

Exploration of Metadata Practices in Digital Collections of Archives with Arabian Language Materials

Saleh Aljalalmah

Basic Education College

The Public Authority for Applied Education and Training

Kuwait

sh.aljalalmah@paaet.edu.kw

Oksana L. Zavalina

Department of Information Science

University of North Texas

United States of America

Oksana.Zavalina@unt.edu

ABSTRACT

A high proportion of materials held by archives in Arabian Gulf and included in digital collections are oral histories, manuscripts, and other language content. As metadata is important for resource discovery, this study aimed to develop understanding of the current state of metadata practices in digital collections of archival institutions in the Arabian Gulf region. It also explored perspectives (including attitudes and possible barriers) for development of large-scale regional portals that would facilitate discovery of Arab digital archives (including language collections) by aggregating metadata. This research project used semi-structured interviews of the managers of 4 out of 5 digital language archives in Kuwait. Results provide insights into perspectives of metadata interoperability among archives and suggest the need for metadata training, and documenting metadata creation guidelines. Findings contribute to evaluating the feasibility of and planning for future functional regional aggregations of cultural heritage digital collections.

CCS CONCEPTS

• Information systems → Database administration • Applied computing → Computers in other domains → Digital libraries and archives • Applied computing → Document management and text processing → Document management → Document metadata • General and reference → Evaluation

KEYWORDS

Arabic language archives, information organization, information access, metadata interoperability

ACM Reference format:

Saleh Aljalalmah and Oksana L. Zavalina. 2023. Exploration of Metadata Practices in Digital Collections of Archives with Arabian Language Materials. In Proceedings of the 2nd International Workshop on Digital Language Archives (LangArc-2023), ACM/IEEE Joint Conference on Digital Libraries. USA, 4 pages. <https://doi.org/10.12794/langarc2114295>

1 INTRODUCTION AND LITERATURE REVIEW

Archives are cultural heritage institutions whose main function is to provide access to information. This access is enabled through information organization, which includes development and application of data content standards that guide metadata creation, for example, Describing Archives: A Content Standard (DACS). In information science and practice, the term metadata refers to bibliographic records that represent materials held by cultural heritage institutions (archives, libraries, and museums). Metadata is considered according to the type, function, domain, etc. (e.g., [4],[17]). Metadata must provide easy access and retrieval for the users, as well as support for the work tasks of collection managers [18]. Frameworks for evaluating metadata quality formulated metadata quality criteria (e.g., [7]).

Metadata schemes (e.g., Dublin Core, Encoded Archival Description, Machine Readable Cataloging: MARC, Open Language Archives Community Metadata: OLAC) include metadata element sets accompanied by metadata creation guidelines. Organization of information also entails development and application of the controlled vocabularies for names of persons and institutions (e.g., Union List of Artist Names, Thesaurus of Geographic Names), subject, genre, language, and other terms (e.g., Library of Congress Subject Headings, OLAC Discourse Type Vocabulary, Glottolog), and classification systems (e.g., Dewey Decimal Classification: DDC, etc.). In addition to these international standards for information organization, institutions often develop their own local metadata schemes, controlled vocabularies, and guidelines, or create adaptations of existing standards for their digital collections to better meet their target audience's needs.

To facilitate access and improve user experience, metadata records that represent materials held by archives, libraries, and museums are brought together in aggregations that serve as centralized points of access. Well-known examples of such aggregations include multinational (Europeana), and regional (e.g., Digital Library of the Caribbean). For a portal like that to function properly and support resource discovery, metadata aggregated in a portal needs to be interoperable. Metadata interoperability can be defined as “the compatibility of two or more systems such that they

can exchange information and data and can use the exchanged information and data without any special manipulation” [8]. Quality of metadata has been evaluated and discussed in relation to its interoperability when brought together into aggregations of digital content (e.g., [20]).

Metadata interoperability is commonly achieved through development of mappings between different metadata schemes, and aggregation-wide metadata guidelines (e.g., metadata application profile) that metadata harvested into the aggregation needs to conform to (e.g., [9], [14]). The first step in this process is the background exploration of metadata practices and standards used by institutions that will likely participate in the portal as contributors. The first centralized portal – the Digital Library of the Middle East that aims to provide access to digital content from the Arabian Gulf archives, libraries, and museums – was launched in 2021.

Cultural heritage institutions in the Middle East have a long history of using international information organization standards. For example, Egyptian Organization for Standardization and Quality Control helped translate international information organization standards since 1957; Jordan Library and Information Association modified Dewey Decimal Classification system to better meet the needs of regional users since 1970s [10]. Arabian Gulf countries embrace the opportunity to share their knowledge with other countries through digital collections. The digitization movement in the region was pioneered by Qatar that launched its digital library in 2012 (Qatar National Library, no date). UNESCO special envoy for basic and higher education believes that digitization efforts could help the world better understand Arab culture [5].

Arabic language, along with e-government and information retrieval is among the most common research topics of the articles by authors from the Arabian Gulf published in top journals [19]. At the same time, there is shortage of investigations into the information organization practices in Arabian Gulf countries, especially in archives. So far, only one published paper and one poster abstract focused on the creation or adoption of metadata, metadata quality assurance, metadata interoperability in Arabian Gulf institutions, including one Kuwaiti archive that provides access to Arab language and culture data [2, 3]. No research has yet been published on aggregation of metadata in the portals that provide central point of access to collections of cultural heritage institutions, including archives. Our exploratory study examined the status of the organization of knowledge in archives in an Arabian Gulf country that can be characterized as digital language archives as they:

- have been developed largely by collecting oral histories after the Gulf war of 1990 when most Kuwaiti archival collections perished [1]
- provide access to digitized and/or born-digital Arab language materials.

2 METHOD

The research questions addressed by this study were:

- What are the techniques and approaches used in archives’ information organization: metadata schema, data content standards, controlled vocabularies, content management tools, search options, metadata harvesting, etc.?
- What are the archive metadata managers’ perceived readiness for and barrier to aggregating metadata in regional portals that would facilitate discovery of Arab digital archives (including language collections)?

This study focused on one Arabian Gulf country: Kuwait. Archives were selected based on criteria which included location of headquarters in Kuwait, and availability of one or more digital collections managed by the archive. At the time of data collection, only 5 archives in Kuwait met these criteria. Potential respondents were selected from the lists of employees available on their institutions’ websites. Interviewing those employees of Kuwaiti archives who make decisions about information organization in digital collections allowed to identify similarities and differences, as well as opportunities and challenges for providing access to digital collections via large-scale centralized portals. The interview recruitment email was sent in two languages (English and Arabic) and invited to respond in the language of respondent’s choice. The response rate constituted 80%, with representatives of 4 archives participating in the study.

We used email interviews as they allow participants to find time in their schedule to provide more thoughtful, reflective responses [13]. Previous studies on metadata-related topics relied on email interviews and found this approach effective [15]. The semi-structured interview questions were sent to the participants in both English and Arabic and participants had the freedom to choose the language of their answers.

3 FINDINGS AND DISCUSSION

The participants’ views regarding large scale portals that would aggregate metadata records representing items in the Arabian Gulf countries’ digital collections were very positive overall. All 4 participants believed this to be a necessary development. Respondents also raised some concerns, specifically emphasizing the associated costs. Despite Arabian Gulf countries’ generally strong economies, governments are not necessarily ready to invest in such projects due to various reasons, including the post-pandemic shift of focus to other areas, and traditional lack of government support to archives. Participants also pointed out the lack of the workforce with the expertise and preparation necessary to design and implement projects like this in the Arabian Gulf countries, as there are no established large-scale aggregations in the region. Participants suggested that as a possible solution, which would however increase project costs, international experts with experience implementing and maintaining such portals could be hired. Relying on professionals from other countries is already an established practice.

The study revealed a variety of training levels among those responsible for metadata in digital collections. One respondent reported having a graduate degree in Library and Information Science, another in Library and Information Technology, and one

more in Computer Science. One participant had an International Baccalaureate 2-year degree in unrelated field although took some archiving coursework. In addition to formal education, one archive's metadata manager reported that they were trained by more experienced colleagues/professionals, another one stated that they received personal on-the-job training, one participant was entirely self-trained in the archive metadata management tasks of their job, and the 4th participant reported a combination of self-training and attending workshops. Respondents had a wide range of years of experience in the field: one worked in archives since as early as 1989 and the most recently hired of our study participants started in 2008.

In the digital collections of two archives, searching is available in both Arabic and English languages. Two other archives only provide Arabic-language search capability. In 3 digital archives that participated in the study users can print and save/download digital items, as well as send them over email. Sharing on social media is also available in 2 archives. One archive implemented only the search function and no other navigation/interaction functions.

No single digital content management system was found to be used by more than one archive. Respondents were found to rely on VIRTUA, SQL, and Symphony. One archive developed its own in-house digital content management system. None of the participants reported that their metadata records are exposed for harvesting using Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

Three respondents reported using MARC 21 as the metadata scheme for their digital collections. One archive developed a local metadata scheme based on Dublin Core as shown in Figure 1 below, with 13 descriptive metadata elements based on 12 Dublin Core elements. None of the participants used the archive-specific metadata scheme Encoded Archival Description (EAD).

Two Kuwaiti digital language archives used standard Dewey Decimal Classification (DDC), and one developed the DDC-inspired local "Islamic Dewey" classification system. One archive relied on the outdated version 20 of the DDC released in 1989, while 3 newer versions have appeared since then (the latest in 2012) due to the limited budget not allowing for upgrade. One respondent's answer to the question about classification scheme(s) used was invalid as they referred to the Anglo-American Cataloging Rules which is not a classification system.

Two study participants reported using subject headings lists developed by and for Arabian Gulf countries: El-Khazindar list of Arabic Subject Headings (Khazindar, 1983) or The Major List of Arab Subject Headings (<https://www.amazon.com/-/القائمة-الكبرى-العربية-ebook/dp/B07FXFCNNN/>). Maknaz Expanded Thesaurus (<http://en.maknaz.org/>) developed in the Arabian Gulf and available based on subscription is used in one archive as a controlled vocabulary for genre terms and names, in addition to subject terms. Participants did not mention any other controlled vocabularies used and relied on free-text keywords for metadata fields other than those representing aboutness, genre, and creators or contributors.

The image shows a screenshot of a metadata record interface. The record is displayed in a structured form with various fields and labels. The labels are: Date (Islamic), Date (Gregorian), Identifier, Source, Relation, Rights, Type, Creator, Contributor, Title, Description, and Descriptors (subject terms). The record contains the following information:

- Date (Islamic): ٢١٠٠-١١٠٠
- Date (Gregorian): ٢١٠٠-١١٠٠
- Identifier: 710488
- Source: المكتبة الوطنية الكويتية
- Relation: ٩781
- Rights: ٧٧٨
- Type: ٦٥٣
- Creator: مكتبة الوطن
- Contributor: مكتبة الوطن
- Title: تربية الطفل
- Description: كتاب باللغة العربية
- Descriptors (subject terms): تربية الطفل

Figure 1: Metadata record example

To the question about availability of local metadata creation guidelines that are used to document and guide the metadata practices at their institutions, two interview participants responded negatively (with one of them commenting that they follow the guidelines of an existing standard: MARC 21). One interviewee reported having metadata creation guidelines restricted for internal use only on the site of the archive. None of the participants mentioned following the metadata guidelines found in the international standards for archival description: ISAD[G] and ISAAR[CPF].

4 CONCLUSION

This exploratory study's findings will be useful as a benchmark for future research. Results show that some international standards of information organization have been adopted or adapted for the regional needs by those archives in Kuwait that include digital collections of Arabic-language materials and therefore fit the definition of a digital language archive. Our study found lack of metadata creation guidelines documenting local practices in application of the standards or locally developed information organization tools, and lack of participation in metadata harvesting with OAI-PMH or equivalents, which indicate potential challenges for interoperability of metadata. Other potential challenges include limited technical skill sets and financial resources available to support aggregations of archival digital collections.

Future research is needed to investigate metadata quality in digital archive collections with Arabian language content. Future studies will also examine the status of the organization of knowledge in other digital language archives in Kuwait, as more such archives are developed, and in other Arabian Gulf countries: Bahrain, Oman, Qatar, Saudi Arabia, and United Emirates. The observed metadata quality and metadata-related practices in these countries might affect the establishment of large-scale portals that include metadata records from entire region.

To ensure functionality of aggregations the demand for which is growing, Arabian digital language archives will need to take several steps towards ensuring metadata interoperability. This includes adopting and applying unified metadata creation

guidelines. Generating crosswalks that show equivalences between metadata fields used by different archives in the region, as well as mapping between different controlled vocabularies used, is another important step worth considering.

REFERENCES

- [1] Ahmed, S. (2018). "Seeking information from the lips of people": oral history in the archives of Qatar and the Gulf region. *Archival Science*, 18(3), 225-240.
- [2] Aljalalmah, S., & Zavalina, O. L. (2021). Information representation and knowledge organization in cultural heritage organizations in Arabian Gulf: A comparative case study. *Journal of Information and Knowledge Management*, 20(2), 1-20. <https://doi.org/10.1142/S0219649221500507>
- [3] Aljalalmah, S., & Zavalina, O.L. (2021). A case study of information representation in a Kuwaiti archive. In *iConference 2021 Proceedings*. <http://hdl.handle.net/2142/109683>
- [4] Baca, M. (2016). *Introduction to metadata*. 3rd edition. Getty Publications. Retrieved from <http://www.getty.edu/publications/intrometadata/>
- [5] Bade, R. (2010, May 27). Digital library project aims at the Middle East. *Roll Call*. Retrieved from <https://www.rollcall.com/2010/05/26/digital-library-project-aims-at-the-middle-east/>
- [6] القائمة الكبرى - العربية - المجلد الأول [The Major List of Arab Subject Headings (2018): Arabic Edition]. Retrieved from <https://www.amazon.com/-/ar-ؤوس-/القائمة-الكبرى-العربية-ebook/dp/B07FXFCNNN/>
- [7] Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: defining, expressing, exploiting. ALA editions.
- [8] CC:DA (ALCTS/CCS/Committee on Cataloging: Description and Access). (2000). *Task Force on Metadata: Final report*, June 16, 2000. Retrieved from <http://downloads.alcts.ala.org/ccda/tf-meta6.html>.
- [9] Chan, L. M., & Zeng, M. L. (2006). Metadata interoperability and standardization—a study of methodology part I. *D-Lib magazine*, 12(6), 1082-9873.
- [10] Eid, S. (2019). Library metadata standards and Linked Data Services: An introduction to Arab and international organizations. *Journal of Library Metadata*, 19(3-4), 163-185.
- [11] Holloway, M. F. (1959). Patterns of library service in the Middle East. *Library Trends*, 8 (October), 192-208.
- [12] Khazindar, I. A. (1983). *Qā'imāt ru'ūs al-mawdū'āt al-'Arabīyah*: [Alkhazindar List of Arabic Subject Headings]. al-Kuwayt: Yuṭlabu min Dār al-Bahūth al-'Ilmīyah.
- [13] Meho, L. I. (2006). E-mail interviewing in qualitative research: A methodological discussion. *Journal of the American Society for Information Science and Technology*, 57(10), 1284-1295.
- [14] Park, J., & Childress, E. (2009). Dublin Core metadata semantics: An analysis of the perspectives of information professionals. *Journal of Information Science*, 35(6), 727-739.
- [15] Park, J. R., & Tosaka, Y. (2015). RDA implementation and training issues across United States academic libraries: An in-depth e-mail interview study. *Journal of Education for Library and Information Science*, 56(3), 252-266.
- [16] Qatar National Library. (No date). Qatar Digital Library. Retrieved from <https://www.qdl.qa/en/about>
- [17] Riley, J. (2010). Seeing Standards: A Visualization of the Metadata Universe. Retrieved from <http://jenriley.com/metadatamap/seeingstandards.pdf>
- [18] Riley, J. (2017). *Understanding metadata*. Washington DC, United States: National Information Standards Organization (<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>), 23.
- [19] Robinson, B. (2016). Addressing bias in the cataloging and classification of Arabic and Islamic materials: Approaches from domain analysis. In A.B. Click et al. (eds.), *Library and Information Science in the Middle East and North Africa* (pp. 255-269). Munich: De Gruyter Saur.
- [20] Shreeves, S. L., Knutson, E. M., Stvilia, B., Palmer, C. L., Twidale, M. B., & Cole, T. W. (2005). Is quality metadata shareable metadata? The implications of local metadata practices for federated collections. In Hugh A. Thompson (ed.), *Proceedings of the Twelfth National Conference of the Association of College and Research Libraries* (pp. 223-237). Chicago, IL: Association of College and Research Libraries. Retrieved from <http://hdl.handle.net/2142/145>.