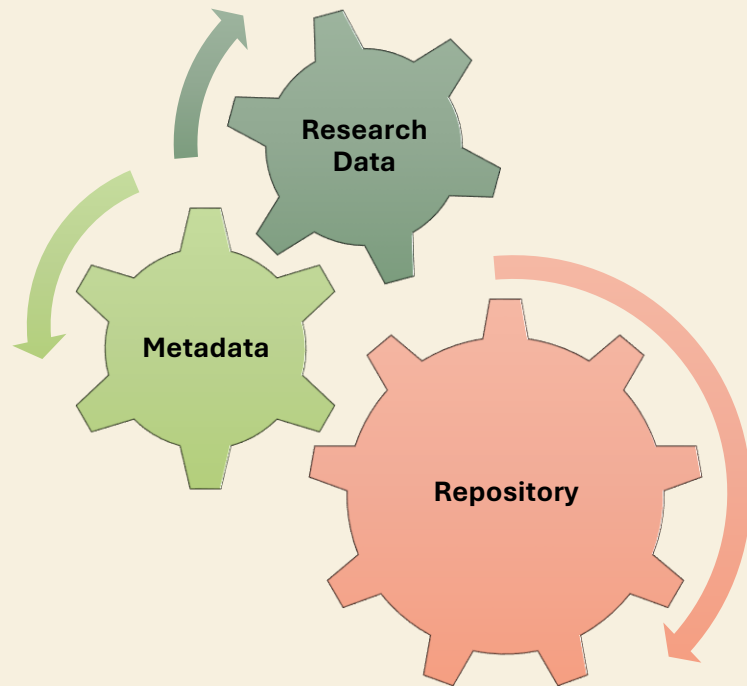
The FAIR principles ensure that data is Findable, Accessible, Interoperable, and Reusable, promoting open and efficient sharing of digital research outputs.

Arabic Open Research Data has a very important role in Digital Humanities.



# Metadata Role in RD Stewardship

Effective research data stewardship requires robust depositing practices and detailed metadata management to ensure data is both discoverable and usable for future research.



# Challenges of Arabic RD Metadata

- **Original Description Preservation:** It's essential to keep original Arabic descriptions to maintain cultural and historical accuracy, but this can be complex due to dialect variations and the intricacies of the script.
- **Metadata Translation:** Translating Arabic metadata into other languages poses risks of misinterpretation or oversimplification, potentially leading to critical information loss.
- **Global Accessibility:** Balancing local specifics with global understandability in metadata is challenging, requiring adjustments to ensure clarity for an international audience.
- **Technical Integration:** Incorporating Arabic metadata into predominantly Latin-script systems involves overcoming right-to-left text orientation and script cursive nature.
- **Metadata Standards Harmonization:** Developing standards that accommodate Arabic specifics while aligning with international norms is complex, especially with differing transliteration schemes and cataloging practices.



# Multi-lingual and transliterated Metadata

## Essential for discovery

- valuable research is often overlooked when published in non-English languages. It can also be helpful for researchers trying to determine whether it is worth their time to request a translation or use a machine translator to read an article, especially for fields and topics where research is scarce.
- It can be frustrating for readers to find translated metadata only to find that the actual content is not available in that language.
- It can also cause your content to be indexed incorrectly or not at all, especially by indexing and discovery services that have strict expectations for metadata and language consistency.



# Metadata Transl(ite)r(ation) Three Challenges



Standards



Systems



Automation

# Research Data Metadata standards



- Metadata standards
- Practices for multi-lingual description.

## MARC 21 record

```
=LDR 00000nam 2200000 4500
=001 123456789
=003 DLC
=005 20230505102030.0
=008 230505s2023 xxu|||||o ||0 eng d
=040 \\$aDLC$beng$erda$cDLC$epn$dDLC
=041 0\\$aeng$bara
=100 1\ $6 880-01$a al-Şalfy, Şadām,$d1975-
=880 1\\$6100-01$a صدام الصلفي,$d1975-
=245 10 $6 880-02 $a Tārīkh wa-huwīyat Jāmi‘at al-Malikah Arwá ‘alá
al-intarnit wa-marāhil al-taṭwīr mundhu al-‘āmm 1998-2021 M / $c
Şadām al-Şalfy.
=880 10$6245-02$a تاريخ وهوية جامعة الملكة أروى على الإنترنت ومراحل
صدام الصلفي / التطوير منذ العام 1998 - 2021 م
=264 \1$a[Place of publication not identified] :$bHarvard Dataverse,
$c2023.
=300 \\$a1 online resource.
=336 \\$atext$btxt$2rdacontent
=337 \\$acomputer$bc$2rdamedia
=338 \\$aonline resource$bcr$2rdacarrier
=500 \\$aTitle from resource description page (viewed May 5, 2023).
=520 \\$aThis dataset documents the development and identity of
Queen Arwa University on the Internet from its first website in 1998
through to 2021, marking significant milestones in digital presence
and educational impact.
=546 \\$aText in English and Arabic.
=650 \0$aEducation$xEffect of technological innovations on.
=650 \0$aWeb archives$zYemen.
=651 \0$aYemen$xHistory.
=856 40$uhttps://doi.org/10.7910/DVN/PISAJ5$zAccess the dataset
here.
```

## DATAcite record

```
<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://datacite.org/schema/kernel-4"
  xsi:schemaLocation="http://datacite.org/schema/kernel-4
http://schema.datacite.org/meta/kernel-4/metadata.xsd">
  <identifier identifierType="DOI">10.7910/DVN/PISAJ5</identifier>
  <creators>
    <creator>
      <creatorName nameType="Personal">al-Şalfy, Şadām</
creatorName>
      <creatorName nameType="Personal" xml:lang="ar">الصلفي،
صدام</creatorName>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en"> Tārīkh wa-huwīyat Jāmi‘at al-Malikah
Arwá ‘alá al-intarnit wa-marāḥil al-taṭwīr mundhu al-‘āmm 1998-2021 M
</title>
    <title xml:lang="ar">تاريخ وهوية جامعة الملكة أروى على الإنترنت
ومراحل التطوير منذ العام 1998 - 2021 م
</title>
  </titles>
  <publisher>Harvard Dataverse</publisher>
  <publicationYear>2023</publicationYear>
  <resourceType resourceTypeGeneral="Dataset">Dataset</
resourceType>
  <descriptions>
    <description xml:lang="en" descriptionType="Abstract">
      Queen Arwa University was established and began its
educational activities at the beginning of 1996 M
    </description>
    <description xml:lang="ar" descriptionType="Abstract">
      تأسست جامعة الملكة أروى ومارست نشاطها التعليمي منذ مطلع
العام 1996 م واستقبلت
    </description>
  </descriptions>
  <rightsList>
    <rights rightsURI="http://creativecommons.org/licenses/by/
4.0/">CC BY 4.0</rights>
  </rightsList>
  <geoLocations>
    <geoLocation>
      <geoLocationPlace>Yemen</geoLocationPlace>
      <geoLocationPlace xml:lang="ar">اليمن</geoLocationPlace>
    </geoLocation>
  </geoLocations>
</resource>
```

## DDI record

```
<DDIInstance xmlns="http://www.ddialliance.org/namespace/
ddi:instance:3_2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.ddialliance.org/
namespace/ddi:instance:3_2 http://www.ddialliance.org/namespace/
ddi:instance:3_2/ddi-instance.xsd">
  <ResourcePackage>
    <StudyUnit>
      <Citation>
        <Title>
          <p xml:lang="en"> Tārīkh wa-huwīyat Jāmi‘at al-
Malikah Arwá ‘alá al-intarnit wa-marāḥil al-taṭwīr mundhu al-‘āmm
1998-2021 M</p>
          <p xml:lang="ar"> تاريخ وهوية جامعة الملكة أروى
على الإنترنت ومراحل التطوير منذ العام 1998 - 2021 م</p>
        </Title>
        <IDNo agency="Harvard Dataverse">10.7910/DVN/PISAJ5</
IDNo>
        <Creator>
          <p xml:lang="en">al-Şalfy, Şadām</p>
          <p xml:lang="ar"> صدام، الملفي</p>
        </Creator>
        <ProductionDate>2023-02-03</ProductionDate>
      </Citation>
      <Abstract>
        <p xml:lang="en">This dataset documents the
development and identity of Queen Arwa University on the internet
from its first website in 1998</p>
        <p xml:lang="ar"> يوثق هذا الملف تطور وهوية جامعة
1998 </p>
      </Abstract>
      <Coverage>
        <TopicalCoverage>
          <p xml:lang="en">Education, Digital History,
University Development</p>
          <p xml:lang="ar"> التعليم، التاريخ الرقمي، تطوير
الجامعات</p>
        </TopicalCoverage>
        <GeographicCoverage>
          <p xml:lang="en">Yemen</p>
          <p xml:lang="ar">اليمن</p>
        </GeographicCoverage>
        <TemporalCoverage>
          <p xml:lang="en">1998-2021</p>
          <p xml:lang="ar">1998-2021</p>
        </TemporalCoverage>
      </Coverage>
      <FundingInformation>
        <FunderName>
          <p xml:lang="en">Queen Arwa University</p>
          <p xml:lang="ar">جامعة الملكة أروى</p>
        </FunderName>
      </FundingInformation>
    </StudyUnit>
  </ResourcePackage>
</DDIInstance>
```

## Dublincore record

```
<?xml version="1.0" encoding="UTF-8"?>
<metadata
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://purl.org/dc/elements/1.1/ http://
dublincore.org/schemas/xmls/qdc/2008/02/11/dc.xsd">

  <dc:title xml:lang="en"> Tārīkh wa-huwīyat Jāmi‘at al-Malikah
Arwá ‘alá al-intarnit wa-marāḥil al-taṭwīr mundhu al-‘āmm 1998-2021 M
</dc:title>
  <dc:title xml:lang="ar">تاريخ وهوية جامعة الملكة أروى على الإنترنت
ومراحل التطوير منذ العام 1998 - 2021 م
</dc:title>

  <dc:creator xml:lang="en">al-Şalfy, Şadām</dc:creator>
  <dc:creator xml:lang="ar">صدام الصلبي، صدام</dc:creator>

  <dc:subject xml:lang="en">Education, Digital History, University
Development</dc:subject>
  <dc:subject xml:lang="ar">التعليم، التاريخ الرقمي، تطوير
الجامعات</dc:subject>

  <dc:description xml:lang="en">This dataset documents the
development and identity of Queen Arwa University on the internet
from its first website in 1998</dc:description>
  <dc:description xml:lang="ar">يوثق هذا الملف تطور وهوية جامعة
الملكة أروى على الإنترنت من أول موقع إلكتروني لها في عام 1998
</dc:description>

  <dc:publisher xml:lang="en">Harvard Dataverse</dc:publisher>

  <dc:date>2023-02-03</dc:date>

  <dc:type xml:lang="en">Dataset</dc:type>

  <dc:format>Electronic Resource</dc:format>

  <dc:identifier>https://doi.org/10.7910/DVN/PISAJ5</dc:identifier>

  <dc:language xsi:type="dcterms:ISO639-2">ara</dc:language>
  <dc:language xsi:type="dcterms:ISO639-2">eng</dc:language>

  <dc:coverage xml:lang="en">1998-2021 - Yemen</dc:coverage>
  <dc:coverage xml:lang="ar">1998-2021 - اليمن</dc:coverage>

  <dc:rights xml:lang="en">CC BY 4.0</dc:rights>
</metadata>
```


# Research data Repository systems



- Current situation of software to support transliterated data
- Supporting the right-to-left
- Unicode
- Indexing and Searching

# Research Data Repository (Dataverse)

## Metadata template

 **Dataverse** Add Data ▾ Search ▾ About User Guide Support Mohamed elzlabany ▾

**Citation Metadata** ▲


<b>Title</b> * <span>?</span>	<i>Custom Instructions:</i> (None - click to add)	<input type="text"/>	
<b>Subtitle</b> <span>?</span>	<i>Custom Instructions:</i> (None - click to add)	<input type="text"/>	
<b>Alternative Title</b> <span>?</span>	<i>Custom Instructions:</i> (None - click to add)	<input type="text"/>	<input type="button" value="+"/>
<b>Alternative URL</b> <span>?</span>	<i>Custom Instructions:</i> (None - click to add)	<input type="text" value="https://"/>	
<b>Other Identifier</b> <span>?</span>	<i>Custom Instructions:</i> (None - click to add)	<b>Agency</b> <span>?</span> <input type="text"/>	<b>Identifier</b> <span>?</span> <input type="text"/>
<b>Author</b> * <span>?</span>	<i>Custom Instructions:</i> (None - click to add)	<b>Name</b> * <span>?</span> <input type="text" value="1) Family Name, Given Name or 2) Orgar"/>	<b>Affiliation</b> <span>?</span> <input type="text" value="Select a research organization"/>
			<b>Identifier Type</b> <span>?</span> <input type="text" value="Select..."/>
		<b>Identifier</b> <span>?</span> <input type="text"/>	
<b>Point of Contact</b> * <span>?</span>	<i>Custom Instructions:</i> (None - click to add)		

## Citation Metadata ▲

<b>Persistent Identifier</b> ?	doi:10.7910/DVN/PISAJ5
<b>Publication Date</b> ?	2023-02-03
<b>Title</b> ?	تاريخ وهوية جامعة الملكة أروى على الإنترنت ومراحل التطوير منذ العام 1998 - 2021 م
<b>Author</b> ?	Al-Slfi, Saddam (Queen Arwa University) - ResearcherID: <a href="https://orcid.org/0000-0002-4006-0193">https://orcid.org/0000-0002-4006-0193</a>
<b>Point of Contact</b> ?	Use email button above to contact.  Al-Slfi, Saddam (Queen Arwa University)
<b>Description</b> ?	تأسست جامعة الملكة أروى ومارست نشاطها التعليمي منذ مطلع العام 1996 م واستقبلت العديد من الطلبة وأعضاء هيئة التدريس المحليين والدوليين وشكلت احدى اهم مؤسسات التعليم العالي الأهلي في تلك الحقبة وضمن أنشطه الطلاب والتطبيق العملي للمواد الدراسية قام مجموعة من الطلاب بإنشاء اول موقع إلكتروني وكان يهتم بالتعريف عن الجامعة وأخبارها واهم كلياتها وأقسامها والعديد من الخدمات التي تراها بدائية في وقتنا الحاضر الا انها كانت مشروع ناجح نال إعجاب الجميع ، وعندما علمت الدكتورة وهيبة فارح رئيس مجلس أمناء الجامعة بالأمر اشرفت بنفسها وبشكل مباشر على مراحل تصميم وبرمجة موقع جامعة الملكة أروى وتكريماً لهؤلاء الطلبة اعتمدت رئاسة الجامعة المشروع كموقع رسمي يمثل نشاط جامعة الملكة على الشبكة العنكبوتية وتم العمل رسمياً بالموقع ونشرة على الإنترنت من مطلع العام 1998 م وكانت اول جامعة يمنية تطلق موقعها الالكتروني على شبكة الإنترنت وحصلت على تقييم جيد وعضويه لدى منظمة اليونسكو للتعليم في حينها. (07-01-2022)
<b>Subject</b> ?	Engineering
<b>Keyword</b> ?	Queen Arwa University <a href="https://qau.edu.ye/ebook/270">https://qau.edu.ye/ebook/270</a>
<b>Related Publication</b> ?	url: <a href="https://qau.edu.ye/ebook/270">https://qau.edu.ye/ebook/270</a>
<b>Depositor</b> ?	Al-Slfi, Saddam
<b>Deposit Date</b> ?	2023-02-03

## Sample record from Harvard dataverse

# Arabic language search

Add Data Search About User Guide Support Sign Up Log In

Metrics 63,103,960 Downloads Contact Share

Q Advanced Search + Add Data

- 🔗 Dataverses (0)
- 📄 Datasets (1)
- 📁 Files (1)

**Publication Year**  
2023 (2)

**Subject**  
Engineering (1)

**Author Name**  
Al-Sifi, Saddam (1)

**Author Affiliation**  
Queen Arwa University (1)

**File Type**  
Document (1)

**Access**  
Public (1)

**1 to 2 of 2 Results** Sort

**تاريخ وهوية جامعة الملكة أروى على الإنترنت ومراحل التطوير منذ العام 1998 - 2021 م** 📄

Feb 3, 2023

Al-Sifi, Saddam, 2023, "تاريخ وهوية جامعة الملكة أروى على الإنترنت ومراحل التطوير منذ العام 2021 - 1998 م", <https://doi.org/10.7910/DVN/PISAJ5>, Harvard Dataverse, V1

تأسست جامعة الملكة أروى ومارست نشاطها التعليمي منذ مطلع العام 1996 م واستقبلت العديد من الطلبة وأعضاء هيئة التدريس المحليين والدوليين وشكلت إحدى أهم مؤسسات التعليم العالي الأهلي في تلك الحقبة وضمن أنشطه الطلاب والتطبيق العملي للمواد الدراسية قام مجموعة من الطلاب بإنشاء أول موقع لـ...

**تاريخ وهوية جامعة الملكة أروى على الإنترنت ومراحل التطوير منذ العام 1998 - 2021 م.pdf** 📄

Feb 3, 2023 - تاريخ وهوية جامعة الملكة أروى على الإنترنت ومراحل التطوير منذ العام 2021 - 1998 م

Adobe PDF - 777.5 KB - MD5: 2da...efb

<https://qau.edu.ye/>

Filename Without Extension: تاريخ وهوية جامعة الملكة أروى على الإنترنت ومراحل التطوير منذ العام 2021 - 1998 م

# Automatic Transliteration challenge



- current practices of (ALA-DIN 31635-ISO 233)
- Transliteration from Roman to Arabic (tools)
- Transliteration from Arabic to Roman (challenges, tools)

# Transliterator software (LOC)

**T** Transliterator 1.6.1  
File Edit Help

Language / Script: Arabic

```

906: : #a 7 #b cbc #c pccadap #d 3 #e ncip #f 20 #g y-nonr
925:0 : #a acquire #b 1 shelf copy #x policy default
955: : #b we48 2011-07-01 z-processor UC to ASME/ME #
      BCCD
010: : #a 2009339919
035: : #a (OCoLC)ocn426166452
040: : #a PUL #c PUL #d PUL #d IXA #d DLC
042: : #a pcc #a lcode
043: : #a a-le---
050:00: #a DS87.2.H37 #b A635 2009
066: : #c (3
245:03: #6 880-01 #a al-'Adālah al-dawliyah bayna qarārāt a
      qarārāt Majlis al-Amn wa-taqārīr lijān al-tahqīq al-dav
      lil-mahkamah al-dawliyah al-khāṣṣah bi-Lubnān al-r
      Ra'īs al-shahīd Rafīq al-Ḥarīrī : #b dalīl shāmil li-kāffē
      mundhu Tishrīn al-Awwal 2004 wa-li-ghāyat Ādhār 2009.
*880:02: #6 245-01//r #a العدالة الدولية بين قرارات الأمم المتحدة وقرارات مجلس الأمن وتقارير لجان التحقيق الدولية
      دليل شامل لكافة قرارات #b والنظام القانوني للمحكمة الدولية الخاصة بلبنان المتعلقة باعتقال الرئيس الشهيد رفيق الحريري
      2009.ولغاية آذار 2004مجلس الأمن منذ تشرين الأول
246:14: #6 880-02 #a 'Adālah al-dawliyah fī qaḍiyat iḡhtiyāl al-Ra'īs al-shahīd Rafīq al-
      Ḥarīrī
*880:14: #6 246-02//r #a عدالة الدولية في قضية اعتقال الرئيس الشهيد رفيق الحريري
250: : #6 880-03 #a al-Ṭab'ah 1.
*880: : #6 250-03//r #a الطبعة 1.
260: : #6 880-04 #a Bayrūt : #b al-Ghazāl lil-Nashr, #c 2009.
*880: : #6 260-04//r #a بيروت: #b الغزال للنشر، #c 2009.

```

FieldsIncluded=100 110 111 130 240 245 246 250 260 264 440 490 600 610 611 630 651 700 710 711 730 740 800  
830  
SubfieldsAlwaysExcluded=uvxy0123456789  
OtherSubfieldsExcludedByTag=246/i 260/c 264/c 650/a  
SubfieldCode=(3  
IncludeFormattingCharactersLcPattern=True

# Punctuation marks:  
# %-U+066A ; cannot transliterate the truncation character  
\*=U+066D  
,=U+060C  
;=U+061B  
?=U+061F

# Hyphenated prefixes:  
wa--=U+0648  
b1--=U+0628  
a1--=U+0627U+0644  
l1l--=U+0644U+0644  
li--=U+0644  
laU+0304--=U+0644  
fiU+0304--=U+0641U+064A  
ka--=U+0643

# Exceptions for specific words  
# Allah  
Al1laU+0304h=U+0627U+0644U+0644U+0647

# Qur'an  
QurU+02BCaU+0304n=U+0642U+0631U+0622U+0646

# lillah  
l1llaU+0304h=U+0644U+0644U+0647

# billah  
b1llaU+0304h=U+0628U+0644U+0644U+0647

# Rahman  
RahU+0323maU+0304n=U+0631U+062DU+0645U+0646

# Ruwat  
RuwaU+0304t=U+0631U+0648U+0627U+0629

UCase 1st    UCase All    Define    Convert highlighted text    Import    Resequence only  
LCase 1st    Replace    Print    Convert whole record    Export    Exit    ?

# Arabic Romanization ALA-LC

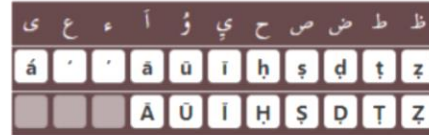
# رؤمنة اللغة العربية ALA-LC

A transliteration tool following the [American Library Association - Library of Congress](#)

أداة لرؤمنة الحروف العربية طبقًا لمعايير جمعية المكتبات الأمريكية - مكتبة الكونجرس

Ad'ham huwa min qāma bf'l dhālika bātbā'h Muḥammad ma'a Qalīl min al-imān bb'thh al-Rasūl al-maymūn  
Al'st'thār bāl'mwāl al-'ashwā'iyah wa-al-tarbiyah aldwlḡh fī al-Kulliyah al-naw'iyah

أدهم هو من قام بفعل ذلك باتباعه محمد مع قليل من الإيمان ببعثة الرسول الميمون  
الإستثمار بالأموال العشوائية والتربية الدولية في الكلية النوعية



Try this example: جوهري، طنطاوي، جمة الشعروالتعريب. وهي جمة قطع من النظم لمشهور شعراء الانجليز، كساها حلة الشعر العربي.

This system follows the [ALA-LC Arabic Romanization Table and Guidelines](#). It is based on the "MLE Simple" model described in Eryani & Habash (2021), with slight modifications. This model, selected for its speed, uses a large corpus of training data to calculate the most frequent Romanization given a certain Arabic input. The system achieves 90% accuracy (ignoring capitalization errors) and 84% exact match accuracy. In out-of-vocabulary scenarios, about 3% of the words, the system falls back on a simple character transliteration technique that Romanizes the input as is, without guessing unmarked vowels or splitting proclitics.

Reference: Fadhī Eryani & Nizar Habash. *Automatic Romanization of Arabic Bibliographic Records*. In Proceedings of the Arabic Natural Language Processing Workshop. 2021. [\[PDF\]](#) [\[BIB\]](#) [\[Github\]](#)



جامعة نيويورك أبوظبي

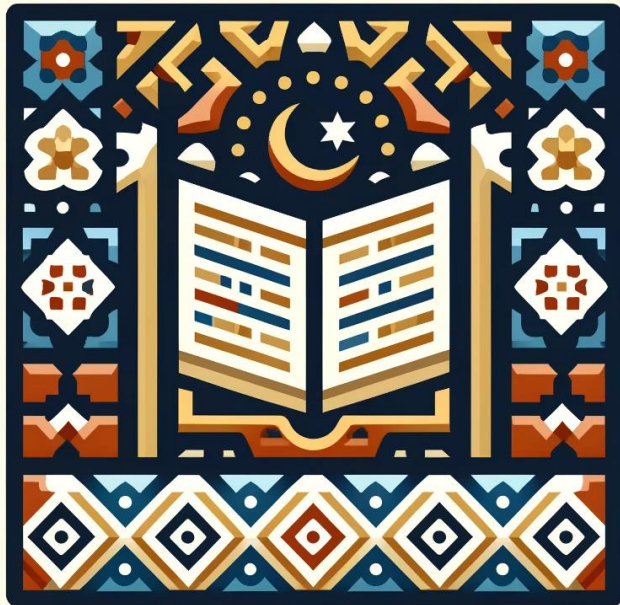


<https://romanize-arabic.camel-lab.com/>

<https://aclanthology.org/2021.wanlp-1.23.pdf>



# Recommended Practices to enhance the transliterated metadata



- **Develop new metadata templates in the research data repository to have fields with more than one language**
- **Use appropriate language fields**
- **Avoid mixing languages in a single field**
- **Aim for consistency between metadata fields and galley**
- **Avoid unnecessary transliteration/romanization**



mh\_zalabany@hotmail.com  
[www.linkedin.com/in/mohamedelzalabany](http://www.linkedin.com/in/mohamedelzalabany)