

User needs in language archives: Findings from interviews with language archive managers, depositors, and end-users

Mary Burke

University of North Texas

Oksana L. Zavalina

University of North Texas

Shobhana L. Chelliah

University of North Texas

Mark E. Phillips

University of North Texas

Language archives, like other scholarly digital repositories, are built with two major audiences in mind. These are depositors of language data and various potential end-users of these materials: researchers (linguistics and others), language communities, students, educators, artists, etc. Being a relatively new phenomenon, language archives have made significant strides forward in providing access to digital language data. With the purpose of identifying the needs of language archive end-users (both met and currently unmet), our interdisciplinary team of linguists and information scientists interviewed language archive managers, end-users, and depositors. This study offers a first look into the decision-making processes and end-user experiences of these groups. To support the continued development of language archives, the exploratory study reported in this article provides empirical data on language archive user needs and supports some anecdotal evidence of known issues facing language archive end-users, depositors, and managers in primarily academic contexts.

1. Introduction and background¹ As language documentation begins to emphasize collaboration with language community members (Czaykowska-Higgins 2009; Nathan & Fang 2013), in the past decade, language archives too have shifted their focus to prioritize language community end-users in addition to researchers and educators. Building from the foundation laid by the Open Language Archives Community (OLAC), the Digital Endangered Languages and Musics Archives Network (DELAMAN) is leading the language archiving community in efforts to “break

¹ This research was funded by the US Institute for Museum and Library Services (IMLS) grant # IMLS LG-87-18-0197. We are thankful to our participants for their time and insights reported here, to the advisory board members for their valuable input, and to our anonymous reviewers who provided thorough feedback on the manuscript.

traditional boundaries between depositors, end-users, and archivists to expand the audiences and uses for archives while involving speaker communities directly in language documentation and archival processes” (Henke & Berez-Kroeker 2016: 428). Conversations are developing around data availability for research reproducibility (Maxwell 2012; Thieberger 2014; 2016; Berez-Kroeker et al. 2018), citation practices (Berez 2015; Gawne et al. 2017), the introduction of linguistics into the broader Open Access Movement (Seyfeddinipur et al. 2019), and the need for appropriate recognition for archiving work (Haspelmath & Michaelis 2014; Thieberger et al. 2016). At the same time, major journals in documentary linguistics have begun to require citations for any primary data cited in their publications and to allow supplementary material (transcriptions, audio, code) to be appended to online publications such as *Himalayan Linguistics* and *Language Documentation & Conservation*.

Despite these great strides, both depositors and potential end-users continue to face challenges when utilizing language archives (Wasson et al. 2016; Sullivant 2020). A survey of existing grammatical descriptions found that, of a hundred grammars and dissertations in a ten-year span, only twenty-two referred the reader to a retrievable archival collection of primary data (Gawne et al. 2017). In light of these observations, linguists have identified the need for a deeper understanding of language archive end-users: their preferences, their needs, and potential challenges they face (Austin 2011; Hildebrandt 2020). Woodbury (2014) advocates for depositors to anticipate the potential audiences of their deposit, annotate the material accordingly, and describe the decision-making process behind what they chose to archive as a guide to future depositors (similar to Austin’s [2013] ‘meta-documentation’). Sullivant (2020) expands on this idea with a comprehensive guide to describing collections of language data to facilitate end-users’ discovery and navigation of archival materials. Seyfeddinipur et al. (2019) recommend that researchers determine and communicate language communities’ unmet needs to language archives in attempts to resolve the barriers to access.

Metadata, broadly defined as ‘data about data,’ is necessary to make these materials accessible and usable. With regard to metadata, the linguistics community has distinguished between information designed to facilitate resource discovery in catalogs and repositories (Good 2002) from annotations and markup of primary data (called ‘thick’ metadata in Nathan & Austin 2004). A metadata record, for example, as shown in Figure 1, contains information about an item: the language, the date it was created, and a basic description of the contents.

| Resource Details | |
|----------------------|--|
| Identifier: | CY978A-Video-06 |
| Title: | Yup'ik Video Collection Box 6 |
| Description: | LKSD (Lower Kuskokwim School District) Video collection; Seal Party, Yup'ik; Seal Party, English; Dog Mushing Yup'ik; Dog Mushing A Way of Life English; Good Fish Catchers Yup'ik; Good Fish Catchers English; Dance Sticks Yup'ik; Dance Sticks English. |
| Comments: | Donated by Oscar and Sophie Alexie, 2011. Date approximate. |
| Contributors: | Alexie, Sophie (author), Alexie, Oscar (author) |
| Date: | 1980 |
| Type: | MovingImage |
| Format: | |
| Coverage: | |
| Language: | Yup'ik, Central Alaskan |
| Linguistic | place names |
| Type: | |
| Collections: | Oscar and Sophie Alexie Collection |

Figure 1. Sample metadata record from the Alaska Native Language Archive²

Any annotation, such as a transcription of an audio recording, would constitute a separate item, rather than metadata for the original audio recording.

The linguistics community has also evaluated the extent to which metadata support the needs for information discovery and access, or the quality of language archive metadata. One early example of a metadata quality evaluation is the OLAC site's ratings of language archives' metadata based on the inclusion of specific elements (title, description, subject, date, and identifier) and use of standard terms for representing attributes of language materials like the discourse type (OLAC Discourse Type³) and roles of participants (OLAC Role Vocabulary⁴) (Hughes 2005). More recently, research on language archive metadata quality explores the use of the Description field in metadata records (Burke & Zavalina 2020a). For example, Harris et al. (2019) describe how a language community representative was able to identify gaps and errors in the metadata of multiple collections of Papua New Guinean languages in PARADISEC. He found "missing interpretations, especially a surprising lack of detail or information" in the Description fields provided by the original depositors, and enriched the records by adding contextual information to facilitate access to these recordings (Harris et al. 2019: 140). Burke & Zavalina

² <https://www.uaf.edu/anla/record.php?identifier=CY978A-Video-06> (Accessed 2020-12-22.)

³ <http://www.language-archives.org/REC/discourse.html> (Accessed 2021-11-01.)

⁴ <http://www.language-archives.org/REC/role.html> (Accessed 2021-11-01.)

(2020a) provide a comparative analysis of the content in the Description fields of metadata records in the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC), the Archive of the Indigenous Languages of Latin America (AILLA), and the Endangered Languages Archive (ELAR), and the metadata creation guidelines given to depositors.

Though ongoing research in this area holds promise for improving access to material in language archives, assessments of end-user needs and metadata quality fall outside the traditional scope of linguistics research. Other social science disciplines focus on understanding end-users' needs (information behavior), presenting information with metadata that suits the context (information organization), and developing best practices for providing access to digital objects (digital curation). To appropriately address this gap, we as linguists can benefit from interdisciplinary collaboration with researchers in such areas to tackle these complex problems. For example, one of the first explorations of this topic took anthropological and user-centered design perspectives into consideration (Wasson et al. 2016). Along with higher-level concerns about the ethics of access, their focus group of a selection of language archive stakeholders identified the following barriers that language archive end-users face:

- lack of contextual information and annotations;
- lack of opportunities for end-users to engage with material;
- lack of culturally relevant categories;
- inability to effectively find data with existing search, browse, and display options; and
- issues with accessibility (in terms of interface language, terminology, and technology).

Many of these issues relate to how materials are presented to end-users, the metadata accompanying them, and end-user interactions with language archives – that is, information organization, information behavior, and data curation, which are core areas of information science. Building on Wasson et al.'s 2016 findings, our interdisciplinary team of linguists, information science researchers, and archivists designed and implemented a research project to accomplish the following goals:

- identify met and unmet needs of language archive end-users and depositors (both actual and potential) and
- understand the challenges language archive managers face and their decision-making processes.

This paper is organized as follows: §2 describes the methodology followed for data collection and analysis; §3 reports the findings; §4 provides context for the findings; and §5 concludes the paper.

2. Study design This qualitative exploratory study took place in two stages: Stage 1, a content analysis, surveyed information regarding metadata, the depositing process,

and search functionalities publicly available on twenty language archive websites⁵ (Burke & Zavalina 2019). The content analysis informed the design of Stage 2: a series of semi-structured interviews and observations conducted in summer and fall of 2019. We interviewed actual and potential end-users of language archives about their experiences depositing and accessing data from language archives. Some participants were observed while browsing language archives and encouraged to comment on their experience using the think-aloud protocol⁶. Stage 2 also included interviews with language archive managers to better understand their decision-making processes, future plans for information organization, and feedback they received from end-users. Findings from the language archive managers' interviews are reported in Burke & Zavalina (2020b) and Burke et al. (2021).

As stipulated by IMLS, the scope of our study focused on language archives within the United States. As such, we chose to recruit participants through public announcements posted to the blog of the Linguistic Society of America's (LSA) Committee on Endangered Languages and their Preservation (CELP) and the LinguistList. In total, sixteen participants, including archive managers (7), depositors (4), and end-users (5), were interviewed. All interviews and observations were conducted in English.

Here, the term *depositors* refers to those who have contributed collections of language data to language archives. *End-users* are those who use language archives for research or education purposes but may or may not have contributed material to language archives in the past. *Language archive managers* are those who are responsible for the maintenance and/or design of language archives. It is common for language archive managers to be linguists, and within any sample of language archive managers, there are inevitably individuals who have also deposited their own materials in language archives and/or have accessed language archives as end-users. Because of the significant overlap between these roles, participants identified the role they felt most comfortable speaking from⁷. Interviews were conducted online via the Zoom videoconferencing tool.

Though we aimed to include participants from each group identified in Burke & Zavalina (2019), we fell short in recruiting participants from language communities due to the bandwidth requirements of Zoom and lack of response to public recruitment posts. The bandwidth required by Zoom limited our ability to accommodate participants with low Internet connectivity. We planned to seek out language community participants at the Collaborative Language Institute (CoLang) in the summer of 2020 but were unable to follow through due to the COVID-19 pandemic.

⁵ Stage 1 included sixteen language archives in the United States, two in Europe, one in Australia, and one in Canada.

⁶ <https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/> (Accessed 2021-11-01.)

⁷ Interview guides for end-users and language archive managers can be found at <https://digital.library.unt.edu/ark:/67531/metadc1757990/> and <https://digital.library.unt.edu/ark:/67531/metadc1757989/>, respectively.

The transcription automatically generated by Zoom was manually corrected using the original recording and then anonymized by redacting personally identifiable information (e.g., participant name, research language, names of archives). Participants' names were replaced with codes, such as *user_1* or *arch_4*, to indicate whether the participant spoke primarily from an end-user or language archive manager perspective, respectively⁸. Anonymized transcripts were then coded for major themes using NVivo 12 Plus, a commonly used qualitative analysis software. Initially based on the interview guide, the hierarchical coding scheme was expanded as appropriate with additional topics discussed in interviews⁹. All interview transcripts were coded by the first author. To ensure validity and minimize coder bias, one third of interview transcripts were coded by two project team members. Interview transcripts with archivists were coded by project team members with practical and theoretical background in information science. Interview transcripts with end-users and depositors were coded by project team members with a background in linguistics and experience as depositors. The level of agreement between the coders was 95.7%.

3. Findings Language archives are built with two major audiences in mind. These are depositors of language data and various potential end-users of these materials: researchers (linguistics and others), language communities, students, educators, artists, etc. Based on the findings of our study (including interviews and observations), this section reviews the concerns that language archive depositors and end-users expressed in relation to their experiences with language archives. These include comments on the depositing process, general accessibility of archives, access for language communities, end-user preferences regarding navigation on language archive websites, alternative dissemination venues for materials, and patterns of language data re-use in research.

3.1 Depositing process Based on the availability of resources and other factors, some repositories offer both self-archiving and mediated deposit options, while others only provide one of the two options. Self-deposit is the process by which a researcher independently uploads their files and fills in metadata to describe these files. In the process of self-deposit, researchers can refer to depositing guidelines posted on the archive's website or other materials (e.g., instructions from training sessions). A mediated deposit is any deposit process in which the depositor receives assistance from the archival staff in uploading material and/or creating metadata. For example, according to Tillman (2017), methods of mediating deposits by university repositories include:

⁸ Both depositors and general end-users were anonymized with the code *user* because these roles commonly overlapped for our participants.

⁹ The hierarchical coding scheme can be found at <https://digital.library.unt.edu/ark:/67531/metadc1757991/>.

- Working directly with researchers who approach the archive about depositing at their own initiative or in response to broader outreach events or targeted outreach, but materials are deposited by information professionals (e.g., librarians);
- Working with heads of departments or other campus units to identify faculty publications, which are then deposited by information professionals;
- Using researchers' CVs and/or websites to identify publications, which are then deposited by information professionals; and
- Harvesting publication metadata from major indexing sites into the archive.

As shown by this listing, the extent of assistance researchers receive as part of the mediated deposit process in repositories can vary substantially, and a mediated deposit process can be initiated by the researcher or by archival staff, with or without the researcher's participation.

Participants shared both positive and negative experiences with language archives' deposit processes. One depositor who had experienced both self-deposits and mediated deposits explained the advantages and disadvantages of self-deposit in the following way: "It's tricky...it gives you autonomy, in the sense of like, 'You can do it!' but it's also really, really time-consuming" (user_7). Another depositor described their experience using archival software to self-deposit as "a really clunky process...like this horrible ring of hell that you occupy trying to get metadata done" (user_5). Similarly, a language archive manager expressed that self-deposit "just transfers work to the depositor" and "forces the depositor to learn more about what the archive system is on their own and go to training workshops and then hope they remember it in two years" (arch_7).

One of the primary issues discussed in interviews was the process required to get data into the language archive. Depositors expressed frustration with navigating specialized software required by some archives: "It is a totally time-consuming multi-step process of 'First start your metadata in [software A], and then finish it off in [software B]'" (user_5). Depositors use their own data management practices while in the field but find it challenging to transfer their metadata into an unfamiliar program. One depositor described their preference for spreadsheets: "Everybody knows how [spreadsheets] work – you don't have to be reminded how to do it. But, if I didn't get to [software B] for a few months, then I had forgotten what I had to do" (user_6). In addition to spreadsheets, depositors who had experienced multiple archival-depositing methods favored Web-based metadata editor tools: "I really did like the [archive] interface for the deposit because it was very intuitive. It is an interface that is user-friendly, and you can just add things as you go" (user_7).

Depositors using SayMore (software widely used to manage language data) reported difficulty ingesting their SayMore metadata into archival software: "You can add new [metadata] fields and name them [in SayMore], but those don't get exported to [software A], so, what's the use?" (user_7). Additionally, the process of representing the relationships between files was often not satisfactory for depositors. Though

linking items is possible within a session, one depositor noted that, for items across multiple sessions, “it will be really useful to be able to say, ‘These two things are connected to one another’” (user_7).

Because depositors consistently expressed that they did not have sufficient time for archiving, some enlisted research assistants to help with data management. They provided their files and the metadata created during fieldwork and asked the research assistant “to do a lot of the processing stuff [in software B], to do most of the metadata” (user_5). Training the research assistant, however, can be similarly time-consuming and challenging: “[Training a research assistant] is very difficult and sometimes you think, ‘it’s better if I just do it myself’” (user_3).

Despite the LSA’s 2018 statement on recognition for archiving in tenure, depositors reported feeling demotivated to archive due to the historic lack of recognition given for archival deposits:

- “It’s all of this time with no reward...you don’t get any recognition hardly for the hours you put into archiving” (user_5).
- “Academics require certain things in order to keep their jobs, or get a job, or get a promotion, and archiving ain’t one of them” (user_6).
- “No, [archiving] doesn’t quite count as research, no” (user_3).

3.2 General accessibility issues for end-users Our study revealed general accessibility issues for end-users (interface language, mobile compatibility, lack of metadata) discussed in this section, in addition to more specific ones detailed in later sections. A major barrier affecting language communities is the language of the archive’s interface. Language archive interfaces and metadata are largely monolingual – only one language archive manager interviewed reported using a multilingual interface and non-English metadata. Though English is widely spoken as a second language, this is not the case for everyone:

My contact language is French in [region], but this is an issue because all of my archiving is done in English. To make my archive [collection] accessible to the broader linguistic community, it needs to be in English, but this means that I’m essentially excluding the entire speech community. (user_5)

Another depositor noted, “[Multilingual interfaces] would be something that I think would increase usability across the board...it would increase the use for people who we’re collecting the data from” (user_7).

Depositors are concerned that language communities will be unable to access archival material without a functional mobile interface. For example, one depositor explained the difficulty of accessing an archive on their mobile phone in the field: “Let’s be honest, Internet connectivity, not so much. People don’t have computers. I don’t think [archive] has the best mobile interface” (user_5). Depositors worry that the multistep process may discourage end-users who are less familiar with accessing data from archival platforms. For example, “[end-users] have to get a username in order to just view the things” (user_5). When end-users are able to access the data,

some items are only available for download or view in a software-specific file format (e.g., .eaf, .flectext) that cannot be opened on mobile devices, rather than streamable or plain text formats (e.g., .mp4, .mp3, .pdf, or .txt).

Respondents who use or intend to use data deposited by others in language archives noted that deposits usually lacked adequate context (usually provided in metadata records) for them to put the data to use in their own projects. For example, end-users commented:

- “I find [searching for data] very difficult, sort of hit or miss...I don’t know if the resource is what I’m looking for until I go dig around in it” (arch_7).
- “I wish the descriptions were more descriptive” (user_1).
- “There wasn’t a lot of super informative information about what was in that file” (user_8).

End-users of language archives expressed frustration with the difficulty in downloading data from archives: “[I need to] do a ‘smash and grab,’ download everything at once, figure out how to process it.... How would I systematically gather data?” (user_1). End-users also indicated the need for expanded search functionalities, such as the ability to refine search results (e.g., display only audio with accompanying transcription). One end-user gave the following example: “[I want] to search by transcript because if I’m undertaking a project that’s text-based only, I need transcriptions” (user_1). To determine whether they would like to download it, end-users commented that it would be useful if “you could see a preview of what the data actually are...basically just having text and the annotations themselves be shown [on the] front end” (user_5).

Participants of our study noted that once the data are downloaded, machine readability is another major factor affecting the usability of data: “So many things are possible if [data]’s starting out as machine readable” (user_8). For example, how a FLEx project is exported can significantly impact the pre-processing work required: “If people even just uploaded their FLEx XML, we cut down hours or weeks of work in terms of being able to access the data” (user_8). The encoding of fonts and special characters is also paramount for interpreting language data. Another end-user explained, “Some of the languages have their own script. We might want to put some words in phonetics [IPA]. With the arrival of Unicode, [we] have the ability to put in a good transcription that actually reflects the way that language works” (user_6). Without these annotations, primary data are accessible only to those who are familiar with that writing system; other end-users will need a transcription in IPA or another writing system in addition to a translation. End-users were not able to easily determine what portion of the data had been annotated:

- “You don’t often know what the ELAN files contain until you download them...then you open it up and realize it doesn’t have anything you were looking for” (user_5).

- “It’d be super useful to know which are the files that are fully glossed and which are the ones that are partially glossed? And when we say ‘partially,’ how partial is that? How much data is it actually? How many words? How many clauses? And then, for those words, how many of them have a translation? How many of them have morphological segmentation? How many of them have glossing?” (user_8).
- “The majority of the videos lacked subtitles, so I only got through because I knew [language]” (user_2).

Finally, depositors noted their dissatisfaction with the usability of their data: “As much as we’d like to think otherwise, language archives are still a data graveyard.... There’s such a long way to go to actually make archive data into something usable” (user_5). One depositor described their effort to encourage end-users to access their deposit: “Every example in the grammar is tagged with its [identifier] in the archive. If somebody is interested in this particular sentence...they could look at it in its broader context...so I’m hoping that it will give people a reason to access the archive” (user_5).

3.3 Language archive managers’ perspectives on language community access

Our interviews revealed steps that language archive managers are taking to encourage language communities to access or deposit in archives. One archive manager described ensuring their metadata creation guidelines were easy to follow with community depositors in mind: “We created those guidelines cognizant of the fact that many of our depositors, all of them really, were not native English speakers, often not in any way trained in data management” (arch_6). Archive managers also described future plans to accommodate language communities, such as creating profiles for speakers, which may include “...a picture of the person...a little bio about their lives and themselves...demographic information that is useful for searching – male, female, tribal affiliation...a bunch of links to all the stuff that they’re in, which is generated automatically” (arch_7). Speaker-centered representations like this may benefit language community end-users who may “just wanna hear a story or [say] ‘Look, that’s my uncle talking!’” (arch_2). Another archive manager discussed their efforts to survey the language communities using the archive:

We did a survey a year ago, inviting people to share their thoughts, but it was about 90% non-Indigenous people who responded. It’s harder to access Indigenous people and communities, who are our ideal audience, but it’s quite hard to access them and find out why they are or aren’t using it.... My feeling is that people in [region] communities where the languages are spoken aren’t necessarily using the materials, but it’s very hard to identify something that’s not happening. I feel like that’s going on, but how do you track that down? Are they not accessing it because they don’t know about it? They don’t know how to use it? They don’t see the value in it? (arch_1)

Language archive managers reported additions to available metadata fields to facilitate language communities' searching, such as including coordinates "because some of the obvious geographic information just doesn't apply in [region] communities" (arch_2). Dates may also be represented in a variety of ways: "Time is different, the calendar system is different, so if you want to respect local calendars, you can't use our twelve-month system." They explained that "for anything that's community-oriented, you have to add that in" (arch_2).

Another approach to accommodating language communities is to create separate, community-controlled resources containing all or some of the same items as the language archive. This practice, also called 'co-archiving,' may manifest as a brick-and-mortar community library of printed materials and hard copies (e.g., CDs, DVDs), or as another Web platform, for example, a Mukurtu instance. Despite the clear benefit of communities controlling access and representation of materials, archive managers reported that "the understanding of how [co-archiving] should work for access is not still not clear" (arch_5). To improve this understanding, they suggested developing a "more unified approach to what it means to have 'local control' in many places, [where] there are different political factions within the community, and it's not at all clear that a particular faction has the authority to say that they have control, or that they can make those kinds of decisions" (arch_5). Language archive managers strive to provide access for language communities but are, in some cases, hindered by insufficient information about the barriers these communities face or by factors outside of their control (e.g., the digital divide).

3.4 Preferences for navigation on language archive websites Navigation – different ways of searching and browsing in a language archive – is enabled by metadata that represent materials. For example, for the end-users to be able to find a language material by the name of the depositor, that piece of information should be included in the metadata record (e.g., in the Contributors field as shown in Figure 1). End-users and depositors who participated in our study identified their preferences for tools that support information discovery incorporated in the design of existing language archives' websites. Multiple participants noted the usefulness of a map-browsing feature with points indicating the location of a language. They found this feature helpful for navigating to materials they are interested in for research purposes:

- "As a linguist looking for particular languages, the map function is usually the most straightforward way to get what I need" (user_5).
- "I used to use the map to navigate because I know where the languages roughly are, and, personally, I'm very visual" (user_7).

Further, many noted the benefit browsing by map could offer non-linguists: "For exploration, I think [the map] would be nice for non-academics using the archive" (user_2). Browsing by map was widely regarded by our participants as a user-friendly feature of many language archives.

Unsurprisingly, our respondents often reported primarily searching and browsing based on the language name. As such, a language archive website with "a landing

page that had all the names of the languages organized alphabetically” was preferable because the end-user “could just click on that [language] and it takes you to the resources on that language” (user_7). End-users of language archives may also be interested in sociolinguistic information. One participant explained, “I was able to search via where the interview had taken place, which wasn’t necessarily the birthplace of the speaker...I needed to be sure of the dialect, and that wasn’t metadata that was recorded” (user_2). Sociolinguistic researchers may also desire information on “other languages spoken by the consultants and by the interviewers that might have an impact” (user_2) to be included in descriptive metadata. Another navigation feature of interest identified by our end-user and depositor respondents was the links between related items and representation of the specific kind of relationship between them. End-users wanted to know, “if you have an audio file and ELAN file...what’s their relationship? Are they children of the same parent? Which is the text?” (user_4).

In addition to sharing attitudes toward existing navigation features in language archives, participants discussed their ideal tools and interfaces for searching and browsing. For example, one end-user described a layout consisting of “a blank search bar, a button where you say ‘advanced search’...boxes to tick and tick, and then a place where you can go for regular expressions if you want” (user_4). Though keyword-search and advanced-search options are present in most language archives, the ability for end-users to enter regular expressions and combine fields (e.g., resource type and language) is not universal. After an initial search in the language archive, end-users may want to sort the results. One end-user gave the following hypothetical: “You’re interested in child-directed speech, and so you look up something like that. And then you’d get lines from texts in different languages. And you could sort it by language” (user_5). End-users advocated for autocompletion for both search terms and metadata entry, explaining that it may “push [end-users and depositors] towards the standard in a smooth way without them feeling like it’s some big burden” (user_4).

3.5 Disseminating data beyond language archives Community control and accessibility issues are part of the reason that linguists are turning toward less formal methods of archiving alongside depositing in language archives. For example, our study findings indicate that social media such as YouTube, WhatsApp, and Facebook are preferred ways of disseminating materials to language communities. This is explained by the ubiquitous nature of these tools and their functionalities for social groups. Depositors choose to rely on social media tools for disseminating (often, in addition to formal archiving) despite understanding the known advantages of archives over proprietary social media in terms of sustainable preservation, security, and searchability. As one depositor explained, “the IT people were saying ‘You shouldn’t do YouTube because it could be closed at any moment.’ And that’s all true, but it’s so easy for those people [in language communities] to use” (user_6).

Beyond social media, depositors were found to commonly use general-purpose open-access repositories (e.g., Zenodo) and personal website platforms (e.g., Wix, WordPress, Weebly) because of the comparative ease of uploading material to these services in contrast to the more involved process of depositing in an archive. One

participant plainly noted, “You come to depositors and you say, ‘Okay, you have to fill in all this metadata – hundreds and hundreds of fields for thousands of files,’ and then, people don’t like it” (user_4). However, just like with social media, the longevity and security of data deposits are not guaranteed, and the navigation interface does not take into account the needs of end-users searching for language data. This makes these options preferable only as a mirror of the material housed in a trusted repository.

Depositors are often aware that these alternative venues for dissemination of language data are not sufficient to replace the formal archiving process: “If I told the NSF, ‘I’m going to upload all my materials to YouTube,’ they’d be like, ‘No, hell no’” (user_5). However, some participants noted the relative ease of posting material on social media or a personal website in comparison to archiving formally, explaining:

Somehow, on normal social media like Instagram.... You can just upload it...a lot of the transaction costs for the depositor are extremely low. And yet the findability and the accessibility is extremely high. [With archiving,] the burden on the depositor is extremely high. And yet the findability and accessibility is extremely low. (user_4)

The interactive component of social media not only facilitates depositors’ dissemination of their material, but also provides depositors some insight into how their materials are being used or accessed by language communities. In our interviews and observations, depositors illustrated this advantage with the following examples:

- “The other day...I saw in my [Facebook] newsfeed, a video that I had made with my consultant who lives in [city] of her telling proverbs, and it was on YouTube. Somebody else had shared it in that group and they were all like, ‘This is great – It’s so good!’ and I’m thinking, ‘This is how the community actually accesses materials – not through an archive’...I feel like nobody has ever downloaded anything I’ve ever put in the archive at all” (user_5).
- “My impression would be that people in [language] communities would have little idea of what is online, which is why, every so often, I do one of these [Facebook] posts telling people on Facebook about everything I put up in an archive. I was trying to show people how they could access the archives, but I’m not sure it’s going to work.... When I’m trying to get my videos back into the community, I don’t even attempt to tell them how to do it through the archives, but what I have been doing is setting up YouTube channels” (user_6).

These posts may serve to direct potential end-users to the language archive, or another place where the archived material is duplicated, such as a YouTube channel. A language archive manager also voiced support for the tactic of using social media to alert end-users – particularly language community members – to the materials available in a language archive: “Rather than narrow-casting, you’re broadcasting:

you're spreading [archival data] into different places...people won't necessarily find out unless they know the [depositor]" (arch_6).

3.6 Patterns of language data re-use As an alternative to accessing language archives, linguists often obtain data through informal channels of scholarly communication. For example, a researcher may email another researcher working on a related language to request data (e.g., an example of a causative construction). One end-user commented, "Every project I've worked on has been data that I've gotten through connections with somebody" (user_8). While this is a common method for researchers who know each other, this sort of communication may not be possible for colleagues not yet connected through professional channels. Another participant compared accessing archival data to more traditional sources, similar to those in use before archival data was digitally available:

A reference grammar...is an archive, if you think about it, just in paper form – it's a collection of all the information [linguist] has been able to gather and put together in some coherent form, and I use her grammar in much the way I would use a digital archive: I search through it, I extract examples, or her summary of examples, and I use that as hypothetical data for the types of work I do in linguistics, like writing a paper about relative clauses. So we've been using archives already, even before the digital manifestation of large-collection archives took more center stage. (arch_2)

Researchers who access archival data may choose to communicate informally with the original depositor: "The first thing I would do [when searching for language data] is contact the depositors and contributors to the archive to ask for their permission to use those data, whether they're publicly accessible or not" (user_3). When researchers re-use archival data, they are, of course, expected to cite it. One participant noted that this is easier when "the deposit has a DOI," or a digital object identifier (user_4). They went on to explain the convention of one archive that allows end-users to cite "individual sentences from the corpus just using the DOI...the deposit's DOI-001 or 002," for example (user_4). Many archives offer automatically generated, recommended citations for items or collections, though they may not include DOIs.

4. Discussion The study reported here aimed to capture the perspectives of language archive depositors and end-users on their experiences with language archives. Among our respondents, needs that are currently met include stable repositories for language data that satisfy funding agency requirements and geographic browsing functionalities. Actual and potential end-users and depositors discussed unmet needs, such as the inability to bulk download and barriers to language community access. In terms of information organization, our findings revealed end-users' issues with search and browse functionalities, interface navigation, and lack of detail included in metadata. Findings related to data-sharing and citation practices give insight into the information behavior of end-users and depositors. Finally, many of our

findings pertain to digital curation, like the depositing process, archiving workflows, and longevity and machine readability of materials.

Many of the unmet needs stem from lack of funding for language archives (Burke et al. 2021) and training. Though funding to sustainably support language archive infrastructure is less easily obtained, some current initiatives offer partial solutions. This section reviews some responses to the issues discussed: mediated deposit options, formal and informal training in archiving, and novel approaches to disseminating archival data. The language archiving community has been developing some of these responses for years, while other tactics are emerging more recently.

4.1. Mediated depositing process As observed in our study, depositors may feel overwhelmed by a self-deposit option depending on their previous experience with metadata creation, the items being archived, available time, and their technical skills. In self-deposit workflows, depositors are responsible for creating metadata. Many archives offer a mediated deposit process wherein metadata are created collaboratively by both the depositor and archival staff (Tillman 2017). With either option, depositors are able to contact the archival staff with any questions about software, file types, or metadata creation that are not addressed in the archive's training materials. Research demonstrates that mediated deposits can increase the rate of deposits and decrease the barriers to researchers contributing to institutional repositories: the amount of time required and lack of confidence and familiarity with the interface and metadata creation (Daoutis & Rodriguez-Marquez 2018). As these are the primary issues participants of our study noted, many depositors may prefer a mediated deposit process.

Offering a mediated deposit option requires significantly more labor from the archive staff and was estimated to dramatically increase costs (Bevan 2007), which may not be possible without additional funding directed to language archives. However, overspending from projected costs are more than twelve times higher for institutions offering self-deposit options when compared with repositories with mediated deposits (Burns et al. 2013). Further, Kurtz (2010) found that metadata in mediated deposits was more complete, consistent, and accurate across the board. The studies discussed here focused on broader institutional repositories rather than specifically on language archives. However, we believe it is reasonable to assume their findings are applicable to language archive depositors, many of whom are faculty members at universities.

4.2 Non-degree training in language archiving A number of the difficulties language archive depositors reported in our study may be partially addressed by increased training. Recently, several language archives have created training for depositors and potential depositors. These courses may include general content in addition to recommendations conforming to an individual archive's specifications, such as the Endangered Languages Documentation Programme (ELDP) trainings, which provide background in language documentation methods and experience with relevant hardware, software, and archiving (ELDP 2020). Similarly, the Collaborative Digital Language Archiving Curriculum offered by CoRSAL (Computational Resource for

South Asian Languages) is designed to guide language communities through the archiving process, including recommended software and ethnographic factors of language documentation (CoRSAL 2020). See also the thorough instructions and variety of resources made available to depositors via several language archives' websites (e.g., AILLA¹⁰, PARADISEC¹¹, California Language Archive¹²). Finally, the online course Archiving for the Future¹³ developed by AILLA provides a comprehensive background in digital curation for language documenters planning to deposit in any archive, from file-naming conventions and formats to appraisal and considerations for long-term discoverability (Kung et al. 2020).

In addition to online courses, linguists may hold training workshops for language community members in person rather than virtually. The Collaborative Digital Language Archiving Curriculum was developed through multiple iterations of face-to-face workshops offered by Chelliah and Burke through June 2019 to February 2020 in Manipur and Delhi. Also in northeast India, the Training and Resources for Indigenous Community Linguists (TRICL) program aims to connect researchers and language communities to support each other in language documentation activities (recording, analysis, transcription, translation) as well as archiving matters (Centre for Cultural-Linguistic Diversity Eastern Himalaya n.d.). Dependent on factors such as time zones, availability of technology, and bandwidth capacity, these face-to-face options may be more accessible to language communities seeking training than their virtual counterparts.

4.3 Language-archiving training as part of academic degrees As archiving primary language data becomes solidified as standard practice, this training will continue to be necessary for future linguists. In addition to non-degree training courses, some courses in linguistics programs offered at universities include content on language archiving and metadata creation (Berez 2015). Field Methods and Tools courses introduce students to data management techniques in addition to field methods and linguistic analysis software. Some provide students with hands-on experience recording language data from language users and maintaining a shared repository of recordings and derivative material (transcriptions, elicitation notes, etc.) throughout the course. Textbooks and handbooks on language documentation have sections dedicated to managing files and preparing for archiving (e.g., Bower 2008; Chelliah & de Reuse 2012).

It is not only linguists who need this training. Because language archives are so often part of institutional repository infrastructures, information professionals also require adequate knowledge and skills to effectively handle language data. Given the rapidly growing volume of deposits and known information access issues in

¹⁰ <https://www.ailla.utexas.org/site/depositors> (Accessed 2021-11-01.)

¹¹ <https://www.paradisec.org.au/deposit/> (Accessed 2021-11-01.)

¹² <https://cla.berkeley.edu/for-depositors.php> (Accessed 2021-11-01.)

¹³ <https://archivingforthefuture.teachable.com/> (Accessed 2021-11-01.)

language archives, we believe degree programs for information professionals should be reviewed to ensure they prepare graduates for work with language archive depositors and end-users. One way to achieve this is by offering combined courses bringing together linguistics and information science students. Preliminary results of experimentation in this area are reported by Zavalina and Chelliah (Buchanan et al. 2020), who describe their experiences co-developing and co-teaching a course with the focus on language-archiving issues to both linguistics and information science students. In this experimental course, students worked in interdisciplinary teams to develop metadata creation guidelines and metadata for different kinds of language data. Though bringing both groups to a common understanding proved to be challenging, the teaching team identified improvements for future offerings of the course.

A promising alternative to combined courses is incorporating specialized modules aimed at supporting language archiving into courses for information professionals on digital libraries, digital humanities, or data curation. For example, Zavalina and Burke incorporated a module on language documentation, language archiving, and linguistics-specific metadata into a spring 2021 advanced course in the information science program at the University of North Texas. A similar effort is underway at Tribhuvan University of Nepal and Southern Illinois University Edwardsville (SIUE); students from a wide range of backgrounds receive training in creating digital humanities exhibits and get hands-on experience creating metadata and annotations for language data (Hildebrandt et al. 2019). Such collaborations aim to bridge the knowledge gap between linguists and information professionals by developing common ground and shared terminology. By extending the training available in information science, information professionals will be better prepared to provide access to language data through digital repositories. Further, interdisciplinary collaborations may help language archive managers keep apprised of changing legislation of particular relevance to language archives, such as the General Data Protection Regulation (GDPR) in the European Union.¹⁴

4.4 Multidimensional dissemination strategies Publicizing the outputs of language documentation projects is identified as an integral component of our field (Himmelmann 2006; Woodbury 2011). Both language archives and depositors are exploring innovative ways to bring awareness to their archival collections. As discussed in §3.5, depositors often post about their language archive collections on social media. These posts often include links to the archive, other locations where the material is available (e.g., YouTube, a personal website), or even a subset of the archived material (e.g., a video or album of photographs). This is especially effective when depositors are connected to language community members on social media. Similarly, language archives post regularly on social media to feature new collections and direct Web traffic to the archive, such as the example shown in Figure 2.

¹⁴ <https://gdpr-info.eu/> (Accessed 2021-11-01.)

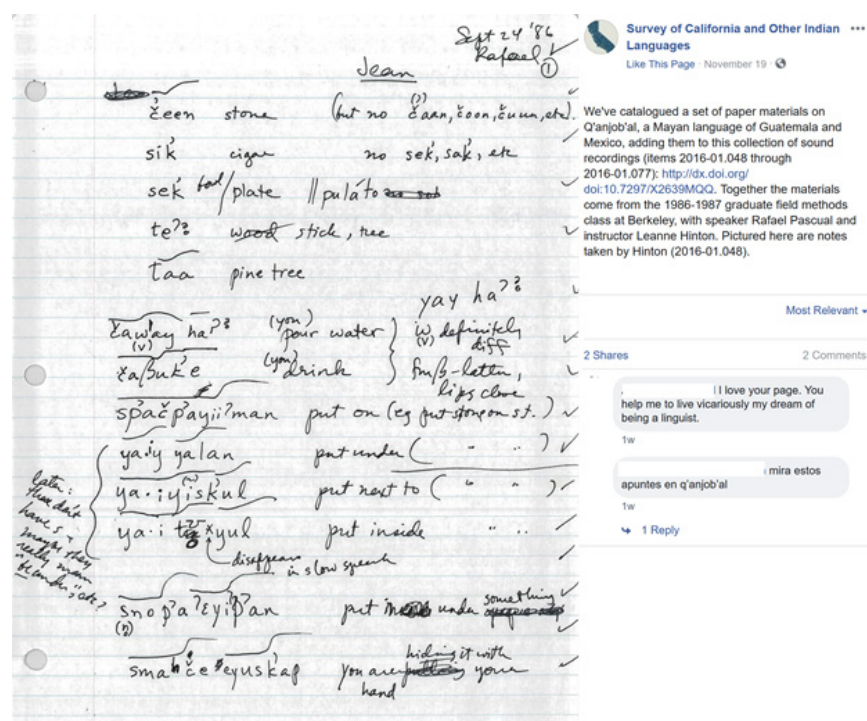


Figure 2. Example of a social media post used to promote collections (Survey of California and Other Indian Languages 2020)

Social media outreach like this Facebook post from the California Language Archive is used to highlight collections in the archive, particularly those which involve prominent linguists.

Depositors may also create guides to their archival collections and publish them as collection overviews in *Language Documentation & Conservation*. Similar to traditional archival finding aids (Cox 2008), these collection guides provide additional context for the collection, such as the sociolinguistic and cultural situation (Oez 2018), project history (Salffner 2015), and grouping of the material (Franjeh 2019). Gawne (2018) and Hildebrandt et al. (2019) synthesize multiple collections housed in different repositories. These collection guides offer end-users the mechanisms of navigating these deposits by defining the tags used (Oez 2018) and highlighting linguistic points of interest (Caballero 2017). Depositors are also able to indicate the level of annotation of primary data (Franjeh 2019) or future plans for additional annotation (Vaughan 2020), which was identified as crucial information by participants of our study. Depositors can provide more extensive context than would be possible in the free-text collection-level description field in archives. Detailed collection guides point readers to potential applications for re-using archival

data, whether for linguistic or cultural research. For a comprehensive discussion of collection guides' contents, see Sullivant (2020).

5. Concluding remarks Since the first digital language archives were created in the 1990s, we have seen exemplary applications of archival data in language revitalization and repatriation (Bischoff et al. 2013; Baldwin et al. 2016; Shepard 2016; Harris et al. 2019). Being a relatively new phenomenon, language archives have made significant strides forward in the overall accessibility of digital language data. To support their continued development, the exploratory study reported in this article provides empirical data on language archive users' needs and supports some anecdotal evidence of known issues facing language archive end-users, depositors, and managers in primarily academic contexts.

This article reviews possibilities and initiatives – both ongoing and emerging – for addressing these and other relevant issues. This in-depth qualitative study included only sixteen English-speaking participants. Although the design of this study was impacted by the funding agency's focus on the United States, it included six participants from other countries who use language archives. Future research should expand the breadth of populations by recruiting participants from diverse backgrounds (e.g., non-English-speaking participants or participants from Asia, South America, and Africa). Though indirect representation of language communities was achieved through interviews with language archive depositors who advocated for their language community collaborators, we were not able to directly represent language community members in our study, which constitutes a limitation. Future research in this area will need to center on the needs of these communities to ensure comprehensive representation.

Information behavior research has described the needs of historians and other humanities researchers (e.g., Stone 1982; Case 1991; Bates 1996; Duff & Johnson 2002), but linguists have not yet been considered in this research area. While there is an overlap in concepts between library-led digital archiving and community digital archiving, the goals and challenges, ethical considerations, scale, content, and impact are different for language archive end-users and depositors. The outputs of language documentation projects include item types novel to librarians and archivists (e.g., procedural texts, word lists, interlinear-glossed text), and end-users of language archives are not traditionally served by information professionals. Therefore, additional research is needed to develop understanding of the unique needs of these populations.

Our experience interviewing language archive end-users, depositors, and managers highlighted the reciprocal relationship between language documentation, archiving, and digital curation. Continued interdisciplinary collaboration between linguists and information professionals will maintain this progress. In much the same way that the widespread availability of digital recording technology impacted the trajectory of language documentation methodology in the 1980s and 90s (Simons 2017), we can see that language archives are poised to play a central role in pushing our field forward.

References

- Austin, Peter K. 2011. Who uses digital language archives? *Endangered Languages and Cultures*, April 29. (<http://www.paradisec.org.au/blog/2011/04/who-uses-digital-language-archives/>) (Accessed 2020-12-17.)
- Austin, Peter K. 2013. Language documentation and meta-documentation. In Jones, Mari C. & Sarah Ogilvie (eds.), *Keeping languages alive: Documentation, pedagogy and revitalization*, 3–15. Cambridge: Cambridge University Press.
- Baldwin, Daryl, David J. Costa, & Douglas Troy. 2016. Myaamiaataweenki eek-incikoonihkiinki eeyoonki aapisaataweenki: A Miami language digital tool for language reclamation. *Language Documentation & Conservation* 10. 394–410. (<http://hdl.handle.net/10125/24713>)
- Bates, Marcia. 1996. The Getty end-user online searching project in the humanities: Report no. 6: Overview and conclusions. *College & Research Libraries* 57. 514–523.
- Berez, Andrea L. 2015. Reproducible research in descriptive linguistics: Integrating archiving and citation into the postgraduate curriculum at the University of Hawai'i at Mānoa. In Harris, Amanda, Nick Thieberger, & Linda Barwick (eds.), *Research, records and responsibility: Ten years of PARADISEC*, 39–51. Sydney: University of Sydney Press.
- Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David I. Beaver, Shobhana Chelliah, Stanley Dubinsky, Richard P. Meier, Nick Thieberger, Keren Rice, & Anthony C. Woodbury. 2017. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56(1). 1–18. [doi:10.1515/ling-2017-0032](https://doi.org/10.1515/ling-2017-0032)
- Bevan, Simon J. 2007. Developing an institutional repository: Cranfield QUEprints – A case study. *OCLC Systems & Services* 23(2). 170–182.
- Bischoff, Shannon, Ivy Doak, Amy Fountain, John Ivens, & Audra Vincent. 2013. Creating grass roots digital Coeur d'Alene resources: The COLRC. *Northwest Journal of Linguistics* 7(4). 1–23.
- Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. New York: Palgrave MacMillan.
- Buchanan, Sarah A., Najim A. Babalola, Shobhana L. Chelliah, Adam Kriesberg, Sarah Pratt, Katherine M. Wisser, & Oksana L. Zavalina. 2020. Transforming the archival classroom for a connected reality. In Dali, Keren, Suliman Hawamdeh, & Hannah C. Gunderman (eds.), *Proceedings of the Association for Library and Information Science Education Annual Conference: ALISE 2020*, 396. Seattle: ALISE. (<http://hdl.handle.net/2142/108854>)
- Burke, Mary & Oksana L. Zavalina. 2019. Exploration of information organization in language archives. *Proceedings of the Association for Information Science and Technology* 56(1). 364–367. [doi:10.1002/prai.2019.30](https://doi.org/10.1002/prai.2019.30)


- Burke, Mary & Oksana L. Zavalina. 2020a. Descriptive richness of free-text meta-data: A comparative analysis of three language archives. *Proceedings of the American Society for Information Science and Technology* 57(1). doi:[10.1002/pra2.429](https://doi.org/10.1002/pra2.429)
- Burke, Mary & Oksana L. Zavalina. 2020b. Identifying challenges for information organization in language archives: Preliminary findings. In Sundqvist, Anneli, Gerd Berget, Jan Nolin, & Kjell Ivar Skjerdingsstad (eds.), *Sustainable digital communities: 15th International Conference, iConference 2020, Borås, Sweden, March 23–26, 2020 proceedings* (Lecture Notes in Computer Science), vol. 12051, 622–629. Cham: Springer. doi:[10.1007/978-3-030-43687-2_52](https://doi.org/10.1007/978-3-030-43687-2_52)
- Burke, Mary, Oksana L. Zavalina, Mark E. Phillips, & Shobhana L. Chelliah. 2021. Organization of knowledge and information in digital archives of language materials. *Journal of Library Metadata* 20(4). 185–217. doi:[10.1080/19386389.2020.1908651](https://doi.org/10.1080/19386389.2020.1908651)
- Burns, C. Sean, Amy Lana, & John M. Budd. 2013. Institutional repositories: Exploration of costs and value. *D-Lib Magazine* 19(1/2). doi:[10.1045/january2013-burns](https://doi.org/10.1045/january2013-burns)
- Caballero, Gabriela. 2017. Choguita Rarámuri (Tarahumara) language description and documentation: A guide to the deposited collection and associated materials. *Language Documentation & Conservation* 11. 224–255. (<http://hdl.handle.net/10125/24734>)
- Case, Donald Owen. 1991. The collections and use of information by some American historians: A study of motives and methods. *The Library Quarterly: Information, Community, Policy* 61(1). 61–82.
- Centre for Cultural-Linguistic Diversity Eastern Himalaya. n.d. *TRICL*. (<http://ccldeh.org/tricl/>) (Accessed 2020-12-17.)
- Chelliah, Shobhana L. & Willem J. de Reuse. 2012. *Handbook of descriptive linguistic fieldwork*. New York: Springer.
- CoRSAL (Computational Resource for South Asian Languages). 2020. *Collaborative digital language archiving curriculum*. (<https://corsal.unt.edu/curriculum>) (Accessed 2020-12-17.)
- Cox, Richard J. 2008. Revisiting the archival finding aid. *Journal of Archival Organization* 5(4). 5–32. doi:[10.1080/15332740802153245](https://doi.org/10.1080/15332740802153245)
- Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation & Conservation* 3. 15–50. (<http://hdl.handle.net/10125/4423>)
- Daoutis, Christine Antiope & Maria de Montserrat Rodriguez-Marquez. 2018. Library-mediated deposit: A gift to researchers or a curse on open access? Reflections from the case of Surrey. *Publications* 6(2). 20. doi:[10.3390/publications6020020](https://doi.org/10.3390/publications6020020)
- Duff, Wendy M. & Catherine A. Johnson. 2002. Accidentally found on purpose: Information-seeking behaviors of historians in archives. *The Library Quarterly: Information, Community, Policy* 72(4). 472–496. doi:[10.1086/lq.72.4.40039793](https://doi.org/10.1086/lq.72.4.40039793)

- ELDP (Endangered Languages Documentation Programme). 2020. *About ELDP trainings*. (<https://www.eldp.net/en/our+trainings/about/>) (Accessed 2020-12-17.)
- Franjeh, Michael. 2019. The languages of northern Ambrym, Vanuatu: A guide to the deposited materials in ELAR. *Language Documentation & Conservation* 13. 83–111. (<http://hdl.handle.net/10125/24849>)
- Gawne, Lauren. 2018. A guide to the Syuba (Kagate) language documentation corpus. *Language Documentation & Conservation* 12. 204–234. (<http://hdl.handle.net/10125/24768>)
- Gawne, Lauren, Barbara F. Kelly, Andrea L. Berez-Kroeker, & Tyler Heston. 2017. Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation & Conservation* 11. 157–189. (<http://hdl.handle.net/10125/24731>)
- Good, Jeff. 2002. *A gentle introduction to metadata*. (<http://www.language-archives.org/documents/gentle-intro.html>) (Accessed 2020-12-17.)
- Harris, Amanda, Steven Gagau, Jodie Kell, Nick Thieberger, & Nick Ward. 2019. Making meaning of historical Papua New Guinea recordings: Collaborations of speaker communities and the archive. *International Journal of Digital Curation* 14(1). 136–149. doi:10.2218/ijdc.v14i1.598
- Haspelmath, Martin & Susanne Maria Michaelis. 2014. Annotated corpora of small languages as refereed publications: A vision. *Diversity Linguistics Comment*, March 11. (<http://dlc.hypotheses.org/691>) (Accessed 2020-December-10.)
- Henke, Ryan & Andrea L. Berez-Kroeker. 2016. A brief history of archiving in language documentation, with an annotated bibliography. *Language Documentation & Conservation* 10. 411–457. (<http://hdl.handle.net/10125/24714>)
- Hildebrandt, Kristine. 2020. Archives: Perspectives from three scholars of Tibeto-Burman. (Presented at the Computational Resource for South Asian Languages 4th Annual Meeting, 1 October.) (<https://digital.library.unt.edu/ark:/67531/metadc1727526/>) (Accessed 2021-11-01.)
- Hildebrandt, Kristine, Tanner Burge-Buckley, & Jacob Sebok. 2019. Language documentation in the aftermath of the 2015 Nepal earthquakes: A guide to two archives and a web exhibit. *Language Documentation & Conservation* 13. 618–651. (<http://hdl.handle.net/10125/24914>)
- Hughes, Baden. 2005. Metadata quality evaluation: Experience from the Open Language Archives Community. In Chen, Zhaoneng, Hsinchun Chen, Qihao Miao, Yuxi Fu, Edward Fox, & Ee-peng Lim (eds.), *Digital libraries: International collaboration and cross-fertilization*, 320–329. Berlin: Springer. doi:10.1007/978-3-540-30544-6_34
- Kung, Susan Smythe, Ryan Sullivant, Elena Pojman, & Alicia Niwagaba. 2020. *Archiving for the future: Simple steps for archiving language documentation collections*. (<https://archivingforthefuture.teachable.com/>) (Accessed 2020-12-17.)
- Kurtz, Mary. 2010. Dublin Core, DSpace, and a brief analysis of three university repositories. *Information Technology and Libraries* 29(1). 40–46. doi:10.6017/ital.v29i1.3157


- LSA (Linguistic Society of America). 2018. *Statement on evaluation of language documentation for hiring, tenure, and promotion*. (https://www.linguisticsociety.org/sites/default/files/Evaluation_Lg_Documentation.pdf) (Accessed 2020-12-17.)
- Maxwell, Mike. 2012. Electronic grammars and reproducible research. In Nordhoff, Sebastian (ed.), *Electronic grammaticography*, 207–235. Honolulu: University of Hawai i Press.
- Nathan, David & Peter K. Austin. 2004. Reconceiving metadata: Language documentation through thick and thin. In Austin, Peter K. (ed.), *Language documentation and description*, 179–187. London: SOAS.
- Nathan, David & Meili Fang. 2013. Re-imagining documentary linguistics as a revitalization-driven practice. In Jones, Mari C. & Sarah Ogilvie (eds.), *Keeping languages alive: Documentation, pedagogy, and revitalization*, 42–56. Cambridge: Cambridge University Press.
- Oez, Mikael. 2018. A guide to the documentation of the Beth Qustan dialect of the Central Neo-Aramaic language Turoyo. *Language Documentation & Conservation* 12. 339–358. (<http://hdl.handle.net/10125/24773>)
- Salfner, Sophie. 2015. A guide to the Ikaan language and culture documentation. *Language Documentation & Conservation* 9. 237–267. (<http://hdl.handle.net/10125/24639>)
- Seyfeddinipur, Mandana, Felix Ameka, Lissant Bolton, Jonathan Blumtritt, Brian Carpenter, Hilaria Cruz, Sebastian Drude, Patience L. Epps, Vera Ferreira, Ana Vilacy Galucio, Brigit Hellwig, Oliver Hinte, Gary Holton, Dagmar Jung, Irmgarda Kasinskaite Buddeberg, Manfred Krifka, Susan Kung, Miyuki Monroig, Ayu'nwi Ngwabe Neba, Sebastian Nordhoff, Brigitte Pakendorf, Kilu von Prince, Felix Rau, Keren Rice, Michael Riessler, Vera Szoelloesi Brenig, Nick Thieberger, Paul Trilsbeek, Hein van der Voort, & Anthony Woodbury. 2019. Public access to research data in language documentation: Challenges and possible strategies. *Language Documentation & Conservation* 13. 545–563. (<http://hdl.handle.net/10125/24901>)
- Shepard, Michael Alvarez. 2016. The value-added language archive: Increasing cultural compatibility for Native American communities. *Language Documentation & Conservation* 10. 458–479. (<http://hdl.handle.net/10125/24715>)
- Simons, Gary. 2017. Developing infrastructure for interoperating digital language archives. (Presented at the Computational Resource for South Asian Languages 1st Annual Meeting, 17 November.) (<https://digital.library.unt.edu/ark:/67531/metadc1065401/>) (Accessed 2021-11-01.)
- Stone, Susan. 1982. Humanities scholars: Information needs and uses. *Journal of Documentation* 38(4). 292–313.
- Sullivant, Ryan. 2020. Archival description for language documentation collections. *Language Documentation & Conservation* 14. 520–578. (<http://hdl.handle.net/10125/24949>)
- Survey of California and Other Indian Languages. 2020. We've catalogued a set of paper materials on Q'anjob'al, a Mayan language of Guatemala and Mexico. Facebook, November 19. (<https://www.facebook.com/surveycla/photos/a.471169259487/10160680078059488/?type=3&theater>) (Accessed 2020-12-17.)

- Thieberger, Nick. 2014. Digital humanities and language documentation. In Gawne, Lauren & Jill Vaughan (eds.), *Selected papers from the 44th Conference of the Australian Linguistic Society, 2013*, 144–159. Melbourne: University of Melbourne.
- Thieberger, Nick. 2016. What remains to be done: Exposing invisible collections in the other 7,000 languages and why it is a DH enterprise. *Digital Scholarship in the Humanities* 32(2). 423–434. [doi:10.1093/lc/fqw006](https://doi.org/10.1093/lc/fqw006)
- Thieberger, Nick, Anna Margetts, Stephen Morey, & Simon Musgrave. 2016. Assessing annotated corpora as research output. *Australian Journal of Linguistics* 36(1). 1–21. [doi:10.1080/07268602.2016.1109428](https://doi.org/10.1080/07268602.2016.1109428)
- Tillman, Ruth Kitchin. 2017. Where are we now? Survey on rates of faculty self-deposit in institutional repositories. *Journal of Librarianship and Scholarly Communication* 5. eP2203. [doi:10.7710/2162-3309.2203](https://doi.org/10.7710/2162-3309.2203)
- Vaughan, Anthony R. 2020. Finding Hawu: Legacy data, finding aids and the Alan T. Walker Digital Language Collection. *Language Documentation & Conservation* 14. 357–422. (<http://hdl.handle.net/10125/24925>)
- Wasson, Christina, Gary Holton, & Heather S. Roth. 2016. Bringing user centered design to the field of language archives. *Language Documentation & Conservation* 10. 641–681. (<http://hdl.handle.net/10125/24721>)
- Woodbury, Anthony C. 2011. Language documentation. In Austin, Peter K. & Julia Sallabank (eds.), *Cambridge handbook of endangered languages*, 159–186. Cambridge: Cambridge University Press.
- Woodbury, Anthony C. 2014. Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. *Language Documentation and Description* 12. 19–36.


Mary Burke
mary.burke@unt.edu

 orcid.org/0000-0002-6498-6820


Oksana Zavalina

 orcid.org/0000-0002-3354-4923

Shobhana Chelliah

 orcid.org/0000-0002-2576-1848

Mark Phillips

 orcid.org/0000-0002-9679-6730