Dublin Core's DCMIType ‘PhysicalObject’ and its use across the Open Language Archives Community

HUGH PATERSON III

University of North Texas

Abstract: This study explores the composition of linguistic and anthropological language-focused artifact records which use the DCMIType term ‘PhysicalObject’. However, the results are broadly applicable to all users of Dublin Core. Dublin Core's DCMIType vocabulary is an important access point for resource discovery. It allows access to resources based on experiential modality types, e.g., Moving Image, Sound, Text, Still Image, Software, etc. Previous research reporting on DCMIType ‘PhysicalObject’ suggests that it is challenging for information professionals to apply consistently. This study affirms that the semantics of ‘PhysicalObject’ can be confusing with regard to non-digital resources. Further, the term’s DC provided definition is limiting due to its applicability only to inanimate objects. Over four hundred and fifty thousand records representing sixty-one data providers from the Open Language Archives Community (OLAC) were analyzed. Across the range of OLAC providers, currently only sixteen records use the DCMIType term ‘PhysicalObject’.

Untapped potential exists for the use of ‘PhysicalObject’ by language-scholars to describe their scholarly work. There are opportunities for greater technical descriptions of language documentation collections. Cultural heritage stewards could integrate records of physical objects with existing OLAC records. A richer understanding of the applicability of ‘PhysicalObject’ within repository records can lead to a more diverse participation in cultural heritage aggregators. In the case of the OLAC aggregator, richer records benefit ethnolinguistic minorities who are seeking artifacts and information pertinent to their cultural heritage.

Introduction

This study investigates the semantics, utility, and use of the DCMIType vocabulary term *PhysicalObject*. The DCMIType vocabulary is frequently used within Qualified Dublin Core (QDC) metadata schemas to indicate the modality, materiality, and interactive qualities of a described resource. Since 1999, Dublin Core has increased in popularity. It is commonly positioned as the default metadata schema in open source digital library software, e.g., DSpace. Therefore, due to its simplistic nature and its default position in repository software, the Dublin Core Metadata Schema sees a wide range of deployments in digital archives, digital libraries, and metadata record aggregators. This discussion is centered around metadata schema’s in use in language archives, but it has implications further afield.

In the following sections, I look at the semantics, utility, and use of *PhysicalObject,* one of the twelve terms within the DCMIType vocabulary.[1] This study is based on aggregated language resource metadata

[1] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-7

Published by Society of American Archivists, March 2024.

records from the Open Language Archive Community (OLAC). OLAC's application profile is based on QDC (Bird and Simons 2003, 2004) and is therefore a valid dataset to investigate the current use of *PhysicalObject* across more than sixty data providers (Simons and Bird 2003b; Bird and Simons 2022).[2] The current study is unique not just in regards to the appraisal and evaluation of OLAC data providers, but it also addresses a very rarely discussed term of the DCMIType vocabulary.

The results of this study show that the use of *PhysicalObject* is infrequent, misunderstood/misapplied, and underutilized. A great opportunity exists to use *PhysicalObject* to meet previously articulated goals for metadata related to language resource discovery. Scholars at the E-MELD workshops[3] in the early 2000s overlooked the potential to use *PhysicalObject* in the metadata records documenting the workflows of language scholars even though participants specifically mentioned that they were interested in metadata about the physical devices used in the creation of audio and video artifacts.

Literature Review

The current study builds upon previous studies of other DCMIType vocabulary terms using the same data sources (Paterson III 2023a, 2022). The current analysis addresses a literature gap in two areas: first, the evaluation of records provided to the Open Language Archive Community (OLAC) aggregator related to physical objects; and second, the scholarly discussion of Dublin Core's DCMIType vocabulary value *PhysicalObject*.[4] Scholarly discourse around the DCMIType vocabulary itself falls into two major areas: first, the creation of the vocabulary and the semantics of the included terms; second, assessments around the use of the vocabulary and its terms. A review of cataloging and description practices for physical objects more broadly by memory institutions is out of scope.

The structure and evaluation of metadata records describing language resources within stewardship organizations has a long history (Bird and Simons 2001; Johnson and Dwyer 2002; Simons and Bird 2003a; Bird and Simons 2003; Hughes 2004; Broeder et al. 2012; Klassmann et al. 2006; Broeder and Wittenburg 2006) but has seen a recent increase of attention (Aljalahmah and Zavalina 2023; Burke et al. 2022; Huber 2023; Paterson III In Press, 2023a, 2023b, 2023c, 2023d, 2022, 2021a). Memory institutions have institution-specific artifact description practices and bespoke metadata application profiles to support institutional practices (Burke and Zavalina 2019, 2020; Burke et al. 2020). Aggregators are websites performing a service. They take transformations sourced from unique application profiles and provide a common interface with specific points of entry for select audiences. By studying records provided to an aggregator such as the OLAC aggregator (Hughes 2004; Paterson III 2023b, 2023c, 2023a, 2022), broad trends can be investigated about how these institutions understand their metadata and how they value the engagement of audiences via the information provided within the aggregator.

The DCMIType term *PhysicalObject* is strikingly absent from most of the scholarly literature reporting on Dublin Core use and the DCMIType vocabulary (Park 2006, 2009; Zavalina 2011; Park and Childress 2009; Ward 2002, 2004, 2003). This absence is not without explanation. Dublin Core is for the most part

[2] http://language-archives.org
[3] https://web.archive.org/web/20050209105520/http://e-meld.org
[4] http://purl.org/dc/dcmitype/PhysicalObject

used in *digital* repositories—stewarding digital objects which are often assumed to have *no physical characteristics*.

In the first known study to focus on the use of the DCMIType term *PhysicalObject*, Zavlin and Zavalina (2023) report on how students in an information and library science program negotiate the description of physical paintings—image resources with physical characteristics (carriers). They point out that students sometimes want to classify physical paintings with the DCMIType term *PhysicalObject* rather than *StillImage.* Assessing the validity of a record classifying a painting as a *PhysicalObject* in contrast to the DCMIType value *StillImage* raises issues grounded in the semantics and definitions of the DCMIType vocabulary terms.

The specific definition provided by the DCMI Usage Board and the illustrative comments accompanying the definition of *PhysicalObject* have changed over the years. For example, Guenther (1999) documents an early sketched definition as:

> **Physical Object**: a non-human object or substance. This category includes objects that do not fit into any of the other categories on this list. In addition these objects must be approached physically to make use of them. For example - a computer, the great pyramid, a sculpture, wheat. Note that digital representations of, or surrogates for, these things should use image, text or one of the other types.

This definition is noteworthy for two reasons. First, it shows that the category was to be a catch all after other logical options were ruled out. Second, it clearly draws a line between human and non-human. I return to the non-human point in the Discussion section. A more recent version of the the definition from the DCMI[5] states:

> **Definition [version 003]**[6]: An inanimate, three-dimensional object or substance.
> **Comment**: Note that digital representations of, or surrogates for, these objects should use Image, Text or one of the other types.

In the Discussion section, I discuss some of the ambiguities around the current definition and its semantics.

The Study and Data

There were several motivations for this study. First, I wanted to investigate the possibility that the claims of Zavlin and Zavalina (2023) were having a real world impact. Second, there is a lack of scholarly resources describing use-cases wherein *PhysicalObject* has been tested or used. Third, I wanted to investigate, and theorize how the DCMIType value *PhysicalObject* could be used when searching through metadata records representing language resources. Many application profiles, such as the OLAC application profile, do not provide illustrative examples that include the use of DCMIType vocabulary terms. By getting a better understanding of actual use, mis-use, and possible alternatives, one gains a

[5] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/dcmitype/PhysicalObject

[6] As of this writing, this is the current version. I have labeled it version 003 as it is the third version I have found published by DCMI.

better understanding of the implicit theory underlying the DCMIType vocabulary and the resulting impacts.

The OLAC aggregator makes its records publicly accessible via individual HTML pages and zip files of the entire aggregation.[7] At the time of the research (April-June 2023) a faceted search tool implemented by Simons and Bird (2011) was also available for navigating OLAC records.[8] Since the time this research was originally conducted, the faceted browsing capabilities of the OLAC aggregator have been deprecated and are no-longer accessible. However, the faceted browsing capabilities including the cross-data-provider query for DCMIType *PhysicalObject* is shown in Figure 1. For reproducibility, the data collected during this research are available via Zenodo as Paterson (2023e).

The data was collected and analyzed via individual record inspection based on record availability on the 23rd of April 2023. At that time, there were 480,812 total metadata records (across 61 data providers). Only 16 records used the DCMIType value *PhysicalObject*. A summary is presented in Table 1. These 16 records were provided by only three data providers.

The Speech and Language Data Repository (SLDR)[9] applied the term to an XML data set. The California Language Archive: Berkeley (CLA), used the term on metadata records containing samples of flora, a preserved goldfinch, and beetles. The American Philosophical Society (APS) applied the term to records whose associated artifacts included: plant specimens, potsherds of a canteen pot, and an arrow head. These are further discussed in the next section.

[7] Since 2021, the server has experienced reliability issues. The latest known dataset for the entire aggregation is available via Zenodo (Paterson III 2021b).

[8] http://search.language-archives.org

[9] SLDR seems to have been folded into ORTOLANG. https://www.ortolang.fr/en/home

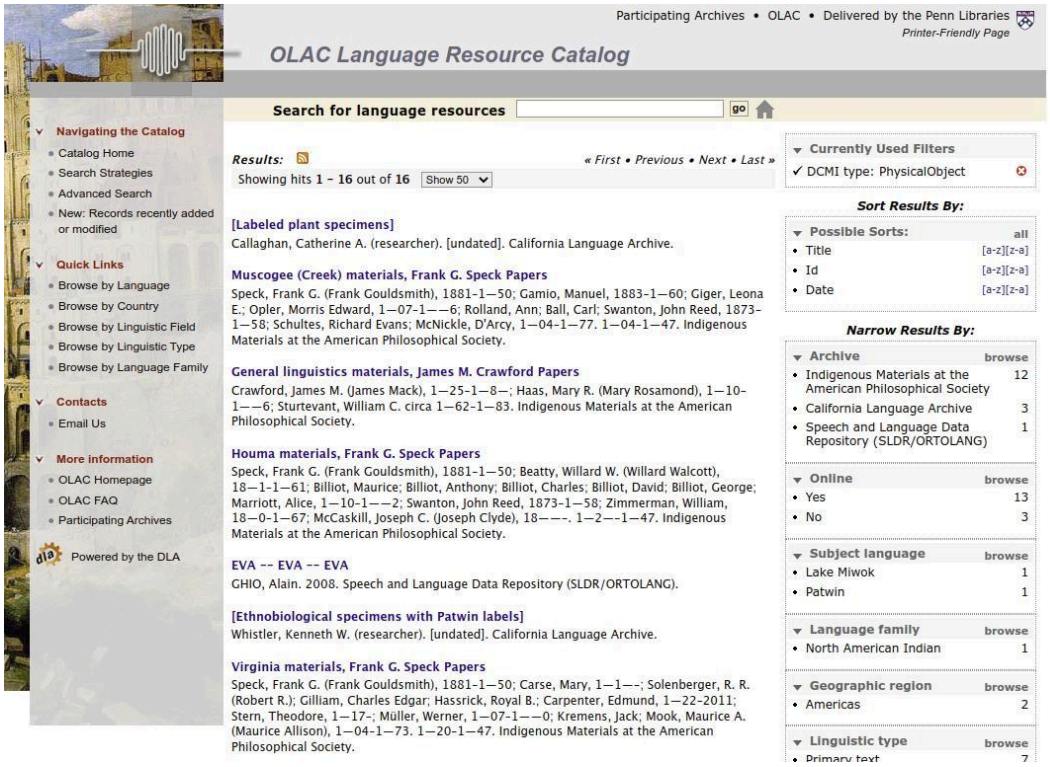

Figure 1. OLAC record explorer with faceted browsing.

Table 1. Overview of Analyzed Records

Data Provider	DCMITypes indicated on metadata record	Artifact Type based on Description	Most appropriate DCMIType
SLDR	PhysicalObject	XML Dataset	Dataset
CLA	PhysicalObject	Biological Specimens	Collection
CLA	PhysicalObject	Biological Specimens	Collection
CLA	PhysicalObject	Biological Specimens	Collection
APS	Sheet music, PhysicalObject, Text	Botanical Specimens	Collection
APS	PhysicalObject, Text, Reports, Charts, Specimens	Ornaments from biological resources	Collection
APS	PhysicalObject, Text, Correspondence, Notes, Specimens, Reports	Bone and Wood Tools; Model	Collection
APS	Notes, Specimens, PhysicalObject, Text	Botanical Specimens	Collection
APS	Correspondence, Notes, PhyscialObject, Text, StillImage, photographs, Essays, Newspaper Clippings, Specimens	Textual Materials	Collection
APS	Drafts, Text, PhysicalObject, Potsherds	Potsherds	Collection
APS	Potsherds, PhysicalObject, Text	No indication	Collection
APS	Specimens, Correspondence, Newspaper Clippings, PhysicalObject, Text	Botanical Specimens	Collection
APS	Correspondence, Disks, Notes, Text, PhysicalObject, Essays, Drafts,	Sound Materials, Textual Materials	Collection
APS	Text, PhysicalObject, StillImage, Drafts, Notes, Disks, Essays, Photographs, Correspondence	Textual Materials	Collection
APS	Botanical Specimens, Notes, Specimens, Text, PhysicalObject	Botanical Specimens	Collection
APS	Newspaper clippings, Postcards, Brochures, Specimens, Stillimages, Text, PhysicalObject	Arrow Head	Collection

Discussion

In this section I discuss four relevant issues. I first discuss the data from the study; second, some issues in the definition of the DCMIType term; third, I provide some illustrated examples of how metadata records could use the term in a definition compliant way and facilitate some communicative goals for language scholars; forth, in Section 4, I provide a metadata template supporting the analysis presented in Section 3.

1. Data Discussion

Analysis of the records revealed three issues. The first appears to be a clerical error in the SLDR provided record. An XML file is not *realia*, i.e., a three-dimensional object or substance. The other two kinds of issues have relevance when considering the best practices of using Qualified Dublin Core.

The second issue is that in several cases the items with physical dimensionality are represented by a record which describes a collection of artifacts, often including texts. The logic for the grouping of artifacts together is well-formed. Artifacts and intellectual pursuits researched together are still associated via the provenance of the artifacts. Per the descriptions provided, their association in physical spaces, folders, or boxes reflect a consistency with the principle of original order. In fifteen of the sixteen

metadata records, to follow best practice and be compliant with the Dublin Core One-to-One Principle,[10] the records should use the DCMIType value *Collection* rather than *PhysicalObject*.

The recommended advice (Hillmann 2005a) for Qualified Dublin Core can be tricky to navigate. For example, in general, unqualified Dublin Core can have an unlimited number of each element applied to an artifact record. The applicability of an unlimited number of repeatable elements to an artifact's descriptive metadata record has been known as the repeatability principle (Hillmann 2005b). Baker (2000) articulates it as a "founding principle that 'every element is optional and repeatable'". However, when considering qualifiers and their terms—specifically those terms within a single taxonomy which are in a mutually exclusive relationship with each other, such as the DCMIType vocabulary—the element can not be repeated with the same qualifier vocabulary. For example, an item is not both primarily engaged with as a *Sound* and as a *Still Image*. The DCMIType vocabulary is supposed to apply to the primary interactive type. If two types are equally valid, then the type *Collection* likely applies because there are really two separate things being described. Therefore, in a sense, this limits the repeatability principle, but it does not nullify it—especially in unqualified applications of Dublin Core. The DCMIType vocabulary terms are defined in such a way that they only describe the primary engagement modality of the described artifact. In a Qualified Dublin Core based application profile, a second, but unqualified use of the DC Type element could be used to refine the primary engagement modality.

These nuances in Qualified Dublin Core are not well described in the Dublin Core literature, and few scholars are exploring these kinds of relationships because the current DCMI Usage Board has assumed that the Dublin Core community has moved to a linked data/RDF model for description. This assumption ignores the prolific use of OAI-PMH (Gaudinat et al. 2017) and its required Dublin Core implementation in an XML format (Lagoze et al. 2015). To complicate the OLAC data context, the OLAC application profile's guidance documents also confuse the repeatability principle and extend the repeatability principle to the mutually exclusive terms of the DCMIType vocabulary (Simons, Bird, and Spanne 2008a, 2008b). The OLAC application profile adds additional qualifiers/refinements to the Dublin Core Type element. While this is not problematic, the OLAC documentation's examples and usage guides show and encourage the use of multiple DCMIType vocabulary terms. Personal communication with some of the founding members of the OLAC application profile has revealed that while the ideal record should be Qualified Dublin Core (this is also stated in Simons, Bird, and Spanne 2008a), several participating data providers have curatorial practices which bundle resources together and only index these complex multi-work resources via a single record (Johnson and Dwyer 2002; Burke and Zavalina 2019; Burke et al. 2020). These OLAC data providers desired to characterize bundles in ways that went beyond thinking about them and describing them with the DCMIType term *Collection*. That is, stakeholders felt the need to state that it was a "collection of something". This same felt need by stewards in other kinds of memory institutions resulted in the development of the *cld:itemType* property defined as part of the *Dublin Core Collection Description Application Profile* (DCCAP).[11] It also aligns with DCMIType term use within the context of *Collection Description in MODS*.[12] However, while both DCCAP and MODS invoke the use of *DCMIType: Collection* in addition to other terms, the DCMIType term *Collection* remains underutilized across the OLAC aggregator (Paterson III 2022). The result is that OLAC metadata, even though it is in a

[10] https://www.dublincore.org/resources/glossary/one-to-one_principle
[11] https://www.dublincore.org/specifications/dublin-core/collection-description/collection-application-profile
[12] https://www.loc.gov/standards/mods/v3/mods-collection-description.html

Qualified Dublin Core format, is often unclear with regard to type declaration; therefore, it is not reusable outside of the OLAC aggregator because the records do not align with general assumptions for Qualified Dublin Core records.

With regard to the discussion of the contrasts between records provided by the California Language Archive (CLA), only one DCMIType value was present. However, when consulting the description field it became evident that the record described several artifacts. This means that the record contains the expected number of elements and qualifiers, though the application of the selected DCMITerm is errant. In contrast with the CLA records, those provided by the American Philosophical Society, do not use the DCMITerm *Collection* but followed the OLAC application profile by applying several DCMITerms. In a sense, then, they are conformant to the OLAC application profile but the records violate the One-to-One Principle and as such have a lower reusability beyond the OLAC aggregator.

The third issue discovered in the metadata records relates to the utility of the DCMIType definition for *PhysicalObject* and the validity of the term in descriptive contexts. The current definition is: "An inanimate, three-dimensional object or substance." Using this definition, it is clear that artifacts such as potsherds and arrow heads are physical objects. However, are *birds*, *plants*, and *insects* inanimate? How about ornaments made from plants? When we consider that earlier definitions of *PhysicalObject* excluded only humans, but then the definition was "clarified" to exclude animate things, the definition and conceptual context around *inanimate* in the Dublin Core context deserves some clarification from the DCMI Usage Board. I further discuss limitations of the definition of *PhysicalObject* in the next section.

2. *Definition of PhysicalObject*

In contrast to historial iterations of the definition of *PhysicalObject* which centered around the *human* versus *non-human* distinction, the current definition centers around the meaning of *animacy*. Looking towards a practical definition of *animacy*, it is essential to note that it has at least three dimensions. First is the life versus death dimension. For example, birds are animate when they are living, but do they become inanimate when dead? That is, can records describing artifacts which were once alive but are no-longer alive be validly listed as *PhysicalObjects* under the current definition?

Second is the has-never-been-alive versus the has-once-been-alive distinction. For example, in the reviewed records one of the indicated physical objects was an ornament crafted out of grass. The grass was at one time living. However, it is now dead and has had a state change into something else, i.e., it was formerly grass and is now a crafted ornament. Perhaps what is crucial when evaluating the applicability of the term *PhysicalObject* is that there is a state change. If this is true, then perhaps the life versus death dimension is only a specialized case of state change. This could mean that there are several different but valid types of state changes which move artifacts from *animate* to *inanimate*.

The third dimension is that of worldview. Some types of artifacts may be animate in one *worldview* or *natural-human-language* while inanimate in another. Many of the world's languages indicate these distinctions overtly through linguistic patterns. Table 2 shows a simple three part ontology for grouping

various kinds of artifacts according to different possible understandings of *animacy*.[13] The implication is that, based on the worldview of the cataloger, an appeal to the *animate* versus *inanimate* definition can lead to a cross-cultural ambiguity.

Table 2. Overview of Analyzed Records

Pattern	Animate	Inanimate
1	Human	Non-human (Animal/Plant/Other)
2	Human/Animal	Plant/Other
3	Human/Animal/Plant	Other

For example, the Papuan language Amele determines animacy following pattern one (Roberts 1998), while Doku (a.k.a., Lengo), a Solomonic language, follows pattern two (Unger 2008, 38–39).[14] In some languages the situation is complex because animate versus inanimate is not a primary distinction. For example, the human versus nonhuman pattern exists (alongside other patterns) in Tarok, an East Benue-Congo language primarily used in Plateau State, Nigeria. This distinction is indicated when nouns use a ‘ u-’ prefix for singular forms and ‘o-’ prefix for plural forms. Personification, or the rise to animate/human status of other things is achieved by employing the prefixes on any noun. Personification is not a mere figure of speech but a worldview that life is unified and deeply inter-dependent; and that everything is potentially alive (p.c., Selbut Longtau, 22 March 2023). Worldview opinions also impact classification in the biological sciences. For example, bacteria are considered living, while viruses are considered non-living (Koonin and Starokadomskyy 2016).

The definition of *PhysicialObject* has an impact in a range of preservation contexts even if it hasn’t had ample discussion in scholarly literature. The definition of *PhysicialObject* has implications for how bones are classified in anthropological collections (Thomas 2015; Ousley, Billeck, and Hollinger 2005; Watson 2003). Bones are larger physical objects when compared to DNA, but the conceptualization of DNA as data is an issue with which many scholars have struggled (Alpaslan-Roodenberg et al. 2021). The inability to use the DCMIType vocabulary for animate artifacts cajoles scholars to classify them as *data* and possibly use the DCMIType term *Dataset*. Beyond *data* some anatomical (often attached to medical schools) and zoological collections contain preserved/lifeless but previously animate specimens. If one considers plants as animate materials, then seed libraries and seed banks also have collections of physical artifacts. These artifacts fall into the ambiguous area via the current DCMIType definition. To address some aspects of the current situation, Darwin Core (Endresen and Knüpffer 2012; Baskauf et al. 2016; Baskauf and Webb 2016), a metadata schema based on Dublin Core, has extended the DCMIType vocabulary by adding the term *Organism*.[15] Even though *Organism* may solve some problems, using this term requires application profile developers to add additional namespace to their application profile. Reuse contexts may not be prepared to add additional metadata elements from these namespaces. Clear

[13] Based on conversations between the DCMI Usage Board and the developers of Darwin Core (Cox 2021), my understanding of the current definition of *PhysicalObject* aligns the term with type three in the ontology within Table 2.

[14] Sinhala, a language primarily used in Sri Lanka, is reported to also follow pattern two.

[15] http://rs.tdwg.org/dwc/terms/Organism

definitional semantics at DCMI would allow for the greatest amount of clarity in metadata aggregation and reuse.

Returning to the language archives use case of this study in the next section, I discuss how language resource stewards can use *PhysicalObject* for metadata records related to audio and video recording tools to document equipment and workflows in the production process.

3. *Possible uses in language resource stewardship*

PhysicalObject could be used to create metadata records for the recording tools used in a language documentation project (e.g., microphones, digital audio recorders, video cameras, etc.). This has apparently, to date, been overlooked as recording tools have been conceptualized as metadata to the sound or video artifact. However, with the DCMIType *PhysicalObject,* a metadata record can be created for the physical tool and then relationships can be made to appropriate records for the sound or video artifacts.

At the beginning of the 21st century, domain-based digitally-focused language resource stewardship was in its infancy. In that context, the US National Science Foundation sponsored a series of workshops to develop best practices within language scholarship and preservation of the evidentiary record. These were known as the E-MELD workshops. Their impact was international and highly discussed across the scholarly practice of language documentation and field linguistics.

As evidenced in proceedings of the 2006 E-MELD workshop, Heidi Johnson led the E-MELD Working Group for Corpus Management and Metadata.[16] Among other things, they ideated on the 'the perfect corpus management tool'. This tool would be used to transmit metadata and language resources from the desks of scholars to digital libraries at stewardship institutions. One necessary feature was the ability to enter information (i.e., metadata) for analytical products in analysis and production pipelines. They specifically mentioned: 1) referencing the physical device that recorded the digital artifacts, 2) referencing contributors, and 3) the ability to create records for software, platform, and fonts as needed. At the time of the E-MELD workshops, the OLAC application profile was also being developed. In the conversations at that time, it was perceived that the relationship between a recording device and a recording was an attribute of the recording. Digital libraries were also very content focused. That is, metadata records were supposed to have digital artifacts associated with them. The OLAC application profile was understood (by some) to not be sufficiently descriptive for indicating recording or digitization equipment (Broeder et al. 2012; Van Uytvanck, Stehouwer, and Lampen 2012). This, among other reasons, motivated Johnson and others to use ISLE Meta Data Initiative (IMDI)[17] based metadata schemas in the language resource stewardship organizations they managed (Johnson and Dwyer 2002; Broeder and Wittenburg 2006; Broeder et al. 2012; Trippel and Zinn 2018). As far as I know, no recommendation was made that recording and digitizing devices be given their own metadata records and subsequently described and related to relevant information object records which are more frequently understood to be the preservable artifacts in scholarly language research activities.

[16] https://web.archive.org/web/20080208060149/http://emeld.org/workshop/2006/proceedings.html
[17] https://archive.mpi.nl/forums/t/imdi-metadata-information/2933

Paterson (In Press) uses bibliographic models, such as OpenWEMI (Coyle 2022) and Dublin Core metadata, to describe and illustrate the relationships around transcriptions, a commonly understood artifact type in language scholarship. These are illustrated in Figure 2. When recording devices are added, this further implements Dublin Core in a definition compliant way. I illustrate this in Figure 3 where audio recording devices are given both a type and token record. These records are then related to the digital audio artifact through a Dublin Core Source relationship. One reason this approach may not have seen uptake is that there are not any templates or examples of it in obvious locations relevant to metadata application profile documentation. In the next section, I provide such a template.

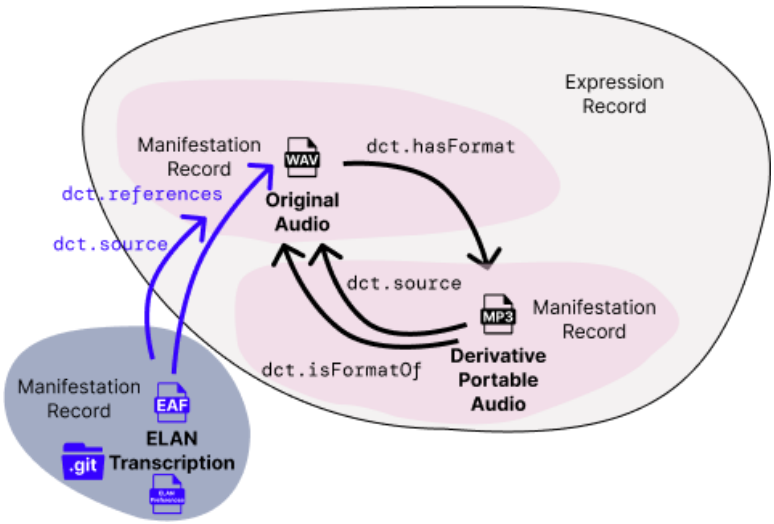

Figure 2. Illustrating relationships between common digital scholarly artifacts in language scholarship

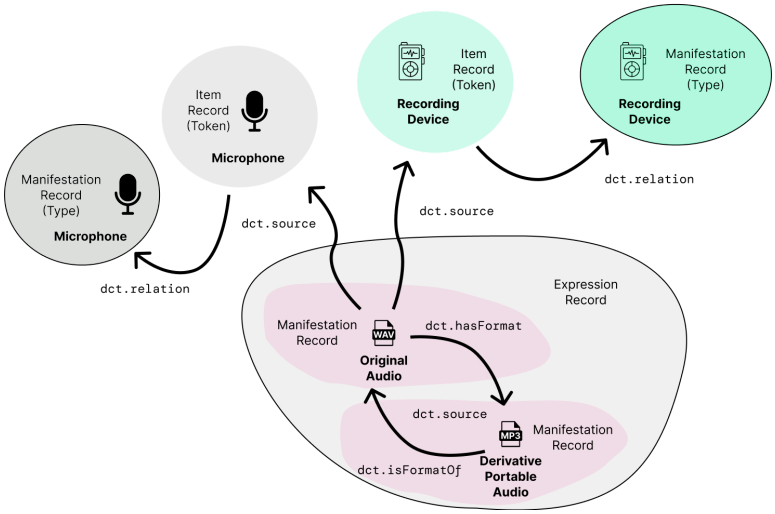

Figure 3. Illustrating relationships between records for audio and records for recording devices

4. Template

A template for working within the OLAC application profile is sketched out for a microphone example in Table 3. In the first column, the Qualified Dublin Core element is presented, followed in the second column by an explanation of the kind of applicable information. Finally in the third column, I present an example of using a microphone I own. Table 3 is then followed by an XML formatted example (Figure 4).

Table 3. Template for physical recording devices

Dublin Core Element	Usage Explanation	Example Value
title	What is the advertised label? Include the Advertised label.	Røde M3
type	DCMIType term	PhysicalObject
extent	The extent should provide the physical dimensions.	390 grams (empty); 22.5cm long and 3.3cm wide.
abstract	A summary of the object. For microphones should include: Voltage, Connection, Pickup Pattern, Specifications.	Black microphone with an XLR cable plug at the bottom. The mic has an internally shock-mounted 1/2-inch condenser capsule. It supports phantom power via 9v battery or P48 phantom power. It has a three way mechanical switch to activate a padding of (-10dB, -20dB) or a high-pass filter (80Hz). Good for recording musical instruments or vocal performances with a cardioid polar pattern via a mic stand.
accrualMethod	What was the method by which this resource was acquired? Consider if any of the accrual methods are in the Accrual Method Vocabulary.	Purchase
provenance	"A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation." Statements of use need not be included if they are inferable from records. For example, I used this mic in Mexico and Nigeria but if I have recordings connected to this mic then those use instances are inferable.	Purchased new from B & H Photo in 2010. Hugh Paterson III used this mic in Mexico and Nigeria.
identifier	Local inventory number.	Paterson #069
identifier	Item level serial number; usually the model number is in the title.	S/N 0027437
date	Date of acquisition or a general date if other more specific dates are not used.	2010
available	Date the product came on the market.	2001-10-02
created	Date of manufacture or creation.	
replaces	A link to the item this item replaced	
requires	What dependencies does this object have for its utility to be actuated? Each requirement should be in its own element.	Phantom Power (9 volt battery or 48 volt), Mic stand.
requires	What dependencies does this object have for its utility to be actuated?	XLR cable
issued	Date of formal issuance of the resource.	

conformsTo	An established standard to which the described resource conforms.	
rightsHolder	The entity for which the rights statement applies.	Hugh Paterson III
rights	On a microphone, "ownership" is the most appropriate rights statement in this context.	Owner
relation	This is a Token Record (WEMI: Item) and should be linked to a WEMI: Manifestation record a.k.a Type Record	Link to General record for all Røde M3 microphones.

```
<olac:olac>
  <dcterms:title xml:lang="en">Røde M3</dcterms:title>
  <dcterms:identifier>Item</dcterms:identifier>
  <dcterms:identifier>Paterson #069</dcterms:identifier>
  <dcterms:identifier>S/N 0027437</dcterms:identifier>
  <dcterms:identifier
xsi:type="dcterms:URI">https://rode.com/en-us/microphones/live/m3</dc
terms:identifier>
  <dcterms:type
xsi:type="dcmitype:DCMIType">physicalObject</dcterms:type>
  <dcterms:provenance>Purchased new from B & H Photo in 2010.  Hugh
Paterson III used this mic in Mexico and
Nigeria.</dcterms:provenance>
  <dcterms:date xsi:type="dcterms:W3CDTF">2010</dcterms:date>
  <dcterms:rightsHolder>Hugh Paterson III</dcterms:rightsHolder>
  <dcterms:rights>Ownership</dcterms:rights>
  <dcterms:requires>XLR Cable</dcterms:requires>
  <dcterms:requires>Phantom Power (9 volt battery or 48
volt)</dcterms:requires>
<dc:description>Cardioid polar pattern; 22.5cm long and 3.3cm wide.
Black in Color.</dc:description>
  <dcterms:abstract>XLR cable plugs into the bottom. The mic has an
internally shock-mounted 1/2-inch condenser capsule. It supports
phantom power via 9v battery or P48 phantom power. It has a three way
mechanical switch to activate a padding of (-10dB, -20dB) or a
high-pass filter (80Hz). Good for recording musical instruments or
vocal performances with a cardioid polar pattern via a mic
stand.</dcterms:abstract>
  <dcterms:accrualMethod>Purchase</dcterms:accrualMethod>
  <dcterms:available
xsi:type="dcterms:W3CDTF">2001-10-02</dcterms:available>
</olac:olac>
```

Figure 4. XML of an OLAC Record containing content from Table 3.

Conclusion

In this paper, I have shown how the DCMIType term *PhysicalObject* is currently used by OLAC data providers. I have investigated the validity of those uses and pointed out some of the semantic issues with the definitions provided by the DCMI Usage Board. The few cases where realia are present in the collections and motivate the use of *PhysicalObject* suggest that OLAC data providers may have allies (with related collections) in cultural heritage memory institutions with physical holdings. Further, people studying language-in-use might benefit from these connections between language artifacts and realia. I have also proposed a valid use within language resource stewardship contexts to further make transparent the workflows used in language documentation. The use is novel in that it requires memory institutions to create records for resources which are not directly in their stewardship. However, the creation of metadata records representing artifacts not currently in holdings is something some institutions already do.

Contrary to claims that OLAC and Dublin Core metadata is simple or insufficient for detailed language resource description (Broeder et al. 2012; Van Uytvanck, Stehauwer, and Lampen 2012; Austin 2013; McCrae et al. 2015), I have shown that the true power underlying the metadata model upon which OLAC is built has not been fully explored. That is, I have shown that Dublin Core via the OLAC application profile can accommodate more of the goals articulated at E-MELD workshops than was previously shown. I argue that the semantics of the DCMIType term *PhysicalObject* is poorly defined and that the DCMI Usage Board should address the insufficiency. In general, I believe that the use of *PhysicalObject* may be useful in metadata transformations where Dublin Core is the target format and the source format indicates a digitization or transfer process including the machinery used in the transfer.[18]

References

Aljalahmah, Saleh, and Oksana L. Zavalina. 2023. “Exploration of Metadata Practices in Digital Collections of Archives with Arabian Language Materials.” In *Proceedings of the International Workshop on Digital Language Archives: LangArc-2023*, edited by Oksana L. Zavalina and Shobhana Lakshmi Chelliah, 11–14. Denton, Texas: University of North Texas. doi:10.12794/langarc2114295.

Alpaslan-Roodenberg, Songül, David Anthony, Hiba Babiker, Eszter Bánffy, Thomas Booth, Patricia Capone, Arati Deshpande-Mukherjee, et al. 2021. “Ethics of DNA Research on Human Remains: Five Globally Applicable Guidelines.” *Nature* 599 (7883). Nature Publishing Group: 41–46. doi:10.1038/s41586-021-04008-x.

Austin, Peter K. 2013. “Language Documentation and Meta-Documentation.” In *Keeping Languages Alive: Documentation, Pedagogy and Revitalization*, edited by Mari C. Jones and Sarah Ogilvie, 3–15. Cambridge, UK; New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139245890.003.

Baker, Thomas. 2000. “A Grammar of Dublin Core.” *D-Lib Magazine* 6 (10). doi:10.1045/october2000-baker.

Baskauf, Steven J., and Campbell O. Webb. 2016. “Darwin-SW: Darwin Core-Based Terms for Expressing Biodiversity Data as RDF.” *Semantic Web* 7 (6). IOS Press: 629–43. doi:10.3233/SW-150203.

[18] One such metadata standard is PBCore which nests digitization metadata within records about sound artifacts. https://pbcore.org/elements/pbcoreinstantiationdocument

Baskauf, Steven J., John Wieczorek, John Deck, and Campbell O. Webb. 2016. “Lessons Learned from Adapting the Darwin Core Vocabulary Standard for Use in RDF.” Edited by Pascal Hitzler and Krzysztof Janowicz. *Semantic Web* 7 (6): 617–27. doi:10.3233/SW-150199.

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.

———. 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.

———. 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.

———. 2022. “The Open Language Archives Community: A 20-Year Update.” *The Electronic Library* 40 (5): 507–24. doi:10.1108/EL-08-2022-0192.

Broeder, Daan, 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)*, edited by Victoria Arranz, Daan Broeder, Bertrand Gaiffe, Maria Gavrilidou, Monica Monachini, and Thorsten Trippel, 1–4. Istanbul, Turkey: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/workshops/11.LREC2012%20Metadata%20Proceedings.pdf.

Broeder, Daan, and Peter Wittenburg. 2006. “The IMDI Metadata Framework, Its Current Application and Future Direction.” *International Journal of Metadata, Semantics and Ontologies* 1 (2). Inderscience Publishers: 119–32. doi:10.1504/IJMSO.2006.011008.

Burke, Mary, and Oksana Zavalina. 2019. “Exploration of Information Organization in Language Archives.” *Proceedings of the Association for Information Science and Technology* 56 (1): 364–67. doi:10.1002/pra2.30.

Burke, Mary, and Oksana L. Zavalina. 2020. “Identifying Challenges for Information Organization in Language Archives: Preliminary Findings.” In *Sustainable Digital Communities: Proceedings of the 15th International Conference, iConference 2020, Böras, Sweden, March 23–26, 2020*, edited by Anneli Sundqvist, Gerd Berget, Jan Nolin, and Kjell Ivar Skjerdingstad, 622–29. Lecture Notes in Computer Science 12051. Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-030-43687-2_52.

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 (April). University of Hawaii Press: 1–24. http://hdl.handle.net/10125/74669.

Burke, Mary, Oksana L. Zavalina, Mark Edward Phillips, and Shobhana Chelliah. 2020. “Organization of Knowledge and Information in Digital Archives of Language Materials.” *Journal of Library Metadata* 20 (4). Routledge: 185–217. doi:10.1080/19386389.2020.1908651.

Cox, Simon. 2021. “Why Is DCType:PhysicalObject ‘Inanimate’?” Issue 101. *Gihub Space: Dcmi/Usage*. https://github.com/dcmi/usage/issues/101.

Coyle, Karen. 2022. “Works, Expressions, Manifestations, Items: An Ontology.” *The Code4Lib Journal* 53 (May). https://journal.code4lib.org/articles/16491.

Endresen, Dag Terje Filip, and Helmut Knüpffer. 2012. “The Darwin Core Extension for Genebanks Opens up New Opportunities for Sharing Genebank Datasets.” *Biodiversity Informatics* 8 (1): 12-‑ 29. doi:10.17161/bi.v8i1.4095.

Gaudinat, Arnaud, Jonas Beausire, Megan Fuss, Elisa Banfi, and Julien Gobeill. 2017. “Global Picture of OAI-PMH Repositories through the Analysis of 6 Key Open Archive Meta-Catalogs.” arxiv. 1708.08669. arxiv. https://arxiv.org/pdf/1708.08669.pdf.

Guenther, Rebecca, ed. 1999. *Dublin Core Resource Types List*. Washington, D.C: Library of Congress. https://www.loc.gov/marc/dc/typelist-19991118.html.

Hillmann, Diane I. 2005a. “Dublin Core Qualifiers.” In *Using Dublin Core*, Section 5. Dublin, Ohio: DCMI. https://web.archive.org/web/20060109122452/http://dublincore.org/documents/usageguide/qualifiers.shtml.

Hillmann, Diane I. 2005b. “Using Dublin Core.” Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/usageguide/#whatis.

Huber, Christian. 2023. “Why It Can Be Difficult to Make Historic Language Recordings Accessible: A View from a Corpus of Historic Dialect Recordings.” In *Proceedings of the International Workshop on Digital Language Archives: LangArc-2023*, edited by Oksana L. Zavalina and Shobhana Lakshmi Chelliah, 15–18. Denton, Texas: University of North Texas. doi:10.12794/langarc2114302.

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 3334. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-30544-6_34.

Johnson, Heidi, and Arienne Dwyer. 2002. “Customizing the IMDI Metadata Schema for Endangered Languages.” In *Proceedings of the International LREC [Language Resources and Evaluation Conference] Workshop on Resources and Tools in Field Linguistics*, edited by Peter Wittenburg. Las Palmas, Canary Islands, Spain: LERC / European Language Resource Association (ELRA). https://www.mpi.nl/lrec/2002/papers/lrec-pap-05-JohnsonDwyer.pdf.

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

Koonin, Eugene V., and Petro Starokadomskyy. 2016. “Are Viruses Alive? The Replicator Paradigm Sheds Decisive Light on an Old but Misguided Question.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 59 (October): 125–34. doi:10.1016/j.shpsc.2016.02.016.

Lagoze, Carl, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. 2015. *The Open Archives Initiative Protocol for Metadata Harvesting*. 2.0. Open Archives Initiative. http://www.openarchives.org/OAI/openarchivesprotocol.html.

McCrae, John P., Penny Labropoulou, Jorge Gracia, Marta Villegas, Víctor Rodríguez-Doncel, and Philipp Cimiano. 2015. “One Ontology to Bind Them All: The META-SHARE OWL Ontology for the Interoperability of Linguistic Datasets on the Web.” In *The Semantic Web: ESWC 2015 Satellite Events*, edited by Fabien Gandon, Christophe Guéret, Serena Villata, John Breslin, Catherine Faron-Zucker, and Antoine Zimmermann, 271–82. Lecture Notes in Computer Science (9341). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-25639-9_42.

Ousley, Stephen D., William T. Billeck, and R. Eric Hollinger. 2005. “Federal Repatriation Legislation and the Role of Physical Anthropology in Repatriation.” *American Journal of Physical Anthropology* 128 (S41): 2–32. doi:10.1002/ajpa.20354.

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.

———. 2009. “Metadata Quality in Digital Repositories: A Survey of the Current State of the Art.” *Cataloging & Classification Quarterly* 47 (3–4): 213–28. doi:10.1080/01639370902737240.

Park, Jung-ran, and Eric Childress. 2009. “Dublin Core Metadata Semantics: An Analysis of the Perspectives of Information Professionals.” *Journal of Information Science* 35 (6): 727–39. doi:10.1177/0165551509337871.

Paterson III, Hugh J. 2021a. “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.

———. 2021b. “OLAC Nightly Data Dump (XML) from 18 July 2021.” XML. OLAC: Open Language Archives Community. Zenodo. doi:10.5281/ZENODO.5112131.

———. 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*. Chicago, IL: Society of American Archivists. https://www2.archivists.org/am2021/research-forum-2021/agenda#peer.

———. 2023a. “An OLAC Perspective on Services: The Forgotten Language Resources.” In *Proceedings of the International Conference on Dublin Core and Metadata Applications (2022)*. Dublin Core Metadata Initiative. doi:10.23106/DCMI.953120677.

———. 2023b. “OLAC and Serials: An Appraisal.” In *Proceedings of the 2nd International Workshop on Digital Language Archives: LangArc 2023*, edited by Oksana L. Zavalina and Shobhana Lakshmi Chelliah, 28–32. Denton, Texas: University of North Texas. doi:10.12794/langarc2114296.

———. 2023c. “Towards Making Shared Metadata Interoperable across the Open Language Archives Community.” In *Proceedings of the 2nd International Workshop on Digital Language Archives: LangArc 2023*, edited by Oksana L. Zavalina and Shobhana Lakshmi Chelliah, 33–37. Denton, Texas: University of North Texas. doi:10.12794/langarc2114297.

———. 2023d. “Diversity and Identity: Categories for OAI Data-Providers in the Open Language Archives Network.” *Proceedings from North American Symposium on Knowledge Organization (NASKO)* 9 (1): 18–31. doi:10.7152/nasko.v9i1.16301.

———. 2023e. “Supporting Evidence for OLAC DCMIType PhysicalObject.” XML, PDFs, and JPGs. OAI-PMH Records, Summary Tables, and Screenshots. Zenodo. doi:10.5281/zenodo.10153055.

———. In Press. “Considering WEMI and the Digital Artifacts Around Language Documentation Transcripts.” In *Proceedings of the 21st Dublin Core Conference: Metadata Innovation: Enabling Knowledge in a Data-Intensive World (DC-2023)*. Daegu, South Korea. [Pre-print]: Dublin Core Metadata Initiative. https://hughandbecky.us/Hugh-CV/talk/2023-wemi-language-documentation-transcripts/.

Roberts, John R. 1998. “GIVE in Amele.” In *The Linguistics of Giving*, edited by John Newman, 1–34. Typological Studies in Language 36. Amsterdam: John Benjamins Publishing Company. doi:10.1075/tsl.36.02rob.

Simons, Gary F., and Steven Bird. 2003a. “Building an Open Language Archives Community on the OAI

Foundation." *Library Hi Tech* 21 (2): 210–18. doi:10.1108/07378830310479848.

———. 2003b. "The Open Language Archives Community: An Infrastructure for Distributed Archiving of Language Resources." *Literary and Linguistic Computing* 18 (2): 117–28. doi:10.1093/llc/18.2.117.

———. 2011. "OLAC: Accessing the World's Language Resources." Poster presented at the Linguistic Society of America Annual Meeting, Pittsburgh, 6–9 January 2011. https://scholars.sil.org/sites/scholars/files/gary_f_simons/poster/olac_poster_for_lsa_2011.pdf.

Simons, Gary F., Steven Bird, and Joan Spanne, eds. 2008a. "OLAC Metadata Usage Guidelines." OLAC Notes. Open Language Archive Community. Http://web.archive.org/web/20230424083006/http://www.language-archives.org/NOTE/usage-20080711.html. Internet Archive. http://www.language-archives.org/NOTE/usage-20080711.html.

———. 2008b. "Type." *OLAC Metadata Usage Guidelines*. OLAC Notes. Open Language Archive Community. http://www.language-archives.org/NOTE/usage.html#Type.

Thomas, Martin. 2015. "Bones as a Bridge between Worlds: Responding with Ceremony to the Repatriation of Aboriginal Human Remains from the United States to Australia." In *Conciliation on Colonial Frontiers: Conflict, Performance, and Commemoration in Australia and the Pacific Rim*, edited by Kate Darian-Smith and Penelope Edmonds. Routledge Studies in Cultural History. New York: Routledge.

Trippel, Thorsten, and Claus Zinn. 2018. "Lessons Learned: On the Challenges of Migrating a Research Data Repository from a Research Institution to a University Library." In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 146–50. Miyazaki, Japan: European Language Resources Association (ELRA). https://aclanthology.org/L18-1023.

Unger, Paul. 2008. "Aspects of Lengo Grammar." M.A. Thesis, Trinity Western University: Trinity Western University. https://drive.google.com/file/d/1h-_S0Dmfr7QBNXKW5KpgkWtWGL9tETwX/view.

Van Uytvanck, Dieter, Herman Stehouwer, and Lari Lampen. 2012. "Semantic Metadata Mapping in Practice: The Virtual Language Observatory." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 1029–34. Istanbul, Turkey: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/437_Paper.pdf.

Ward, Jewel Hope. 2002. "A Quantitative Analysis of Dublin Core Metadata Element Set (DCMES) Usage in Data Providers Registered with the Open Archives Initiative (OAI)." M.S. in Information Science, Chapel Hill, NC, USA: University of North Carolina at Chapel Hill. https://ils.unc.edu/MSpapers/2816.pdf.

———. 2003. "A Quantitative Analysis of Unqualified Dublin Core Metadata Element Set Usage within Data Providers Registered with the Open Archives Initiative." In *Proceedings of the 2003 Joint Conference on Digital Libraries, 2003*, 315–17. Houston, TX, USA: IEEE Comput. Soc. doi:10.1109/JCDL.2003.1204883.

———. 2004. "Unqualified Dublin Core Usage in OAI‐PMH Data Providers." *OCLC Systems & Services: International Digital Library Perspectives* 20 (1). Emerald Group Publishing Limited: 40–47. doi:10.1108/10650750410527322.

Watson, Nicole. 2003. "The Repatriation of Indigenous Remains in the United States of America and Australia: A Comparative Analysis." *Australian Indigenous Law Reporter* 8 (1). Indigenous Law

Centre, Law School, University of New South Wales: 33–44. https://www.jstor.org/stable/26479533.

Zavalin, Vyacheslav, and Oksana L. Zavalina. 2023. “Exploration of Accuracy, Completeness and Consistency in Metadata for Physical Objects in Museum Collections.” In *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*, edited by Isaac Sserwanga, Anne Goulding, Heather Moulaison-Sandy, Jia Tina Du, António Lucas Soares, Viviane Hessami, and Rebecca D. Frank, 83–90. Lecture Notes in Computer Science 13972. Cham: Springer Nature Switzerland. doi:10.1007/978-3-031-28032-0_7.

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). Routledge: 104–28. doi:10.1080/19386389.2011.629957.