Hugh Paterson III. 2023. Diversity and Identity: Categories for OAI data-providers in the Open Language Archives Network. NASKO, Vol. 9. pp. 18-31.

Hugh Paterson III — University of North Texas & University of Oregon

Diversity and Identity: Categories for OAI data-providers in the Open Language Archives Network

Abstract

This work analyzes the network typology of data-providers who use the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) to engage in ethnolinguistic information-resource stewardship. The Open Language Archive Community's (OLAC) network is analyzed addressing: (1) the ontological nature of OAI data-providers, chiefly that not all data-providers are archives; (2) the classificatory nature of the data-providers in contrast to existing OLAC categories of *personal* and *institutional*; and (3) the impact of classification/description on the social-understanding about those providers. That is, discrete classificatory terminology does not exist within the target OLAC user community. A broader understanding of the classificatory distinctions among cultural heritage organizations would enable depositors to select the most appropriate institutions for cultural heritage preservation. Two classification taxonomies are presented for the data-providers. The taxonomy terms are applied to the members of the network: (1) as a lens by which one may understand metadata quality discrepancies across data-providers; (2) to identify strong and weak areas within the network; and (3) to identify network growth potential in contrast to the historically involved network participants. The developed taxonomies are applicable to cultural heritage networks outside of the set of OLAC data-providers and contribute to broader metadata quality discussions in the Library-Archive-Museum (LAM) community.

Introduction

Metadata quality across aggregated record sets and harvested record sets is a well discussed topic in the literature (Stvilia et al. 2004; Ward 2004; Bui and Park 2006; Park 2006; Palmer, Zavalina, and Fenlon 2010; Zavalina 2011; Palavitsinis, Manouselis, and Sanchez-Alonso 2014). Less well discussed is how network typologies of data-providers impact the reported results in metadata quality studies. Understanding network typologies in aggregate data contexts can have several benefits for network managers and other stakeholders. However, defining appropriate categories for data-provider network members can be challenging. Too few or too many categories and the narrative evidenced via the data analysis becomes difficult to interpret. This study looks at the 60 plus members of the Open Language Archive Community (OLAC) and proposes two taxonomies relevant to cultural heritage institutions stewarding language resources. This stands in contrast to an existing two-way distinction that the OLAC application profile (OLAC-AP) provides. By exploring the diversity of the networked data-providers, this study does three things. First it addresses an awareness gap among network participants related to who is involved. Second, it explores the classification of data-providers for the purpose of network health and growth potential. Third, by re-evaluating terminology used by the network of data-providers it allows for a more holistic discussion about the kinds of network stakeholders and their long-term roles related to language resources in the cultural heritage domain.

The classification of data-providers is an important business function across industries. For example, Bessembinder and colleagues (2019) discuss the classification of climate data-providers and the meteorological services enabled via their data sharing. Exegy (2019) classifies financial data-providers and the level of business services bundled with data access. In these contexts, the classification of data-providers is used to indicate the authority weight and inform business operations about the metadata quality. Within the context of research on cultural heritage institutions (CHI), the social classification or "framing" of providers has implications for research evaluating CHI information retrieval systems. That is, the social assent of categories and divisions of Knowledge Organization (KO), using Hjørland's (2008) broad sense of KO, impacts the interpretation of metadata records, which are in fact KO products in Hjørland's (2008) narrow sense. I propose that a theory of KO must account for types of information resources as well as types of business models (social functions) used by organizations in a larger KO ecosystem.

The presented taxonomies applied to data-providers are applicable not only to OLAC, which was the analyzed network, but also to other data-sharing networks in the cultural heritage space.

Background

OLAC is a federation of 60 plus data-providers sharing metadata records (Bird and Simons 2022) via the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH)[1] for consumption and display via a common aggregator[2]. OAI-PMH was designed to allow programmatic harvesting of metadata records (Shreeves, Kaczmarek, and Cole 2003). Typical applications of OAI-PMH include cross-institutional metadata harvests, metadata transmission between systems within the same institution, and networks of institutions contributing to a common data store. In the last case, the common data store is often coupled with a user interface for searching the collection of records from across institutional data-providers. This is the basic architecture behind OLAC and other large cultural heritage and scholarly communications discovery portals such as: *Digital Public Library of America* (DPLA)[3], *Europana*[4], *Directory of Open Access Journals*[5], and the *Platform for Open Data*[6].

The consumption of records from diverse types of data-providers suggests, at least in cases of metadata aggregation, that various KO practices (in the narrow sense) for resource description are brought together for display and engagement. These practices can and often do represent different kinds of KO approaches, e.g., impacting metadata quality assessment (Manghi, Candela, and Pagano 2010) or user experience design (Chopey 2005; Zavalina 2011, 2012). Therefore, it is important to account for these various local KO approaches when studying aggregated data; e.g., OLAC metadata as was presented by Paterson (2022).

The presented taxonomies clarify several complex conceptual contrasts. The first is the distinction between the terms *personal* and *institutional* as they are used in the OLAC-AP. Second is the social use of the term *archive* by language-scholars. Third are the terms *Data Provider*, *Service Provider* and *Repository* as used in the context of OAI-PMH.

Institutional vs personal

The OLAC-AP is based on both OAI-PMH and Dublin Core (Bird and Simons 2003, 2004, 2001; Simons and Bird 2003). The OAI-PMH schema has a container, `<description>`, which is used to describe the data-provider[7]. One option provided by the OAI-PMH documentation allows application profiles to further define a schema for use within the `<description>` container to provide network-specific information about the data-provider (Lagoze et al. 2005). Since its establishment in 2001, the OLAC-AP has defined and required this component's use

[1] https://www.openarchives.org/pmh
[2] http://search.language-archives.org
[3] https://dp.la
[4] https://www.europeana.eu/en
[5] https://doaj.org
[6] https://pod.stanford.edu
[7] The OAI-PMH documentation uses the term *repository*. For clarity throughout this paper, I use the term *data provider*.

by data-providers to identify their nature as either *personal* or *institutional* (Simons and Bird 2008, §3). The OLAC-AP defines the terms as follows:

> *Institutional* indicates that the repository is operated by an institution that is committed to maintaining it in the future, even after the individuals currently associated with it are no longer involved.
>
> *Personal* indicates that the repository is being operated by an individual (or a group of individuals) without the commitment of an institution for maintenance far into the future.

The OLAC-AP definition for *personal* implies that a collective of individuals should be classified as *personal*. This is counterintuitive based on the common usage definition of personal. These terminological choices have created a challenging situation to evaluate data-providers clearly and objectively. For example, would a department of colleagues providing data together in a single feed be *personal* or *institutional*? Similarly, would an ad-hoc network of researchers, such as the *Rift Valley Network*, be equally classified as *personal*? In the former case, a department seems to be closer to *institutional* than an unincorporated association of researchers, yet even a set of colleagues may not have the same lasting duration as an organization with a preservation mandate. Prior to work by Paterson (2021a), no data-provider self-identified as *personal*, yet several of the data-providers were clearly departmental in scope and maintained by a single person.

Archive

A second point of terminological confusion revolves around the term *archive*. Within OLAC documentation, all data-providers are discussed as *archives*. OLAC documentation drafters have used inclusive language choosing to cluster different types of institutional data-providers together. The choice maps well to the concepts of *open* and *community* found in the OLAC name, but the language of inclusion here does not acknowledge the diversity of the kinds of current or potential data-providers. OLAC's use of the term *archive* leads to a very interesting question: "What is an archive?" Is an archive an institution with a preservation mandate, as is commonly used in scholarly literature (Featherstone 2006; Seyfeddinipur et al. 2019; Burke et al. 2022; Matthews 2016)? Or is an archive a set of records often with a common origin and intra-record relationships, as is discussed by Duranti (1997) and others (Jenkinson 1937; Johnston and Schembri 2006)? The formative role that OLAC has had in the language-scholar community has strongly influenced the concept of *archive* among language-scholars.

Language-scholars use the term *archive* differently, e.g., some use it to mean a set of associated records with links (Ratner and MacWhinney 2016; Johnston and Schembri 2006), while others use the term in reference to an organization with a preservation mandate (Franchetto and Keren 2014; Skilton 2021)[8]. This indicates that, at least among language-scholars, there isn't a unified concept behind the term *archive*. This characterization of the language-scholar community is supported by results from a survey where 370 language-scholars responded to the question: "Have you archived your lexical dataset?" One hundred respondents replied "yes", but only 13 had made a deposit to a "long term stewardship institution" (Paterson III 2015).

Repository, data-provider, and service provider

The OAI-PMH documentation defines the terms *Data Provider*, *Service Provider*, and *Repository*. Within the OAI-PMH context, a corporate entity implementing data sharing technology is the *Data Provider*. The server technology implementing the access is called the

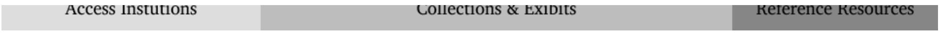
[8] A third usage also appears in the literature, *to archive*, meaning to make a submission to a collecting place.

Repository. Finally, *Service Providers* use metadata harvested via OAI-PMH as a basis for building value-added services. Clarifying the fine distinctions between the various technical and broader social use contexts are vital to terminological clarity within the presented taxonomies.

Methodology

To evaluate and classify data-providers, I investigated their descriptions as provided on the OLAC website[9]. I also considered their names and web presence. When I considered the social function each data-provider was attempting to fill, three broad functional but mutually exclusive categories emerged: (1) institutions which in some way provide access to resources; (2) collections and exhibits (displays); and (3) reference resources. I classified data-providers into the first group if they were institutions which stewarded resources which may be acquired under some access policy. In contrast, I placed data-providers that never possess resources into group three. That is, group three data-providers either merely list additional resources, or they are the resource, e.g., Several encyclopedia-like resources provide a record for each article within the larger work. Finally, the remaining data-providers were focused on interactive engagement with resources or telling a narrative about the resources. I put these data-providers into group two. In Figure 1, I arranged the three categories along a continuum where the left side is more likely to have the resource while the right side is less likely to have the actual resources described.

Figure 1. Taxonomy of broad categories.



The three emergent social functions each pertaining to a category are: *engagement with resources* (access institutions), *engagement with narrative* (collections and exhibits), and *engagement with facts* (reference resources).

Second Taxonomy

The preservation of resources beyond the efforts of a single person seems to be the major focus of the OLAC-AP *personal*/*institutional* dichotomy. The attempt to maintain this distinction motivated the exploration of a second finer-grained taxonomy. In the process of applying the taxonomy to OLAC data-providers it was realized that (1) an access organization might not be mandated to preserve content; and (2) that these three function-based categories are esoteric and may be challenging for practical use within the OLAC-AP because they are not directly applicable in ways that staff at data-providers can easily apply.

For added perspective on possible kinds of *Access Institutions* participating in OAI-PMH networks the list of providers to the DPLA was consulted[10]. As discussed later, the consultation increased the number of relevant organization types which might be data-providers. However, Libraries, Archives, and Museums (LAMs) constituted a significant portion of data-providers. The rise of digital libraries and metadata sharing has opened up many new conversations between LAMs (Zorich, Waibel, and Erway 2008; Waibel and Erway 2009; Tonta 2008; Matthews 2016; Ke 2016; Roy, Bhasin, and Arriaga 2011; Katre 2011). Many authors approach LAM/G(allery)LAM discussions from the perspective of collaboration and unification. However, Besser points out the traditionally divergent business models of these memory institutions which I found most useful while creating the second taxonomy (Dietz et al. 2005, 23).

[9] http://web.archive.org/web/20230418175254/http://www.language-archives.org/archives

[10] https://pro.dp.la/hubs/our-hubs

> Though libraries, museums, and archives all look like similar repositories housing cultural resources, there are some fundamental differences in mission, in what is collected, in how works are organized, and in how the institution relates to its users.
>
> The traditional library is based upon the individual item, which is generally not unique. Archives manage groups of works and focus on maintaining a particular context for the overall collection. Museums collect specific objects and provide curatorial context for each of them. These distinctions of the fundamental unit that is collected, affect each institution's acquisition policy, cataloging, preservation, and presentation to the public.
>
> Libraries and museums are both repositories, but libraries are user-driven. The role of the library is to provide access to a vast amount of material, which the user freely roams, making his/her own connections between works. Museums, historically, are curator-driven. They have only provided limited access to holdings, usually through a particular interpretative exhibition context, as provided by curatorial and educational staff. The museum provides a framework of context and interpretation, and the user can navigate within that smaller context. Archives tend to be research driven. They are accessible, often by appointment, in non-public spaces. The archivist has identified an area of the collection a researcher might be interested in, but s/he must go through it physically, item by item, to find out more information.

Thus, in the creation of the second taxonomy I adopted Besser's observations distinguishing the institutional characteristics of memory institutions. However, Besser's work pre-dates the widespread use of the term *digital repository* and as such doesn't directly address this important component of the current cultural preservation landscape. In considering how a digital repository (such as *Zenodo*[11] or *OSF*[12]) is different from any of the notions of Besser, I looked at the traditional content management practices in archives and compared them with the best practice OAIS model for data management in digital repositories (CCSDS 2012; and cf. Bel 2012). Archives have historically curated collections as a process of stewardship. As such archivists view the archival collection much like a glacier, slow to move, but changing. This stands in contrast to the OAIS model which calls for repositories to hold exact copies of content as submitted. Artifact curation does not occur within the conceptual model of repositories. With these distinctions in mind, the terms and characteristics outlined in Table 1 were adopted.

Table 1. Access Institutions

Taxonomy Term	***Characterization***
Archive	Preservation, Curation, Community re-engagement
Repository	Faithful distribution of deposits
Library	Patron driven, Temporary community access
Museum	Curator driven, Provides limited access to holdings, Interpretative exhibition
Gallery	Collaborates with creators to sell creative works

Table 1 also contains the term g*allery*. Several authors point to the business model as the critical difference between museums and galleries (Eden Gallery 2021; Hoerle-Guggenheim 2017). Eden Gallery states:

[11] https://zenodo.org

[12] https://osf.io

> Museums and art galleries use entirely different business models to fund their operating costs and make money. The simplified difference between an art gallery and a museum is that a museum is a place of entertainment; it's an activity to visit a museum. However, an art gallery is a business that displays and sells goods. An art gallery, like Eden Gallery, aims to raise the profile of artists who exhibit in its spaces and ultimately sell artworks.

While galleries and museums are listed in Table 1, their roles are notable because these institutions often display their stewarded resources within collections and exhibits. As institutions, their public interactions are dedicated to the interpretation of artifacts. Museums in their curator-driven capacity are unapologetically expressionistic, while galleries, with their sales-oriented business model, measure their success through outcomes grounded in impressionistic interpretations. These two audience-oriented natures in some ways overlap with the types of taxonomic resources in the broad category of collections and exhibits illustrated in Table 2.

Regarding other types of institutions, a review of DLPA data-providers revealed several additional entity types including *Networks*, *Centers*, *Publishers*, *Institutions with collection-specific management practice*, *Institutes*, *Historical Societies*, *Registries*, and *Services*. Each of these entity types, not to mention specific entities, differs in business model implementation and the characteristics by which information resources are stewarded. However, it was decided for this taxonomy that such entities were not viable for inclusion in-and-of themselves. Two reasons for this were: (1) data-providers may be sub-entities of organizations with these names; (2) often, organizations with these names truly do fit into the taxonomy terms available. Choosing a stewardship-oriented and function-based alignment allows data-providers the option to consider their most appropriate identifying term. For example, publishers most frequently align with behaviors consistent with repositories, even if they have a profit-centric model akin to galleries. Different corporate entities may operate a library or archive to the specific ends of the parent organization. It might be the library which is the data-provider rather than the larger organization which is responsible for the OAI-PMH data. Terms appearing in organizational titles can carry brand value rather than conforming with characterizations presented in Table 1. For example, some entities may call themselves a library or an archive but function like a repository.

Within the broad category *Collections and Exhibits,* the terms and characterizations listed in Table 2 were considered. These types of data-providers are focused on a specific narrative and the metadata often supports this goal in its original context. Institutional data-providers may manage several special collections or archives (archives in the sense of a coherent set of related records). However, it might also be the case that data-providers only manage a single collection. Such a collection may provide access to resources (like a repository does) or may only point to source locations (like a bibliography). Considering the *institutional* versus *personal* dynamic available via the OLAC-AP, and the kinds of digital collections appearing across the internet, four distinct taxonomy terms were chosen: *Special Collection*, *Personal Portfolio*, *Lab or Department Portfolio*, and *Project Portfolio*. Exhibits or collections generally have different kinds of arrangements. It follows then that they also have different types of metadata supporting their cohesion and navigation.

Table 2. Collections & Exhibits

Taxonomy Term	***Characterization***
Special Collection	Scope limited by topic or experience
Personal Portfolio	Scope limited to a single person's work
Lab or Department Portfolio	Scope limited to a single organizational unit
Project Portfolio	Scope limited to a single research endeavor

There are several reasons why the taxonomy terms in Table 2 deserve equal placement within a taxonomy also containing the terms in Table 1. First, many of these terms cover a use-case where an individual or community crafts an expressive statement via a collection of creative works. This sense of autonomy is often rescinded when content is committed to an institutional steward. Retaining the ability to craft the experience around resources is one reason scholars do not submit scholarly outputs to institutional stewards. Second, many of these cases would fall under the current OLAC-AP term *personal*, but they are not always *personal* to an individual. Third, there is a terminological ambiguity among language-scholars on how to refer to these types of exhibits and collections (cf. Paterson III 2021b, ftnt. 8). Even among information professionals, the types of collections created by language-scholars could be considered an *archive* in the sense of the term referring to a distinct set of records, which is ambiguous with the usage of *archive* referring to a type of stewardship institution. Content standards like *Describing Archives: A Content Standard* (Society of American Archivists 2013) provide guiding terminology for resolving these kinds of ambiguities. By acknowledging that more than just "archives" are data-providers the diversity of the contributor network is embraced. Also, by acknowledging diversity within the OLAC-AP, contributors must ask themselves if they have taken the necessary steps for long-term resource stewardship. This still allows for situations which express a great deal of autonomy and expressivity through the building of unique interactive narratives.

The third group in the broad taxonomy shown in Figure 1 is *Reference Resources*. As listed in Table 3, I found three kinds of resources which fit into this broad category. The first were encyclopedia-like resources that covered a range of languages, which in the case of the OLAC aggregator's user interface, is a specially indexed access point. The second type of resource was a list of other resources like bibliographies and discographies. I included the OAI-PMH concept of *gateway* within the term Bibliography. OAI-PMH gateways are specific network nodes which grant access to a dataset via another protocol such as Z39.50[13]. The third kind of resource is like the second in that it points to other resources. The term *aggregator* was chosen for these resources in contrast to gateways, which faithfully pass on metadata records. As used here, aggregators are nodes that operate in conjunction with a bibliographic utility (Hillmann 2008; Hillmann, Dushay, and Phipps 2004) which does data transformation or consistency alterations to the record. Both Europeana (Raemy 2020; Neale and Charles 2020) and some nodes within the DPLA network (Lynch and Gibson 2019; Lynch, Gibson, and Han 2020) operate this way. These types of entities do not currently exist in the OLAC network but could. A forward-looking taxonomy should account for these kinds of functions.

[13] http://www.openarchives.org/OAI/2.0/guidelines-gateway.htm

Table 3. Reference Resorces

Taxonomy Term	***Characterization***
Encyclopedic Resource	A topical reference work covering a range of special access points
Bibliography	A list of items
Aggregator	A list of items sourced from other data-providers

Discussion

Table 4 presents a breakdown of the OLAC data-providers applying the two taxonomies. By applying the taxonomies to the network providers, it shows that a significant number of data-providers are *repositories* and *encyclopedic resources*. The distribution suggests that there is current value being realized by the producers of encyclopedic resources. If we assume that publishers most naturally fit into the category of repository, then the number of repositories participating in the network should be about two orders of magnitude higher than the current numbers. It also suggests that curation activities for language resources are not likely to happen among OLAC networks participants. Given the large number of print resources in libraries, greater participation in the network by these types of institutions would add significant value to the network and serve to increase awareness around language resources.

Table 4. Analysis of OLAC data-providers

Taxonomy Term	**Broad Category**	***Instances in the OLAC Network***
Archive	Access institution	17
Repository	Access institution	20
Museum	Access institution	0
Gallery	Access institution	0
Library	Access institution	6
Special Collection	Collections & Exhibits	1
Personal Portfolio	Collections & Exhibits	1
Lab or Department Portfolio	Collections & Exhibits	2
Project Portfolio	Collections & Exhibits	1
Encyclopedic Resource	Reference Resource	14
Bibliography	Reference Resource	1
Aggregator	Reference Resource	0

The scope of the data coverage informs the classification of the data-provider. For example, if the data is of the catalog for the entire institution, then classify according to one of the values in the *access institution* group; if the data provided from an institution is scoped to a specific collection, then classify according to a term within the *collection and exhibit* group. For institutions with more than one collection but not a whole institution's catalog there are at least two options: first the use of multiple OAI-PMH data feeds, or second, the use of the listSet/setSpec mechanism per the OAI-PMH specification. Currently, some OLAC data-providers do provide setSpec data, however, the OLAC-AP and aggregator do not specify or make use of this data. Specifying the use and scope of the OAI-PMH data feed description along with setSpec use would clarify the OLAC-AP for situations where resource stewards such as the *Pacific And Regional Archive for Digital Sources In Endangered Languages* (PARADISEC) and SIL International's *Language & Culture Archives* provide records for resources in their holdings as well as records for resources they know about but are not specifically in their holdings.

The classification of current data-providers suggests that efforts to include more data-providers within the network ought to consider systemic approaches for including data-providers from some of the underrepresented categories. Efforts to include portfolio collections would support the scholarly profiles of scholars and research units. Using appropriate terminology when referencing data-providers serves to: (1) clarify expectations around the long-term availability of resources—especially those representing the ethnolinguistic heritage of endangered language communities. And (2) clarify user experience design research related to information retrieval systems for language resources. Within the field of language-resource stewardship and language scholarship, the different senses of the term *archive* have been conflated, resulting in impacts on reported research results. For example, Yi et al. (2022) present a comparison of "language archives" and their web-facing user interfaces without differentiating the kind of information resource a website presents, e.g., institutions with diverse collections versus a single corpus. They also further conflate websites presenting either sense of *archive* with actual digital asset management infrastructure. The framing of their analysis suggests that the website *is* the archive.

Digital infrastructure facilitating asset storage, discovery, and acquisition is actually a complex construct which varies from implementation to implementation. In contrast to Yi et al.'s (2022) 'equal treatment' of different kinds of web facing entitles in the name of 'language archives', Ferreira et al. (2021) argue that archives and digital displays for interactions are distinct types of entities. They denote specific kinds of purposes for websites highlighting the fact that some websites do not structure their existence around a preservation mandate; in a sense they are ephemeral, even as they provide meaningful community access to resources. These ephemeral websites are often community produced exhibits or the presentation of the scholarly outputs of research labs. Although these websites are not archives in the preservation institution sense, they can equally be data-providers to OLAC for aggregation and increased awareness of the language resources available.

The framing of Yi and colleges' work demonstrates that some language-scholars' understanding of an archive is related to the language-scholar's experience of it—a nod to digital materiality (and cf. Manoff 2006; Leonardi 2010; Jung and Stolterman 2012; Pink, Ardévol, and Lanzeni 2016). It also suggests that a scholar's understanding of preservation *is* the ability to access resources. These are two important and under-explored components in developing successful cultural preservation workflows which involve scholar-driven accessions and descriptions.

Conclusion

The result of this study was that a 12-term taxonomy was developed. When applied, it can be used to better understand the OAI-PMH data-provider network. A more diverse typology of data-providers within the OLAC application profile would serve user communities more effectively and impact the social perspective on stewardship organizations and access channels. The utility of the taxonomy can be realized in contexts beyond the Open Language Archives

Community. By acknowledging diversity, reasonable expectations by language-scholars and data users can be established.

The taxonomy serves at least three functions. First, it allows for a gap analysis by revealing the kinds of OLAC data-providers which have found value through participation and the kinds of data-providers around which possible network-growth opportunities exist. Second, it allows for a more useful metadata quality evaluation by grouping like contributors together. Third, it raises awareness among language-scholars and metadata specialists concerning the differences between data-providers.

Acknowledgments

I am grateful to Becky Paterson and Paul Unger for proof reading comments. I am also indebted to Nick Thieberger, Steven Bird, and anonymous reviewers for thought provoking and encouraging comments. All errors are unfortunate and are my own.

References

Bel, Bernard. 2012. “Implementing the OAIS for Oral/Linguistic Resources: The Speech and Language Data Repository Venture.” In *Journées OAIS*. Lyon, France. https://hal.archives-ouvertes.fr/hal-01053214.

Bessembinder, Janette, Marta Terrado, Chris Hewitt, Natalie Garrett, Lola Kotova, Mauro Buonocore, and Rob Groenland. 2019. “Need for a Common Typology of Climate Services.” *Climate Services* 16 (December): 100135. doi:10.1016/j.cliser.2019.100135.

Bird, Steven, 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*, edited by Thierry DeClerck, Steven Krauwer, and Mike Rosner, 7–18. Université de Toulouse, France: EACL-ACL; elsnet. https://www.aclweb.org/anthology/W01-1506.

Bird, Steven, and Gary F. Simons. 2003. “Extending Dublin Core Metadata to Support the Description and Discovery of Language Resources.” *Computers and the Humanities* 37 (4): 375–88. doi:10.1023/A:1025720518994.

Bird, Steven, and Gary F. Simons. 2004. “Building an Open Language Archives Community on the DC Foundation.” In *Metadata in Practice*, edited by Diane Hillmann and Elaine L. Westbrooks. Chicago: American Library Association.

Bird, Steven, and Gary F. Simons. 2022. “The Open Language Archives Community: A 20-Year Update.” *The Electronic Library* 40 (5): 507–24. doi:10.1108/EL-08-2022-0192.

Bruce, Thomas R., and Diane I. Hillmann. 2004. “The Continuum of Metadata Quality: Defining, Expressing, Exploiting.” In *Metadata in Practice*, edited by Diane I. Hillmann and Elaine L. Westbrooks, 238–56. Chicago, IL: ALA Editions.

Bui, Yen, and Jung-ran Park. 2006. “An Assessment of Metadata Quality: A Case Study of the National Science Digital Library Metadata Repository.” In *Information Science Revisited: Approaches to Innovation, CAIS/ACSI 2006 Proceedings of the 2006 Annual Conference of the Canadian Association for Information Science, York University, Toronto, Ontario. June 1 - 3, 2006*, edited by Haidar Moukdad. Toronto, Ontario: CAIS/ACSI. doi:10.29173/cais166.

Burke, Mary, Oksana L. Zavalina, Shobhana L. Chelliah, and Mark E. Phillips. 2022. “User Needs in Language Archives: Findings from Interviews with Language Archive Managers, Depositors, and End-Users.” *Language Documentation & Conservation* 16: 1–24. http://hdl.handle.net/10125/74669.

CCSDS. 2012. “Reference Model for an Open Archival Information System (OAIS).” Recommended Practice, Issue 2 CCSDS 650.0-M-2. Magenta Book. Washington, DC, USA: Management Council of the Consultative Committee for Space Data Systems. https://public.ccsds.org/pubs/650x0m2.pdf.

Chopey, Michael A. 2005. “Planning and Implementing a Metadata-Driven Digital Repository.” Cataloging & Classification Quarterly 40 (3–4). Routledge: 255–87. doi:10.1300/J104v40n03_12.

Dietz, Steve, Howard Besser, Ann Borda, Kati Geber, and Pierre Lévy. 2005. “Virtual Museum (of Canada): The Next Generation.” Ottawa, Ontario: Canadian Heritage Information Network. https://www.academia.edu/34788702/Steve_Dietz_Howard_Besser_Ann_Borda_and_Kati_Geber_eds_2005_Virtual_Museum_of_Canada_the_Next_Generation_Canadian_Heritage_Information_Network.

Duranti, Luciana. 1997. “The Archival Bond.” *Archives and Museum Informatics* 11 (3–4): 213–18. doi:10.1023/A:1009025127463.

Eden Gallery. 2021. “Art Galleries vs Museums: What’s The Difference?” *Eden Gallery*. https://www.eden-gallery.com/news/art-galleries-vs-museums.

Exegy, Inc. 2019. “What Types of Financial Data Providers Are There?” *Insights*. Exegy, Inc. July 5. https://www.exegy.com/types-of-financial-data-providers.

Featherstone, Mike. 2006. “Archive.” *Theory, Culture & Society* 23 (2–3): 591–96. doi:10.1177/0263276406023002106.

Ferreira, Vera, Leonore Lukschy, Buachut Watyam, Siripen Ungsitipoonpor, and Mandana Seyfeddinipur. 2021. “A Website Is a Website Is a Website: Why Trusted Repositories Are Needed More Than Ever.” In *Proceedings of the International Workshop on Digital Language Archives: LangArc 2021*, edited by Oksana L. Zavalina and Shobhana Lakshmi Chelliah, 1–4. Denton, Texas: University of North Texas. doi:10.12794/langarc1851176.

Franchetto, Bruna, and Rice Keren. 2014. “Language Documentation in the Americas.” *Language Documentation & Conservation* 8: 251–61. http://hdl.handle.net/10125/24606.

Hillmann, Diane I. 2008. “Metadata Quality: From Evaluation to Augmentation.” *Cataloging & Classification Quarterly* 46 (1): 65–80. doi:10.1080/01639370802183008.

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

Hjørland, Birger. 2008. “What Is Knowledge Organization (KO)?” *Knowledge Organization* 35 (2–3): 86–101. doi:10.5771/0943-7444-2008-2-3-86.

Hoerle-Guggenheim, Philippe. 2017. “The Difference Between an Art Gallery and a Museum.” *Medium*. July 28. https://medium.com/@PhilippeHG_NYC/the-difference-between-an-art-gallery-and-a-museum-613b0db6353f.

Hughes, Baden. 2004. “Metadata Quality Evaluation: Experience from the Open Language Archives Community.” In *Digital Libraries: International Collaboration and Cross-Fertilization*, edited by Zhaoneng Chen, Hsinchun Chen, Qihao Miao, Yuxi Fu, Edward Fox, and Ee-peng Lim, 320–29. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-30544-6_34.

Jenkinson, Hilary. 1937. *A Manual of Archive Administration*. London: P. Lund, Humphries & Co., Ltd. http://archive.org/details/manualofarchivea00iljenk.

Johnston, Trevor, and Adam Schembri. 2006. “Issues in the Creation of a Digital Archive of a Signed Language.” In *Sustainable Data from Digital Fieldwork: Proceedings of the Conference Held at the University of Sydney, 4-6 December 2006*, edited by Linda Barwick and Nicholas Thieberger, 7–16. Sydney, Australia: Sydney University Press. https://ses.library.usyd.edu.au/handle/2123/1289.

Jung, Heekyoung, and Erik Stolterman. 2012. “Digital Form and Materiality: Propositions for a New Approach to Interaction Design Research.” In *Proceedings of the 7th Nordic Conference on Human-*

Computer Interaction Making Sense Through Design - NordiCHI '12, 645. Copenhagen, Denmark: ACM Press. doi:10.1145/2399016.2399115.

Katre, Dinesh. 2011. "Digital Preservation: Converging and Diverging Factors of Libraries, Archives and Museums – an Indian Perspective." *IFLA Journal* 37 (3): 195–203. doi:10.1177/0340035211418728.

Ke, Hao-Ren. 2016. "Fusion of Library, Archive, Museum, Publisher (LAMP): The NTNU Library Experience." *Journal of Information Science Theory and Practice* 4 (2): 66–74. doi:10.1633/JISTAP.2016.4.2.5.

Lagoze, Carl, 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. Open Archives Initiative. http://www.openarchives.org/OAI/2.0/guidelines.htm.

Leonardi, Paul M. 2010. "Digital Materiality? How Artifacts without Matter, Matter." *First Monday* 15 (6). doi:10.5210/fm.v15i6.3036.

Lynch, Joshua D., and Jessica Gibson. 2019. "Analyzing and Normalizing Illinois Digital Heritage Hub Type Metadata,". https://www.ideals.illinois.edu/handle/2142/104611.

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

Manghi, Paolo, Leonardo Candela, and Pasquale Pagano. 2010. "Interoperability Patterns in Digital Library Systems Federations." In *Pre-Proceedings of the 2nd DL.Org Workshop—Making Digital Libraries Interoperable: Challenges and Approaches*, edited by Donatella Castelli, Yannis Ioannidis, and Seamus Ross, 67–76. Glasgow, Scotland: European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2010). http://www.dlorg.eu/uploads/Booklets/2nd%20Workshop%20Proceedings/Pre-proceedings-1.pdf#page=73.

Manoff, Marlene. 2006. "The Materiality of Digital Collections: Theoretical and Historical Perspectives." *Portal: Libraries and the Academy* 6 (3): 311–25. doi:10.1353/pla.2006.0042.

Matthews, Joseph R. 2016. *Adding Value to Libraries, Archives, and Museums: Harnessing the Force That Drives Your Organization's Future*. Santa Barbara, California; Denver, Colorado: Libraries Unlimited, an imprint of ABC-CLIO, LLC.

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

Palavitsinis, Nikos, Nikos Manouselis, and Salvador Sanchez-Alonso. 2014. "Metadata Quality in Digital Repositories: Empirical Results from the Cross-Domain Transfer of a Quality Assurance Process." *Journal of the Association for Information Science & Technology* 65 (6): 1202–16. doi:10.1002/asi.23045.

Palmer, Carole L., Oksana L. Zavalina, and Katrina Fenlon. 2010. "Beyond Size and Search: Building Contextual Mass in Digital Aggregations for Scholarly Use." *Proceedings of the American Society for Information Science and Technology* 47 (1): 1–10. doi:10.1002/meet.14504701213.

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

Paterson III, Hugh J. 2015. "Lexical Dataset Archiving: An Assessment of Practice." Poster presented at the 4th International Conference on Language Documentation & Conservation, February 26-March 1, 2015, at the Ala Moana Hotel in Honolulu. https://hughandbecky.us/Hugh-CV/publication/2015-lexical-database-archiving.

Paterson III, Hugh J. 2021a. “From CV to OLAC.” Presentation at the 7th International Conference on Language Documentation & Conservation (ICLDC), 4–7 March 2021, University of Hawaiʻi at Mānoa. https://hughandbecky.us/Hugh-CV/talk/2021-from-cv-to-olac.

Paterson III, Hugh J. 2021b. “Language Archive Records: Interoperability of Referencing Practices and Metadata Models.” M.A. Thesis, Grand Forks, North Dakota: University of North Dakota. Theses and Dissertations. 3937. University of North Dakota Scholarly Commons. https://commons.und.edu/theses/3937.

Paterson III, Hugh J. 2022. “Analysis of OLAC Data Contributors’ Use of DCMIType ‘Collection.’” In *Proceedings of the 15th Annual Society of American Archivists Research Forum*. Chicago, IL: Society of American Archivists. https://www2.archivists.org/am2021/research-forum-2021/agenda#peer.

Pink, Sarah, Elisenda Ardévol, and Débora Lanzeni. 2016. “Digital Materiality.” In *Digital Materialities: Design and Anthropology*, edited by Sarah Pink, Elisenda Ardévol, and Débora Lanzeni, 1–26. London: Bloomsbury Academic. http://site.ebrary.com/id/11135397.

Raemy, Julien Antoine. 2020. “Enabling Better Aggregation and Discovery of Cultural Heritage Content for Europeana and Its Partner Institutions.” M.S. Thesis, Geneva, Switzerland: Haute école de gestion de Genève. https://julsraemy.ch/assets/doc/Mastersthesis_europeana_raemyjulien_FV.pdf

Ratner, Nan, and Brian MacWhinney. 2016. “Your Laptop to the Rescue: Using the Child Language Data Exchange System Archive and CLAN Utilities to Improve Child Language Sample Analysis.” *Seminars in Speech and Language* 37 (02): 074–084. doi:10.1055/s-0036-1580742.

Roy, Loriene, Anjali Bhasin, and Sarah K. Arriaga, eds. 2011. *Tribal Libraries, Archives, and Museums: Preserving Our Language, Memory, and Lifeways*. Lanham: Scarecrow Press.

Seyfeddinipur, Mandana, Felix Ameka, Lissant Bolton, Jonathan Blumtritt, Brian Carpenter, Hilaria Cruz, Sebastian Drude, et al. 2019. “Public Access to Research Data in Language Documentation: Challenges and Possible Strategies.” *Language Documentation & Conservation* 13: 545–63. http://hdl.handle.net/10125/24901.

Shreeves, Sarah L., Joanne S. Kaczmarek, and Timothy W. Cole. 2003. “Harvesting Cultural Heritage Metadata Using the OAI Protocol.” *Library Hi Tech* 21 (2): 159–69. doi:10.1108/07378830310479802.

Simons, Gary F., and Steven Bird. 2003. “Building an Open Language Archives Community on the OAI Foundation.” *Library Hi Tech* 21 (2): 210–18. doi:10.1108/07378830310479848.

Simons, Gary F., and Steven Bird, eds. 2008. “OLAC Repositories.” Open Language Archive Community. http://www.language-archives.org/OLAC/repositories.html.

Skilton, Amalia. 2021. “Ticuna (Tca) Language Documentation: A Guide to Materials in the California Language Archive.” *Language Documentation & Conservation* 15: 153–89. http://hdl.handle.net/10125/24972.

Society of American Archivists. 2013. *Describing Archives: A Content Standard*. 2nd ed. Chicago, Illinois: Society of American Archivists. http://files.archivists.org/pubs/DACS2E-2013_v0315.pdf.

Stvilia, Besiki, Les Gasser, Michael B Twidale, Sarah L Shreeves, and Tim W Cole. 2004. “Metadata Quality for Federated Collections.” In *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)*, 111–25. https://www.ideals.illinois.edu/handle/2142/721.

Tonta, Yaşar. 2008. “Libraries and Museums in the Flat World: Are They Becoming Virtual Destinations?” *Library Collections, Acquisitions, and Technical Services* 32 (1): 1–9. doi:10.1016/j.lcats.2008.05.002.

Waibel, Günter, and Ricky Erway. 2009. "Think Globally, Act Locally: Library, Archive, and Museum Collaboration." *Museum Management and Curatorship* 24 (4). Routledge: 323–35. doi:10.1080/09647770903314704.

Ward, Jewel Hope. 2004. "Unqualified Dublin Core Usage in OAI-PMH Data Providers." *OCLC Systems & Services: International Digital Library Perspectives* 20 (1): 40–47. doi:10.1108/10650750410527322.

Yi, Irene, 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): 1–19. doi:10.5334/johd.59.

Zavalina, Oksana L. 2011. "Contextual Metadata in Digital Aggregations: Application of Collection-Level Subject Metadata and Its Role in User Interactions and Information Retrieval." *Journal of Library Metadata* 11 (3–4): 104–28. doi:10.1080/19386389.2011.629957.

Zavalina, Oksana L. 2012. "Subject Access: Conceptual Models, Functional Requirements, and Empirical Data." Journal of Library Metadata 12 (2–3): 140–63. doi:10.1080/19386389.2012.699829.

Zorich, Diane, Günter Waibel, and Ricky Erway. 2008. "Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums." OCLC Research. doi:10.25333/X187-3W53.