Towards Making Shared Metadata Interoperable across the Open Language Archives Community

Hugh J. Paterson III
Department of Information Science
University of North Texas
United States of America
i@hp3.me

ABSTRACT

This paper presents two methods for connecting aggregated records to their source institutional metadata profiles. The use case of the Open Language Archives Community (OLAC) application profile is considered and evaluated. The design purpose of OLAC is to share knowledge about language resources. To that end, the OLAC metadata application profile supports the exchange of metadata so that it can be aggregated and serve the needs of end-users. Uniformity in the semantic use of elements within the application profile provides the greatest utility for end-users. Discovering the source of semantic diversity remains a challenge. A first step in providing scholars access to the semantics of aggregated metadata is to publish the local metadata profiles used by institutions.

CCS CONCEPTS

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

KEYWORDS

OAI-PMH, Metadata Semantics, Documentation, Open Language Archives Community

ACM Reference format:

Hugh J, Paterson III. 2023. Towards Making Shared Metadata Interoperable across the Open Language Archives Community. In Proceedings of the 2nd International Workshop on Digital Language Archives (LangArc-2023), ACM/IEEE Joint Conference on Digital Libraries. USA, 5 pages. https://doi.org/10.12794/langarc2114297

1 INTRODUCTION

Several scholars [13, 36, 37] note that stewards of language resources must strategize to engage with multiple audiences. That is, stewardship institutions need to consider multiple audiences and communication channels as they look to increase engagement with stewarded resources. This impacts cataloging (metadata record creation), resource discovery, and user interface design. The search and discovery process directly relates to successful stewardship. Undergirding search and discovery success is the issue of metadata quality. Yasser [38] summarizes Zeng and Qin [40] in describing the relationship between metadata quality and the ability of a digital library to meet its goals: "... poorly created metadata records result in poor retrieval and limit accessibility to collections, ultimately exercising a detrimental impact on the continuing adoption and use of a digital library. In consequence, problematic metadata is highly undesirable and needs to be understood for further action in developing remedial solutions."

The Open Language Archives Community (OLAC) has for twenty years [4, 5] provided a metadata application profile to archives and other data providers to help them meet their resource engagement goals. Recent research on language resource stewardship practices has reported on both user interface [39] and the content of description records [10]. These teams of scholars source their evidence directly via the web presentations of language resource stewards. Alongside these efforts, other work has focused on the display and presentation of records via the OLAC interface, which is often a derivative from the native metadata application profiles at institutions [29, 30]. The research that has analyzed OLAC records [29, 30] has focused on the semantics [22, 27] and usage of the metadata elements within records relative to the nature of the artifact being described. As such, it falls broadly into "metadata quality" research which investigates the accuracy, completeness, and consistency in records and across record sets [9, 23, 35]. This more recent work contrasts with previous models measuring metadata quality of OLAC records which used quantitative approaches to measure the number of elements provided per record [14].

Metadata accuracy and consistency has been addressed in large scale aggregation projects in a variety of ways, often including metadata utilities which attempt to regularize records for the benefit of end-users [12, 19, 24, 25, 31]. However, completeness can remain a challenge due to the variety of semantic options. Completeness is a measure of the totality of description versus the total possible description within the metadata schema based on an object's nature. Accessing source metadata schema documentation brings clarity to evaluation processes.

The issue of metadata quality is central to the idea of creating shareable and interoperable data which end-users will find useful in their searches [32]. Resource descriptions need to be interoperable not only at the syntactic level (Dublin Core Elements)

but also at the semantic level, e.g., which elements are used and how the information values in the elements are derived. Consistent metadata quality—including semantics—is important to end-user experience. High-quality metadata is especially impactful when the engagement platform becomes dynamic or when visual representations are dependent on the content within the record. These are critical issues for the OLAC community to address if OLAC is to survive in the digital libraries' "mainstream" as Bird and Simons [4] envision.

2 OLAC METADATA

To fully appreciate the context of OLAC records, a deeper understanding of the processes by which OLAC records are generated is needed. For many OLAC data providers, the metadata records offered for aggregation are transformed (i.e., cross-walked) from a "native" or institutional metadata schema into the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and Dublin Core-based OLAC metadata application profile [1–3, 33]. For example, SIL International's Language & Culture Archives uses a Dublin Core based application profile which has no public documentation but can be investigated via the open-source application RAMP [26].

Archive of the Indigenous Languages of Latin America (AILLA) has no public documentation on its website regarding its metadata schema, but it has been stated that it uses an IMDI based application profile [15]. IMDI (ISLE Meta Data Initiative) is a metadata profile which started out in Europe with projects such as DoBeS [7, 8, 17, 18]. It evolved into CMDI, a modular metadata schema used by various CLARIN entities [6, 11]. Some portion of the metadata schemas of several language resource stewards including The Endangered Languages Archive (ELAR) and The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) can be seen in the LaMeta application's code. Broadly across aggregation efforts cross-walking metadata is a common architectural process data providers support to communicate with aggregators [16]. The conceptual process is illustrated in Figure 1.

The OLAC metadata application profile (OLAC-AP) provides specific access points for end-users to discover and engage with resource records. Data-provider- or institution-specific metadata application profiles may be designed to facilitate institution-specific user interfaces or reporting requirements. These may be in addition to, or in lieu of, access points provided via OLAC interfaces. The use of institution-specific metadata application profiles is not uncommon across digital library projects. This is the very reason that OAI-PMH was created, and that Dublin Core remains so pervasive across the digital libraries landscape—there is a clear need for (1) a generalizable super-set of metadata and (2) interface-building around generalized metadata.

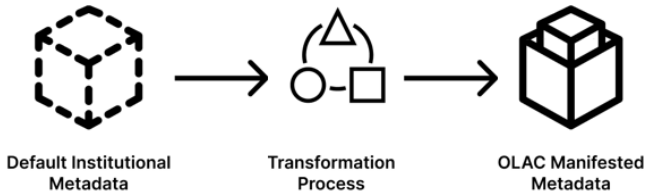

Figure 1: Moving metadata from data-provider "local" formats to the OLAC format.

The semantics of specific fields in institutional metadata application profiles may differ from the semantics of the most appropriate field in the OLAC-AP. Additionally, certain fields in the OLAC-AP may be inferred during the transformation process, e.g., the SIL Language & Culture Archives does not record the DCMI Type in their local application profile but generate this field for OLAC consumption based on several other factors. This means that inconsistencies or low-quality metadata may have several sources. Primary among these are signal noise via the transformation process and low-quality cataloging at the point of data origin. As metadata professionals look at OLAC metadata to evaluate record quality and interface utility for end-users, it is useful to consult the documentation for the transformation process and the institution-specific metadata application profiles. Additionally, institutional metadata application profiles and cataloging practices may evolve over time.

These changes may have different evolutionary cycles from metadata transformation processes. This can leave OLAC metadata in a discombobulated state while metadata is well-formed (to local standards) at data providers. However, many of the institution-specific metadata application profiles for OLAC data contributors are not accessible to the public and neither is documentation on the transformation process. The state of documentation access for OLAC data providers is not entirely out of the norm. Park and Tosaka [28], when investigating metadata aggregators, their application profiles, and the use of Dublin Core, observed that many data providers add to application profiles and frequently do not make their local metadata profiles public. They say: "the survey shows that the use of locally added homegrown metadata elements is allowed in nearly 70% of them. Only about one-fifth of local application profiles (19.6%) are made available online to the public. This means that not only is it difficult to create shareable metadata but also it is very difficult to have a quality assurance mechanism that is shareable beyond the local environment."

3 PROPOSAL

The rest of this paper discusses two ways in which records can be related to the cataloging schema used in their creation by the OLAC network of data providers. By granting access to metadata records via OLAC and access to the documentation for the metadata schemas at participating data providers, institutions support the flourishing of ethnolinguistic minority communities through metadata and language related artifacts, and they also support the scholarly networks which support them.

3.1 Modifying the data-provider description

The first way in which records can be related to the cataloging schema is to add an XML element with the source schema in the data provider's description record. The OAI-PMH implementation guidelines [21, §3.1] outline a series of optional containers. One container type provides information about a data provider. The OLAC-AP has implemented a container for providing identifying information about the data provider [34] as illustrated in Figure 2.

```
1   <description>
2   <olac-archive type="institutional" currentAsOf="
    YYYY-MM-DD"
3   xmlns="http://www.language-archives.org/OLAC
    /1.1/olac-archive"
4   xmlns:xsi="http://www.w3.org/2001/XMLSchema-
    instance"
5   xsi:schemaLocation="http://www.language-
    archives.org/OLAC/1.1/olac-archive
6   http://www.language-archives.org/OLAC/1.1/
    olac-archive.xsd">
7   <archiveURL>www.example.com</archiveURL>
8   <participant name="" title="" email="x@y.z"/>
9   <institution>Entity</institution>
10  <institutionURL>www.example.com</institutionURL
    >
11  <shortLocation>City, Country</shortLocation>
12  <location>Address</location>
13  <synopsis></synopsis>
14  <access></access>
15  <archivalSubmissionPolicy></
    archivalSubmissionPolicy>
16  </olac-archive>
17  </description>
```

Figure 2: OLAC-AP structure for the description of a data provider.

One approach to providing contextual information about the data provider's native metadata application profile is to modify this section of the OLAC-AP to include the title, version, and location of access for the native metadata application profile used by the data contributor. Following the patterns in the existing documentation, something like what is illustrated in Figure 3 would work.

```
1 <sourceMetadataApplicationProfile title="" version="
    documentationURL="" />
```

Figure 3: Placement of OLAC-AP content within the OAI record structure.

This method provides some basic access to the native metadata application profiles of OLAC data providers. This approach, however, has several drawbacks. For example, it does not specify at a record level which metadata schema or cataloging policy was current at the time a record was created. Cataloging policy can also affect metadata quality. However, if the proposed XML element were repeatable, then some change history would be accessible. A fourth attribute for dateActive="YYYY-MM-DD" would then indicate, in addition to the version number of the metadata profile, the dates a version was active. The method also does not address the change cycle in the metadata transformation technology if metadata is also transformed, which most is. A second repeatable element would be needed to track the metadata transformation technology life cycle. Good metadata application profile documentation should track changes, assigning version numbers to documentation versions and include dates of version changes within the documentation.

3.2 Record level association

Using OAI-PMH's built-in record provenance feature [21, §3.4] provides a second solution for addressing the documentation of the native data provider metadata application profile. This approach would require a modification to the current OLAC database. Current OLAC architecture harvests records via OAI-PMH but only writes certain fields and attributes to the SQL database from which the User Interface is driven. The current architecture disregards any data provider information supplied within an OAI-PMH <provenance> container. Unlike the first proposed solution, the second solution applies at the record level. The OAI-PMH <provenance> container has specific elements useful for tracking both changes within the record, for example those conducted by metadata utilities after harvesting but prior to display, and sources of the record [20]. OLAC data providers can use the <provenance> container to acknowledge archival deposit curation activities.

The two options presented need not be considered mutually exclusive. That is, they can be and likely should be used in concerts. The first one provides a general link to a presumably well documented metadata schema and the second one indicates record level provenance.

4 CONCLUSION

In considering the future of OLAC, Bird and Simons [4] say: "we hope to shift from an idiosyncratic community-specific infrastructure to a mainstream infrastructure that interoperates with the global Web of Data". By identifying and linking to data-provider application profiles and implementing provenance tracking for archival records, OLAC further connects with the global availability of bibliographic records. OLAC data, and the narrative of the data providers, moves towards greater transparency and interoperability. Altering OLAC infrastructure to support record level provenance will build upon OLAC's theme of openness. Well-formed provenance records can support a variety of scholarly activity metrics demonstrating scholarly effort.

Provenance recording and semantic inference can provide the mechanisms by which a metadata utility can engage with data providers to support their metadata curation processes at a scale they would not be able to achieve independently. Such a service changes the dynamics around involvement for data providers. Instead of simply providing metadata to OLAC, the ability to receive suggestions from a metadata utility can start to prompt data providers with record level nudges related to quality enhancements.

REFERENCES

[1] Steven Bird and Gary F. Simons. 2001. The OLAC Metadata Set and Controlled Vocabularies. In Proceedings of ACL/EACL Workshop on Sharing Tools and Resources for Research and Education, Thierry DeClerck, Steven Krauwer, and Mike Rosner (Eds.). EACL-ACL; elsnet, Université de Toulouse, France, 7–18. https://www.aclweb.org/anthology/W01-1506

[2] Steven Bird and Gary F. Simons. 2003. Extending Dublin Core Metadata to Support the Description and Discovery of Language Resources. Computers and the Humanities 37, 4 (2003), 375–388. https://doi.org/10.1023/A:1025720518994

[3] Steven Bird and Gary F. Simons. 2004. Building an Open Language Archives Community on the DC Foundation. In Metadata in Practice, Diane Hillmann and Elaine L. Westbrooks (Eds.). American Library Association, Chicago.

[4] Steven Bird and Gary F. Simons. 2021. Towards an Agenda for Open Language Archiving. In Proceedings of the International Workshop on Digital Language Archives: LangArc 2021, Oksana L. Zavalina and Shobhana Lakshmi Chelliah (Eds.). University of North Texas, Denton, Texas, 25–28. https://doi.org/10.12794/ langarc1851171.

[5] Steven Bird and Gary F. Simons. 2022. The Open Language Archives Community: A 20-Year Update. The Electronic Library 40, 5 (2022), 507–524. https://doi.org/10.1108/EL-08-2022-0192.

[6] Daan Broeder, Menzo Windhouwer, Dieter van Uytvanck, Thorsten Trippel, and Twan Goosen. 2012. CMDI: A Component Metadata Infrastructure. In Proceedings of Describing LRs with Metadata: Towards Flexibility and Interoperability in the Documentation of LR (A Worshop at LERC 2012), Victoria Arranz, Daan Broeder, Bertrand Gaiffe, Maria Gavrilidou, Monica Monachini, and Thorsten Trippel (Eds.). European Language Resources Association (ELRA), Istanbul, Turkey, 1–4. http://www.lrec-conf.org/proceedings/lrec2012/workshops/11.LREC2012%20Metadata%20Proceedings.pdf

[7] Daan Broeder and Peter Wittenburg. 2006. The IMDI Metadata Framework, Its Current Application and Future Direction. International Journal of Metadata, Semantics and Ontologies 1, 2 (2006), 119–132. https://doi.org/10.1504/IJMSO.2006.011008.

[8] Daan Broeder, Peter Wittenburg, and Onno Crasborn. 2004. Using Profiles for IMDI Metadata Creation. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), Maria Teresa Lino, Maria Francisca Xavier, Costa Rute, and Raquel Silvia (Eds.). European Language Resources Association (ELRA), Lisbon, Portugal. http://www.lrecconf.org/proceedings/lrec2004/pdf/513.pdf

[9] Thomas R. Bruce and Diane I. Hillmann. 2004. The Continuum of Metadata Quality: Defining, Expressing, Exploiting. ALA Editions. https://ecommons.cornell.edu/handle/1813/7895

[10] Mary Burke and Oksana L. Zavalina. 2020. Descriptive Richness of Free-text Metadata: A Comparative Analysis of Three Language Archives. Proceedings of the Association for Information Science and Technology 57, 1 (Oct. 2020). https://doi.org/10.1002/pra2.429

[11] CMDI Taskforce. 2016. Component Metadata Infrastructure (CMDI) Component Metadata Specification – Version 1.2. Metadata Specification CE-20160880. CLARIN. 37 pages. https://office.clarin.eu/v/CE-2016-0880-CMDI_12_ specification.pdf

[12] Diane I. Hillmann, Naomi Dushay, and Jon Phipps. 2004. Improving Metadata Quality: Augmentation and Recombination. In DC-2004–Shanghai Proceedings. Dublin Core Metadata Initiative — a project of ASIS&T, Shanghai, China, 11-14 October. https://dcpapers.dublincore.org/pubs/article/view/770

[13] Gary Holton. 2012. Language Archives: They're Not Just for Linguists Any More. In Potentials of Language Documentation: Methods, Analyses, and Utilization, Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts, and Paul Trilsbeek (Eds.). Number 3 in Language Documentation & Conservation Special Publication. University of Hawai'i Press, Honolulu, Hawai'i, 111–117. http://scholarspace.manoa.hawaii.edu/handle/10125/4523

[14] Baden Hughes. 2004. Metadata Quality Evaluation: Experience from the Open Language Archives Community. In Digital Libraries: International Collaboration and Cross-Fertilization, Zhaoneng Chen, Hsinchun Chen, Qihao Miao, Yuxi Fu, Edward Fox, and Ee-peng Lim (Eds.). Number 3334 in Lecture Notes in Computer Science. Springer, Berlin; Heidelberg, 320–329. https://doi.org/10.1007/978-3-540-30544-6_34.

[15] Heidi Johnson and Arienne Dwyer. 2002. Customizing the IMDI Metadata Schema for Endangered Languages. In Third International Conference on Language Resources and Evaluation: International Workshop on Resources and Tools in Field Linguistics, Las Palmas 26 - 27 May 2002, Peter Wittenburg (Ed.). LERC, Las Palmas, Canary Islands, Spain. https://www.mpi.nl/lrec/2002/papers/lrec-pap-05-JohnsonDwyer.pdf

[16] Peter Kiraly and Marco Buchler. 2018. Measuring Completeness as Metadata Quality Metric in Europeana. In 2018 IEEE International Conference on Big Data (Big Data). IEEE, Seattle, WA, USA, 2711–2720. https://doi.org/10.1109/BigData.2018.8622487

[17] Alex Klassmann, Freddy Ofenga, Daan Broeder, and Roman Skiba. 2006. IMDI Metadata Field Usage at MPI. Language Archives Newsletter 8 (2006), 6–8. https: //pure.mpg.de/rest/items/item_60279/component/file_60280/content

[18] Alex Klassmann, Freddy Offenga, Daan Broeder, Romuald Skiba, and Peter Wittenburg. 2006. Comparison of Resource Discovery Methods. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odijk, and Daniel Tapias (Eds.). European Language Resources Association (ELRA), Genoa, Italy. http://www.lrec-conf.org/proceedings/lrec2006/ pdf/808_pdf.pdf

[19] Carl Lagoze, Dean Krafft, Tim Cornwell, Naomi Dushay, Dean Eckstrom, and John Saylor. 2006. Metadata Aggregation and "Automated Digital Libraries": A Retrospective on the NSDL Experience. In Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '06). IEEE, Chapel Hill, NC, USA, 230–239. https://doi.org/10.1145/1141753.1141804

[20] Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. 2002. XML Schema to Hold Provenance Information in the "about" Part of a Record. In Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting (2.0 ed.). Open Archives Initiative. http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm

[21] Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. 2005. Guidelines for Optional Containers. In Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting (2.0 ed.). Open Archives Initiative. http://www.openarchives.org/OAI/2.0/guidelines.htm

[22] David Loshin. 2009. Identifying Master Metadata and Master Data. In Master Data Management, David Loshin (Ed.). Morgan Kaufmann, Boston, 129–142. https://doi.org/10.1016/B978-0-12-374225-4.00007-2

[23] David Loshin. 2011. Dimensions of Data Quality. In The Practitioner's Guide to Data Quality Improvement, David Loshin (Ed.). Morgan Kaufmann, Boston, 129–146. https://doi.org/10.1016/B978-0-12-373717-5.00008-7

[24] Joshua D. Lynch, Jessica Gibson, and Myung-Ja Han. 2020. Analyzing and Normalizing Type Metadata for a Large Aggregated Digital Library. The Code4Lib Journal 47 (Feb. 2020). https://journal.code4lib.org/articles/14995

[25] Andy Neale and Valentine Charles. 2020. MS68 Metis Strategic Recommendations M18: Aggregation Strategy. Europeana DSI-4. Europeana Foundation, The Hague, Netherlands. https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20DSI-4%20Aggregation%20Strategy.pdf

[26] Jeremy Nordmoe. 2011. Introducing RAMP: An Application for Packaging Metadata and Resources Offline for Submission to an Institutional Repository. In Proceedings of Workshop on Language Documentation & Archiving, SOAS, London, 18 November 2011., David Nathan (Ed.). SOAS, London, UK, 27–32. https: //www.sil.org/resources/archives/43211

[27] Jung-ran Park. 2006. Semantic Interoperability and Metadata Quality: An Analysis of Metadata Item Records of Digital Image Collections. Knowledge Organization 31, 1 (2006), 20–34.

[28] Jung-Ran Park and Yuji Tosaka. 2010. Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms. Cataloging & Classification Quarterly 48, 8 (Sept. 2010), 696–715. https://doi.org/10.1080/01639374.2010.508711

[29] Hugh J. Paterson III. 2022. An OLAC Perspective on Services: The Forgotten Language Resources. In Proceedings of DC-2022. Dublin Core Metadata Initiative, Online, Preprint. https://hughandbecky.us/Hugh-CV/talk/2022-servicesthe-forgotten-language-resources/DC_2022_Conference_Paper_Paterson_Revisions_pre-print.pdf

[30] Hugh J. Paterson III. 2022. Where Have All the Collections Gone?: Analysis of OLAC Data Contributors' Use of DCMIType 'Collection'. In Proceedings of the 15th Annual Society of American Archivists Research Forum, 21 July, 2021. Society of American Archivists, Chicago, IL. https://www2.archivists.org/am2021/ research-forum-2021/agenda#peer

[31] Julien Antoine Raemy. 2020. Enabling Better Aggregation and Discovery of Cultural Heritage Content for Europeana and Its Partner Institutions. Master of Science. Haute école de gestion de Genève, Geneva, Switzerland. https://julsraemy.ch/ assets/doc/Mastersthesis_europeana_raemyjulien_FV.pdf

[32] Sarah L. Shreeves, Ellen M. Knutson, Besiki Stvilia, Carole L. Palmer, Michael B. Twidale, and Timothy W. Cole. 2005. Is "Quality" Metadata "Shareable" Metadata? The Implications of Local Metadata Practices for Federated Collections. In Proceedings of the Twelfth National Conference of the Association of College and Research Libraries, April 7-10 2005, Minneapolis, MN, H. A. Thompson (Ed.). Association of College and Research Libraries, Chicago, IL, 223–237. http://hdl.handle.net/2142/145

[33] Gary F. Simons and Steven Bird. 2003. Building an Open Language Archives Community on the OAI Foundation. Library Hi Tech 21, 2 (2003), 210–218. https://doi.org/10.1108/07378830310479848

[34] Gary F. Simons and Steven Bird (Eds.). 2008. OLAC Repositories (2008-07-28 ed.). Open Language Archive Community, Dallas, TX. http://web.archive.org/web/20230418175450/http://www.language-archives.org/OLAC/repositories.html

[35] Besiki Stvilia, Les Gasser, Michael B. Twidale, and Linda C. Smith. 2007. A Framework for Information Quality Assessment. Journal of the American Society for Information Science and Technology 58, 12 (Oct. 2007), 1720–1733. https://doi.org/10.1002/asi.20652

[36] Christina Wasson, Melanie Medina, Miyoung Chong, Brittany LeMay, Emma Nalin, and Kenneth Saintonge. 2018. Designing for Diverse User Groups: Case Study of a Language Archive. Journal of Business Anthropology 7, 2 (Nov. 2018), 235–267. https://doi.org/10.22439/jba.v7i2.5605

[37] Anthony C. Woodbury. 2014. Archives and Audiences: Toward Making Endangered Language Documentations People Can Read, Use, Understand, and Admire. In Special Issue on Language Documentation and Archiving, David Nathan and Peter K. Austin (Eds.). Number 12 in Language Documentation and Description. SOAS, London, 19–36. http://www.elpublishing.org/PID/135

[38] Chuttur M. Yasser. 2011. An Analysis of Problems in Metadata Records. Journal of Library Metadata 11, 2 (April 2011), 51–62. https://doi.org/10.1080/19386389.2011.570654

[39] Irene Yi, Amelia Lake, Juhyae Kim, Kassandra Haakman, Jeremiah Jewell, Sarah Babinski, and Claire Bowern. 2022. Accessibility, Discoverability, and Functionality: An Audit of and Recommendations for Digital Language Archives. Journal of Open Humanities Data 8, 10 (March 2022), 1–19. https://doi.org/10.5334/johd.59 .

[40] Marcia Lei Zeng and Jian Qin. 2008. Metadata. Neal-Schuman Publishers, New York. http://catdir.loc.gov/catdir/toc/ecip0816/2008015176.html.