

**Findings from the
Workshop on User-Centered Design of Language Archives
White Paper**

**Christina Wasson (University of North Texas)
Gary Holton (University of Hawai'i at Mānoa)
Heather Roth (University of North Texas)**

25 July 2016

This material is based upon work supported by the National Science Foundation under Grants No. BCS-1543763 and BCS-1543828.



Table of Contents

1. Introduction.....	1
1.1. Goals of White Paper.....	1
1.2. Introduction to Research Topic: User-Centered Design of Language Archives	1
1.3. Research Trajectory Initiated by Workshop.....	2
1.4. The Workshop	2
1.5. Workshop Participants.....	2
2. Review of Language Archives and User-Centered Design	4
2.1. Language Archives	4
2.2. User-Centered Design	9
2.3. Bringing Language Archives and UCD Together.....	14
3. Typology of Language Archives.....	16
3.1. Type 1: Large Language Archives with Global Collections	17
3.2. Type 2: Large Language Archives with Regional Collections.....	17
3.3. Type 3: Language Archives Embedded in Larger Digital Repositories	18
3.4. Type 4: Single Language Community Archives	18
3.5. Type 5: Elder Archives.....	20
4. Stakeholder Groups and Their Perspectives.....	21
4.1. Language Communities	21
4.2. Linguists.....	25
4.3. Archivists	27
4.4. UCD Practitioners	29
4.5. Funding Agency Representatives.....	32
5. Access: Opening the Door	35
5.1. The Twin Missions of Preservation and Access	35
5.2. Opening the Door	35
5.3. Escort.....	35
5.4. Collection Guides.....	36
6. Organizations to Partner with for UCD Guidelines	37
7. Next Steps.....	39
Acknowledgements.....	40
References	41

1. Introduction¹

1.1. Goals of White Paper

This white paper describes findings from the workshop on User-Centered Design of Language Archives organized in February 2016 by Christina Wasson (University of North Texas) and Gary Holton (University of Hawai'i at Mānoa). It reviews relevant aspects of language archiving and user-centered design to construct the rationale for the workshop, relates key insights produced during the workshop, and outlines next steps in the larger research trajectory initiated by this workshop. The purpose of this white paper is to make all of the findings from the workshop publicly available in a short time frame, and without the constraints of a journal article concerning length, audience, format, and so forth. Selections from this white paper will be used in subsequent journal articles. So much was learned during the workshop; we wanted to provide a thorough documentation to ensure that none of the key insights would be lost.

We consider this document a white paper because it provides the foundational insights and initial conceptual frameworks that will guide us in our further research on the user-centered design of language archives. We hope this report will be useful to members of all stakeholder groups seeking to develop user-centered designs for language archives.

The findings presented here are based on careful analysis of video recordings of the February 2016 workshop and transcript summaries of those recordings. The recordings and transcripts have been deposited with the University of North Texas Library's digital repository and are accessible for research and reference; see Section 1.4 for more details.

1.2. Introduction to Research Topic: User-Centered Design of Language Archives

More than half of the world's 7000 or so languages are at risk of no longer being spoken by the end of this century. Online language archives provide the promise of long-term preservation of linguistic materials, and facilitate access to these resources. The information stored may encompass diverse formats, including recordings, manuscripts, and field notes. In some cases, such materials provide the only surviving record of now sleeping languages. Archival resources also document broader cultural information, including traditional knowledge.

One of the complex aspects of online language archives is that they seek to cater to diverse user groups. Most importantly, they can be a resource for members of the language community. These members may use archives for language revitalization efforts, or more broadly for access to their cultural heritage. In addition, language archives are intended for use by linguists, who may combine linguistic data from multiple archives for the purpose of cross-linguistic comparisons. There are usually additional user groups as well. The challenge is to design language archives that accommodate the needs of all user groups. A user-centered design

¹ This material is based upon work supported by the National Science Foundation under Grants No. BCS-1543763 and BCS-1543828. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

process could help make language archives more accessible and more useful to larger numbers of users.

1.3. Research Trajectory Initiated by Workshop

The workshop initiated a research trajectory that brings the fields of language archives and user-centered design (UCD) into dialogue, with the long-term goal of improving users' experiences with language archives and making them more accessible and useful to diverse user groups. Through this research trajectory, we ultimately aim to encourage a paradigm shift in language archives toward the adoption of UCD principles, parallel to the shift that occurred in the corporate world about twenty years ago. The research trajectory as a whole will encompass:

1. The workshop that launched the trajectory
2. A project involving (re-)design of two or more language archives that will enable us to develop guidelines for a user-centered design process for language archives
3. Dissemination of these guidelines to relevant organizations and communities, in order to encourage a paradigm shift, or in other words culture change

The co-organizers of this research trajectory represent the two fields that are being brought into dialogue. Christina Wasson, a linguistic anthropologist, has been active in UCD and design anthropology since 1996 (Wasson, 2000; Wasson, 2002; Wasson, 2005; Wasson & Metcalf 2013; Wasson & Squires 2012; Aiken, Schlieder, & Wasson, 2014). Gary Holton, a linguist, has been involved in international efforts to develop best practices for the digital preservation and access of endangered language materials since 1999 (Dobrin & Holton 2013; Holton, 2011a; Holton, 2011b; Holton, 2012; Holton, 2014; Holton, Berez, & Williams, 2007).

1.4. The Workshop

The immediate goal of the workshop was to map the terrain at the intersection of language archives and user-centered design by engaging representatives of key stakeholder groups in dialogue. Some of the key outcomes that emerged from our conversations were 1) mapping the diverse perspectives of different stakeholder groups, 2) mapping types of language archives and their varying relationships with user groups, and 3) identifying current access issues.

The workshop took place February 20-21, 2016, at the University of North Texas. Further details are at <https://designinglanguagearchives.com>. Video recordings and transcript summaries of the workshop have been placed in the UNT library's digital repository. They are available to the general public; the collection home page is at <http://digital.library.unt.edu/explore/collections/LANGAR/>.

1.5. Workshop Participants

Participants were selected to represent the main stakeholder groups that engage with archives for endangered languages. Many participants belonged to more than one group. They are listed in Table 1. In addition, the workshop was attended by the organizers, Christina Wasson and Gary Holton, and Research Assistant Heather Roth.

Table 1. Invited Workshop Participants

Name	Affiliations
Edward Alexander	Dinjii Zhuh K'yaa (Gwich'in Language Center and Archive) <i>Executive Director</i> Gwich'in Council International <i>Co-Chair</i>
Daryl Baldwin	Myaamia Center <i>Director</i> Miami Tribe of Oklahoma <i>Member</i>
Santosh Basapur	Illinois Institute of Technology (IIT) Institute of Design (ID) <i>Researcher</i> <i>Faculty Member</i>
Andrea Berez-Kroeker	University of Hawai'i at Mānoa – Department of Linguistics <i>Assistant Professor</i> Kaipuleohone – University of Hawai'i Digital Language Archive <i>Director</i>
Shobhana Chelliah	University of North Texas – Program in Linguistics <i>Professor</i> Documenting Endangered Languages (DEL) Program at the National Science Foundation (NSF) <i>Former Program Director, 2012-2015</i>
Susan Kung	University of Texas at Austin Archive of the Indigenous Languages of Latin America (AILLA) <i>Manager</i>
Wesley Leonard	Southern Oregon University – Native American Studies Program <i>Associate Professor and Chair</i> Miami Tribe of Oklahoma <i>Member</i>
Crysta Metcalf	Independent Consultant – Design Anthropologist Interactive Media User Research, Motorola Mobility <i>Former Director</i>
Jennifer O'Neal	University of Oregon Libraries <i>Corrigan Solari University Historian and Archivist</i> National Museum of the American Indian, Smithsonian Institution <i>Former Head Archivist</i>
Felix Rau	University of Cologne – Department of Linguistics <i>Researcher</i> Language Archive Cologne <i>Team Member</i>
Loriene Roy	University of Texas at Austin – School of Information <i>Professor</i> Minnesota Chippewa Tribe – Anishinabe <i>Member</i>
Mandana Seyfeddinipur	Endangered Languages Documentation Programme <i>Director</i> Endangered Languages Archive (ELAR) <i>Head</i> SOAS, University of London – Department of Linguistics <i>Research Fellow</i>
Michael Shepard	Goucher College – Cultural and Environmental Sustainability Program <i>Professor of Practice</i> Mohave Community College <i>Anthropology Instructor</i>
Justin Spence	University of California, Davis – Native American Studies <i>Assistant Professor</i>
Alexander Wadsworth	FirstVoices <i>Manager</i> <i>Information Technology</i> <i>Language Mapping</i>

2. Review of Language Archives and User-Centered Design

This section reviews the aspects of language archives and user-centered design that provided the rationale for the workshop. The contents are based on Gary Holton's and Christina Wasson's presentations at the start of the workshop.

2.1. Language Archives

2.1.1. History of Language Archiving

Language loss and the concomitant threat to linguistic diversity are prominent topics in both the field of linguistics and the wider public discourse. It is commonly reported that more than half of the world's 7000 or so languages are in danger of no longer being spoken by the end of this century. Efforts to save and revitalize Indigenous languages are featured in mainstream media outlets. The number of Indigenous language apps on the Apple App Store grows daily. Funding agencies including the National Science Foundation have responded with dedicated programs focused on endangered language documentation and archiving. Professional organizations have developed around language archiving, including the Open Language Archives Community (OLAC) and the Digital Endangered Languages and Musics Archiving Network (DELAMAN). Workshops have been convened to examine the Potentials of Language Documentation (Leipzig, Germany, 2011); Best Practices in Digital Archiving of Language and Music Data (Gniezno, Poland, 2013); and Utilization of Language Archives in Endangered Language Research, Revitalization, and Documentation (Portland, Oregon, 2015). All this attention on the subject may give the impression that language archiving is a relatively new field, yet in fact language archiving has a long history, especially in the United States. The history presented in this section is offered primarily from the perspective of linguistics. It might be presented differently if it were written from the perspective of Indigenous communities. It is important to keep in mind that language archives emerged within the context of Western conceptualizations of "language" and "archives," as well as a colonial power structure.

Decades before British colonists fought a war of independence, Benjamin Franklin advocated for the creation of archives in the colonies.

"The first drudgery of settling new colonies... is now pretty well over, and there are many in every province in circumstances that set them at ease, and afford leisure to cultivate the finer arts, and *improve the common stock of knowledge*" (Franklin, 1743, emphasis added).

Franklin's call led to the creation of the American Philosophical Society (APS), which eventually came to house some of the major works of documentation of the Indigenous languages of North America. The papers and field notes of Franz Boas—the father of American anthropology—today fill more than 73 linear feet at the APS repository in Philadelphia. A century later saw the creation of the National Anthropological Archives (NAA) at the Smithsonian Institution, under the directorship of John Wesley Powell. Shortly thereafter in 1901 Alfred Kroeber founded the linguistics department at the University of California Berkeley, leading efforts to document the Native languages of California, one of the most linguistically diverse places on the planet. The archival resources collected by Kroeber and his students were preserved by the Survey of

California Indian Languages, founded in 1953 under the directorship of Mary Haas, one of the most influential American linguists of the 20th century. A few decades later, in 1972, Michael Krauss founded the Alaska Native Language Center, creating an archival collection which would grow to include nearly everything written in or about each of Alaska's 20 Indigenous languages.

The history of American linguistics as a discipline is intimately tied to the languages spoken by Indigenous Americans. Efforts to document these languages led European scholars to rethink core notions of human language structure. The golden age of this Americanist linguistics—if such a period exists—can be dated effectively to the founding of the *International Journal of American Linguistics* in 1917. Writing in the inaugural issue, Franz Boas remarked on “the fragmentary character of our knowledge of the languages spoken by the American aborigines,” noting also that “for most of the native languages of Central and South America the field is practically *terra incognita*” (1917: 1). We tend to speak of language endangerment as if it were a new problem, but the threats to Indigenous language vitality were well known to linguists and anthropologists in the early part of the 20th century. The need to document and preserve a record of these languages motivated scores of scholars to devote their careers to language documentation. The records they were able to preserve, recorded using state-of-the-art technologies such as portable phonograph recorders, filled the archives of many prominent academic institutions. For example, Boas' student Melville Jacobs, who taught at the University of Washington beginning in 1928, collected audio recordings and field notes of many endangered languages of the Pacific Northwest. For some languages of Oregon, Jacobs' recordings are now the only surviving record. The Melville Jacobs papers fill nearly 80 cubic feet in the University of Washington Libraries Special Collections, providing a lasting record of these languages.

During this golden age of Americanist linguistics, language archiving was not considered a distinct field; it was simply what one did. As Michael Krauss once pointed out to Gary Holton, the legislation which founded the Alaska Native Language Center to “study the languages native to Alaska” made no specific mention of archiving. It didn't need to. For what would be the purpose of collecting documentation of endangered Alaskan languages if there were no intention of preserving the documentation itself? Language archiving went hand in hand with language documentation.

2.1.2. Digital Language Archiving

Over the past two decades language archiving – specifically digital language archiving – has emerged as a distinct field. This shift is a direct response to two, possibly three, important developments. First, the recognition that many of the world's languages and indeed the world's linguistic diversity are under threat (Krauss 1992) has resulted in a shift within some areas of linguistics toward a focus on basic language documentation. The documentary focus of linguistics in the first half of the 20th century had largely been forgotten by mainstream linguistics in the 1960's, as the profession became fascinated by theoretical models and universalism. Decades passed before Krauss famously warned: “Obviously we must do some serious rethinking of our priorities, lest linguistics go down in history as the only science that presided obliviously over the disappearance of 90% of the very field to which it is dedicated.” Two decades later, Krauss' clarion call has largely been heeded. Entire conferences and journals are now devoted to the study of documentary linguistics. Private and national funding regimes have been created for the sole purpose of supporting language documentation. The

resulting surge in documentation efforts has led to a large body of material in need of an archival home.

Second, at the same time that a group of linguists was reawakening to the problem of endangered languages, this same group also recognized an unprecedented technological transformation. As these linguists rushed to record endangered languages, technology was evolving so quickly that recording devices specified in a grant proposal became obsolete before the proposal was funded. Documentary linguists were thus forced to grapple with technological standards in order to ensure that the records they were creating would be of lasting value. An exponential increase in the ability to create language documentation data in digital form necessitated the development of standards for formats and metadata, so that digital data could be effectively managed. In the first decade of the 21st century, committed groups of linguists, archivists, and technology experts made significant strides toward the creation of standards for digital language documentation and archiving. The Open Language Archives Community (OLAC), founded in 2001, set out standards for descriptive metadata, leading most significantly to the adoption of ISO standard 639-3, which assigned unique 3-letter codes to each of the world's languages. Prior to the adoption of ISO 639-3 in 2007, the ISO standards body recognized only 464 distinct codes covering the world's 7000 or so languages. The NSF-funded Electronic Metastructures for Endangered Languages Data (E-MELD) project further developed standards for formats such as interlinear text (cf. Bird & Simons, 2003). While work on the development of standards continues, the best practices advocated by these initiatives have largely been recognized and adopted in the field. Standards for digital audio formats, text annotation, lexical data, and descriptive metadata are largely agreed upon. Archiving these data is not only common practice but increasingly required by granting agencies, academic programs, and publishers. Moreover, a number of dedicated online language archives have been created to implement these best practices and serve the increasing need for language archiving. They are listed in Table 2.

Table 2: Dedicated Online Language Archives

Founded	Archive	Location
2000	Archive of the Indigenous Languages of Latin American (AILLA)	Austin
2000	Dokumentation bedrohte Sprachen (DoBeS) / Language Archive Cologne	Cologne
2003	Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)	Sydney
2004	Endangered Languages Archive (ELAR)	London
2008	Kaipuleohone Digital Language Archive	Honolulu

The field of online language archiving bridges these two independent developments by recognizing that language documentation is of little value if it is not preserved. In other words, there is little sense in documenting languages if we have no clear strategy for preserving the products of that documentation. However, to date digital language archiving has focused almost exclusively on data preservation rather than access. To the extent that access has been addressed, the focus has been almost entirely on access by the linguistic community. That is, the assumption has been that the creators of language documentation are essentially the same as the consumers of language documentation. In a seminal paper, Himmelmann (1998) suggests that language documentation should be repurposable, that is, useful to linguists who do not actually collect the documentation (see also Woodbury, 2003). In this view, language documentations are primary data which can be used to address theoretical questions posed by persons who did not actually collect that documentation. While this may not seem a radical

proposition, it is quite novel in the field of linguistics. Although linguists routinely cite published data, unpublished data are rarely cited or indeed even accessible to anyone other than their creator. Much recent discussion has focused on the potential consequences of a looming paradigm shift that will transition linguistics to a more data-oriented science. For example, a 2011 workshop hosted at the Max Planck Institute-Leipzig addressed the “potentials” of language documentation by examining how the data accumulated in the massive archive associated with the Documentation of Endangered Languages (DoBeS) project can be used by linguists other than those who collected the data (Seifart et al., 2012).

As the field of online language archiving emerged in the early years of the 21st century, there has been a notable lack of intellectual engagement with the wider field of archiving. To a large extent linguists have gone it on their own, forging standards and best practices with little input from archivists. The reasons for this are clear. When linguists convened in December 2000 to create the Open Language Archives Community, they were motivated by the two pressing developments discussed above: the need to document and preserve endangered languages, and the need to deal with digital data. Traditional archives simply weren't ready. In 2000 the Library of Congress preservation recommendation for digital data was to copy the data to open reel analog tape. For 21st century linguists living and working in an increasingly digital world, this simply wasn't an option. Lacking extant standards, linguists chose to forge their own.

Of course the traditional archiving world soon caught up and awoke to the need for standards for digital archiving. This led to a period of parallel evolution as both archivists and linguists worked largely independently to develop digital best practices. The U.S. National Digital Stewardship Alliance now unites a growing network of institutions committed to long-term preservation of digital data. International standards for trusted digital repositories have been agreed and implemented. And there is an increasing reliance within the archiving world on content specialists who can provide the subject area knowledge to implement digital archives. Equally significant is the increased focus on usability which has emerged within the traditional archiving world. Most notable in this regard is the emergence of a participatory archive framework, which acknowledges that “usability does not denote use alone, but also denotes a deeper level of involvement in the sense of actual participation in the archive and in the archival process” (Huvila, 2008: 25). Although a relatively recent phenomenon, this renewed interest in usability has important implications for language archives. Now that many of the basic questions of preservation have been (at least partially) addressed, language archives are increasingly looking to improve the usability of their resources.

In part this is due to a third, and perhaps most critical, development which has influenced the emergence of digital language archiving: namely, the realization of the important role that archival resources, especially in digital form, can play in language conservation efforts. As work continues to document Indigenous languages and to ensure that documentation is archived, there is a concomitant need to support language maintenance efforts. Often the urgency of language maintenance is such that documentation and conservation activities become closely intertwined. Digital preservation depends crucially on access, namely, “the storage, maintenance, and *accessibility* of a digital object over the long term” (Green, 2003: 199, emphasis added). Access is a fundamental part of the archive mission to identify, preserve and make available relevant resources (Hunter, 2003). In this view access is even critical to the functioning of archives, since it leads to reappraisal and thus improves the quality of resources over time. At times, the goals of preservation and access may appear to be at odds, as the desire to create innovative and attractive products such as multimedia websites and mobile

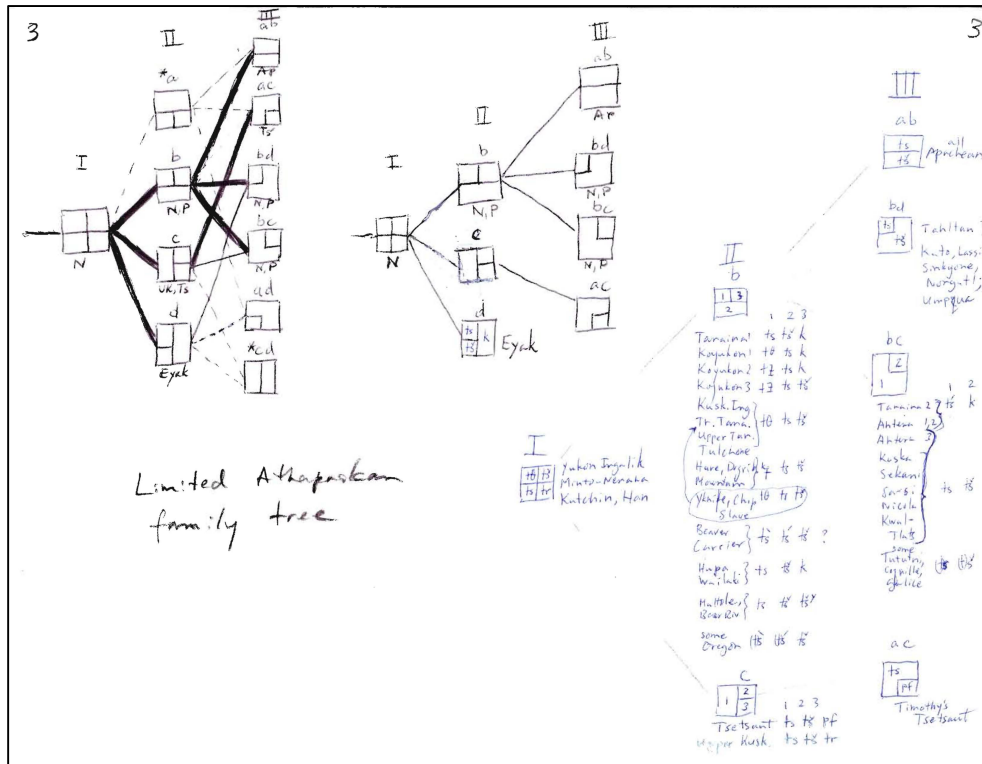
apps favors reliance on proprietary, non-archival formats which can lead to risk of data loss. However, online language archives can reconcile these seemingly competing goals by advocating archival standards to store data in a format which can be easily repurposed using both current technologies and those which will arise in the future (cf. Bird & Simons, 2003). A simple example of this strategy is the recommendation to archive audio files in an uncompressed, non-proprietary linear PCM format while simultaneously creating presentation copies in proprietary formats as needed for a particular technology application.

As language archives have evolved, some archivists have found that they are being used by people other than linguists for purposes which were not originally envisioned by either the creators of the documentation or the archivists. Language archives contain much more than just language data, and users often approach language archives in search of information about topics as devoid of linguistic content as family history and photographs (Holton, 2012). Anecdotal reports suggest that many users of language archives are speakers of the languages being archived or their descendants (Austin, 2011). Archives have made explicit attempts to serve this community. However, the vast majority of these attempts have been top-down in nature, inspired by the archives' vision of the needs of these user communities (Dobrin & Holton, 2013; Holton, 2014). Language archives have yet to attempt to understand their user communities in a structured way. Access is a fundamental part of archiving, and in the post-digital revolution era, access is in one sense easy: digital repatriation is as simple as providing downloadable files. But *access* does not necessarily mean *accessible*. Users of language archives still struggle to interpret materials which are not adequately annotated (see Figure 1 below). This is especially true for recordings that lack transcripts or translations.

Moreover, unidirectional approaches to access are unsatisfactory, especially in the endangered language context. For communities with few if any remaining fluent speakers, language archives represent a repository of linguistic and cultural knowledge. Those language communities do not merely want to be consumers of the knowledge in these repositories; rather, they desire to actively engage with and shape the record. If language archives fail to facilitate this kind of participatory access they risk remaining in an era when, in the words of William Hagan, "to be an Indian is to have non-Indians control your documents from which other non-Indians write their versions of your history" (1978: 135). The wider archiving world is well aware of the benefits that participatory archiving can bring, helping traditionally marginalized communities to "preserve empowered narratives" (Shilton & Srinivasan, 2007: 90). Noting that "archives have continuously evolved to remain relevant" in the face of changing technologies and user demographics, Mary Linn, former curator of Native Languages at the Sam Noble Oklahoma Museum, notes that "endangered language archives are well-placed to participate in and articulate these shifts" (2014: 65). As users repeatedly make clear, language archives are not merely repositories of past knowledge but rather living, forward-looking platforms for propagating Indigenous languages and cultures.

"Our digital archives are only useful if they enable us to have analog meetings and connections between actual people conversing and sharing" (Edward Alexander, quoted in Shepard, 2015: 226).

Figure 1. Document from the Alaska Native Language Archive (item CA961K1969b). Associated metadata description reads: “Critique of previous taxonomic classifications of Athabaskan languages (Hojjer, 1938, 1960, 1963; Pinnow, 1966) presents alternative approaches; examines PCA and Apachean isoglosses, Alaskan and Northern Athabaskan lexicostatistics.”



2.2. User-Centered Design

Although linguists and managers of language archives have repeatedly acknowledged the need – and indeed obligation – to provide improved access to language documentation, they lack a structured methodology for understanding their user communities and translating that understanding into the design of an archive.

A well-developed methodology for understanding user communities does exist, and it could be productively applied to the design of language archives. User-centered design (UCD) is a structured process for ensuring that technologies (and other products) are designed to meet the needs and constraints of their users. It is an interdisciplinary endeavor that involves close collaboration among users, researchers who analyze user needs, designers, and other relevant specialists. UCD emerged in the business context in the 1980s and 1990s as a response to criticisms that designers were developing products and technologies without adequately understanding the needs of users (Norman, 1988; Wasson, 2000). The chief difference from other product design philosophies is that UCD tries to optimize the product around how users can, want, or need to use the product, rather than forcing the users to change their behavior to accommodate the product (Wikipedia, 2016).

In the business world, the adoption of UCD principles constituted a significant paradigm shift about 20 years ago (Robinson, 1993; Robinson, 1994; Ritter, Baxter, & Churchill, 2014). This paradigm shift fundamentally altered business understandings, from the assumption that designers' intuitions were an adequate foundation for new product ideas, to the recognition that the products resulting from such a process had an unacceptably high rate of failure. The new work processes that emerged in industry required designers to base their product ideas on user research in order to mitigate the risk of launching new products.

Anthropologists have played a prominent role in the development of UCD, and the term "design anthropology" emerged to describe anthropologists working in this area (Squires & Byrne, 2002; Wasson, 2000). The role of anthropologists in UCD is to work with user groups to identify their needs and constraints, and then translate those insights into actionable implications for designers and others involved in the development of new products and technologies. Prior to the rise of UCD, designers had mainly worked with cognitive psychologists who tested new product ideas in laboratory settings, focusing on the interaction between one human and one product or technology (Bannon, 1991; Robinson, 1993). Anthropologists, by contrast, observed product use "in the wild," in the actual contexts where people were using those products. Researchers were able to see how interactions among groups of people shaped product use, and how users' engagements with a product were shaped by cultural systems of meaning. From the beginning, anthropological studies in the field of design showed major discrepancies between designers' intended uses of their products, and consumers' everyday behaviors (Suchman, 1995; Wasson, 2000; Wasson & Squires, 2012).

Four key components of the UCD process are:

1. Start by identifying user groups
2. Work with users to identify needs
3. Process is interdisciplinary and collaborative
4. Process is iterative

2.2.1. Start by Identifying User Groups

Early in the UCD process, the UCD team needs to ask:

- What groups of people are currently using the technology?
- What other groups might benefit from using it?

This information is needed in order to ensure that the design responds to the needs of all significant current and potential user groups. It is especially important when a resource is used by groups of people with distinctly different needs. Surprisingly few publications on language archives have sought to systematically identify user groups for language archives. The primary reference seems to be Austin (2011), a blog post on PARADISEC that reported on interviews with managers of six language archives. A few other scholarly articles examine the needs of either linguists or local communities, but do not engage in a holistic investigation of the full range of users.

Linguists and Language Communities. A first pass at identifying the main user groups of language archives might be to say that there are two groups: linguists and language

communities. This is already interesting from a UCD point of view, because the needs of these two user groups are strikingly different and sometimes at odds. An informal review of user logs at the Alaska Native Language Archive conducted by Gary Holton reveals that academic researchers typically seek linguistic information in a form which allows ready comparison with other languages, thus facilitating philological and typological analyses, while members of local language communities often seek information about particular people, such as a story or personal narrative recorded by a relative.

For linguists, language archives offer the raw data which provide the very foundation of linguistic science. Archives may be the only source of data on endangered languages. Where languages are still spoken but in decline, archives may contain data which record a more vibrant stage of the language. Where publications such as dictionaries and reference grammars already exist, language archives provide the raw data which allow verification and falsification of claims made in published sources. Archival data thus provide the scientific underpinning to linguistic research (Gezelter, 2009). However, within the field of linguistics, the ability to effectively use archival language data remains limited (Berez, 2013).

For local language communities, archives provide a repository of cultural heritage. For many small language communities, rapid language shift has brought profound change, and language communities are increasingly turning to archives to support language and culture revitalization efforts. For these users, language may be only one component embedded in a larger matrix which includes history, genealogy, music, and other not strictly linguistic resources. Though language archives may not have been designed with such uses in mind, language archives are often the best sources of such information (Holton, 2012). Documentary linguistics records information about language; yet at the same time, that language provides a medium for the transmission of other culturally relevant information. In most cases, archival linguistic data have been removed from the source communities, complicating access by local language communities. Language archives are often located within academic institutions which may not view local language communities as a primary user group. The return or repatriation of language documentation resources has become an increasing priority, but many of these efforts at repatriation have been top-down, based on archivists' and linguists' perceptions of how and why local language communities access language archives (Dobrin & Holton, 2013; Holton, 2014).

Complexifying the User Groups. The constellation of user groups of language archives actually has a much more intricate structure than just a binary opposition between linguists and language communities. Complexifying factors include:

- Users may belong to multiple groups – for instance, a user may be both a linguist and a member of a language community
- There may be additional user groups – historians, artists, students, the general public, etc.
- There may be factions within user groups
- Some users are yet unborn – there may be a concern for future generations of the language community
- There may be users who aren't users – for instance, members of a language community who don't actually want to use an archive themselves, but want to know it is there for their children and grandchildren

- Other stakeholder groups need to be considered, such as archivists; they play an important role in the design and maintenance of language archives and in guiding users through the archives

Finally, we have some major concerns about how user-centered the very concept of an archive is. As many researchers have noted, archives are constructed within a paradigm of Western science concepts and assumptions (Foucault, 1982; Isaacman, Lalu, & Nygren, 2005; Povinelli, 2011; Stoler, 2010; Zeitlyn, 2012). These concepts may or may not be a good match with cultural models of the Indigenous groups whose languages are being archived. In this sense, there is a colonial aspect to the whole principle of archiving. And the more rigid the conception is, based on predetermined definitions of what is an archive and what are the linguistic categories of analysis, the less it is user-centered for language communities.

This situation creates challenges for the development of a UCD approach to language archives. However, we feel hopeful that there can be creative solutions to designing language archives that accommodate the needs of multiple user groups with disparate interests. To give just one example, a linguist expressed concern that a language community might prioritize funding for access over funding for preservation. Yet a language community member expressed concern in the opposite direction. We envision win-win solutions that could address both needs. For instance, a language archive might have a front end and a back end. The back end could be a simple, stable database structure that would address the preservation concerns and definitions of archivists and linguists. At the same time, there could be multiple front end applications that would facilitate access for various user groups, for instance a game for a mobile device to help teach the language to young people. Such games could include elements to involve young users, such as glitzy graphics and the possibility of “leveling up,” the process of moving to ever higher levels of challenge and reward that, for many users, contributes to the fascination of gaming.

2.2.2. Work with Users to Identify Needs

Once user groups have been identified, the UCD research process begins. The goal is to map user experience. Sometimes this is called user research or design research. The process involves a holistic examination of the lives of users. Part of the research is learning how the technology may play a role in users' lives, but the research is not narrowly focused around just the use of that technology. For instance, we cannot understand language archives without examining broader issues such as the work lives and professional norms of linguists, or the history and cultural values of language communities.

Typical UCD research questions might include:

- What is the broader cultural context within which the technology can help people achieve their goals?
- What are the main ways that people use the technology?
- What would they like the technology to do that it doesn't do?
- What problems do they encounter, and how do they work around those problems?

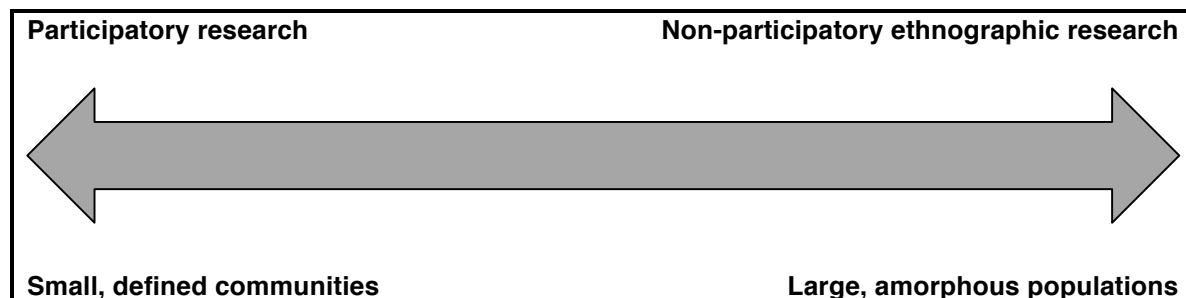
For instance, for particular local language communities, we might ask, what does their traditional language mean to them? What practices do they engage in with regard to the

language? Are other aspects of cultural heritage interwoven? What are their goals and concerns regarding language preservation and revitalization? What forms of technology are most accessible to them? For linguists, we might ask, what are their goals in accessing language archives? What kinds of information and search capabilities are important to them? What kinds of comparisons do they want to be able to make across languages?

The methods of UCD research are based on the anthropological approach termed “ethnography.” Data collection methods typically include participant observation and in-depth interviews. Both observations and interviews are often video or audio recorded, and transcribed; at a minimum, researchers write detailed field notes about them. The UCD team subsequently analyzes the transcripts and field notes (and any other materials collected) in order to identify patterns and themes.

The level of collaboration with user groups varies according to the continuum illustrated in Figure 2. In small communities or organizations, it is often possible to engage in a participatory research process, where members of the community are active members of the UCD team from beginning to end. For the UCD of a technology that is used by a huge, amorphous population, this is not as easy. For instance, when Wasson led a study on car use, there was no way to select a sample of users who could really represent and speak for the whole population of people who drive cars. In the latter case, the research typically follows a non-participatory ethnographic approach, where the UCD team does its best to understand and accurately represent the perspectives of the population, but the study participants are not active members of the UCD team throughout the research process.

Figure 2. Levels of Participation



2.2.3. Process is Interdisciplinary and Collaborative

The third key component of the UCD process is that it is profoundly collaborative and interdisciplinary. UCD is a complex endeavor and there is no one person who has all the skills necessary to create a good product on their own. So a UCD team should include, at a minimum:

- Members of the user groups whose needs the technology is supposed to meet (when the UCD team follows a participatory research approach)
- User researchers, who translate user needs into design recommendations
- Designers, who implement insights from user groups and researchers in the design of the user interface and navigational structure of e.g. a language archive
- Technologists, who do the software development

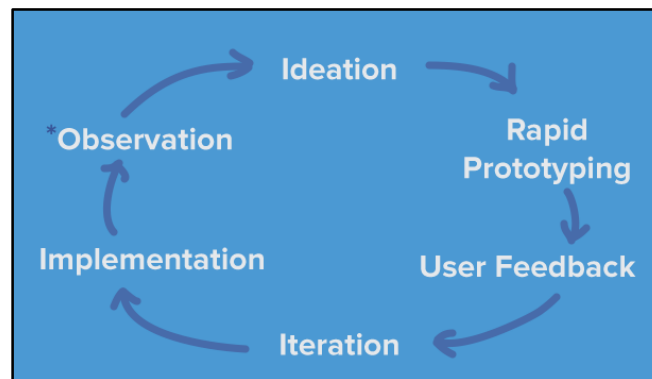
- Other experts and stakeholders as relevant; for language archives, this might include archivists, curators, and experts in preparation of language learning materials

By including all of these voices, a product can be developed that accommodates the needs and constraints of each user group, while also working with the organizational needs of the archive and the constraints of available technologies.

2.2.4. Process is Iterative

The final component of UCD is that it is an iterative process. The UCD literature is full of images with different kinds of circular patterns, as illustrated in Figure 3. The point is to show that the process is cyclical. People who design technology have learned that plans never work quite as expected. While earlier approaches to the design process were top-down and linear, newer approaches are spiral. Best practice in design is now conceptualized as the process of developing a prototype, having users test the prototype, and then modifying the design based on what worked and where users encountered difficulty. The process is usually repeated several times before product launch. Also, because technology changes so rapidly, technology designs have to be revisited every few years and updated to accommodate new developments.

Figure 3. Iterative UCD Process. Image used by permission of UserTesting. Found at <https://www.usertesting.com/blog/2015/07/09/how-ideo-uses-customer-insights-to-design-innovative-products-users-love/>.



2.3. Bringing Language Archives and UCD Together

The application of UCD principles to archive design has particular relevance for language archives. From the outset, the field of language archiving has emphasized the portability of language resources, ensuring that materials do not become locked in obsolete, inaccessible formats, and providing for language identification through standardized codes (Bird & Simons, 2003). Many of the largest language archives in North America target specific regions, including Alaska (Alaska Native Language Archive), Oklahoma (Sam Noble Oklahoma Museum of Natural History), California (Survey of California and Other Indian Languages), and Latin America (Archive of the Indigenous Languages of Latin America). Linn (2014) argues that such regional language archives have been at the forefront of what she refers to as “participatory

delivery” of language resources. This makes language archives an ideal domain in which to apply UCD principles.

As the use of language archives has burgeoned in recent years, making these archives more accessible has become an urgent need. In spite of recent advances in the standards and protocols for language archives, users of these archives often have difficulty accessing the information they seek (Woodbury, 2014). Until now, developers of language archives have been unfamiliar with the principles of user-centered design (UCD). Bringing the fields of language archives and UCD into dialogue has the potential to significantly improve users’ experiences with language archives, in terms of their ability to navigate the sites and obtain the information they seek. Informed by UCD, language archives could be designed or redesigned to accommodate user groups’ cultural practices of data sharing, access, and use.

3. Typology of Language Archives

A key outcome of the workshop was the development of a tentative typology for language archives. It was constructed through a group activity led by UCD practitioner Crysta Metcalf. The typology is valuable because it suggests how a UCD process might be customized to different types of language archive.

Workshop participants grouped language archives into five types; the first four are summarized in Table 3, while the fifth is discussed below. Everyone at the workshop recognized that this typology was a simplification of the actual situation. Nevertheless, the typology is useful in that it helps UCD practitioners learn about patterns of similarity and difference across language archives.

Table 3. Typology of Language Archives

Type	Type 1	Type 2	Type 3	Type 4
Descriptor	Large language archives with global collections	Large language archives with regional collections	Language archives embedded in larger digital repositories	Single language community archives
Contents	Global collection	Regional collection	Often regional but not always	Single language (or possibly 2 or 3 languages from a single community); may include cultural, historical, etc. materials
Primary Mission	Preservation/documentation	Preservation/documentation	Preservation/documentation; may extend to language revitalization	Serving a language community; contributing to language revitalization
Source of Collections	Linguist depositors, typically tied to funded research projects	Linguist depositors	Linguist depositors	May come directly from linguist depositors; may be copies of collections in other archives
Intended Users	Broad base of users	Broad base of users	Broad base of users; parent repository may be required to serve certain constituents, e.g. university serves students, faculty	Language community; may also choose to be accessible to broad base of users
Funding	Private grants	Government or private grants or university support	Parent repository, most often a state institution	Community, maybe grants
Examples	ELAR DoBeS/Language Archive Cologne	PARADISEC Calif Lang Archive AILLA Sam Noble ANLA	U Oregon Libraries U Hawai'i Kaipuleohone	Dinjii Zhuh K'yaa Myaamia Center FirstVoices is conglomeration of Type 4s

Workshop participants noted that language archive Types 1-3 tend to be managed by linguists, while Type 4 will probably hire someone (if they have funds) with a more general background, probably a community member, and their responsibilities will encompass more than linguistics/archiving.

Another dimension of variation was the extent to which a language archive is integrated into an institution such as a university or state archive. Type 3 language archives are the most deeply embedded. While Types 1-2 are associated with an organization, they have more independence. For instance, Susan Kung indicated that even though Type 2 AILLA is part of the University of Texas Library and on the UT Library server, it operates independently from the rest of their digital repositories and collections. Type 4 language archives may or may not be associated with an institution; even when they are, such as the Myaamia Center, which has a relationship with Miami University (Ohio), they maintain a fair amount of independence.

3.1. Type 1: Large Language Archives with Global Collections

The main examples of Type 1, ELAR and DoBeS, are ambitious endeavors to develop global collections of materials on endangered languages. They are or were both funded by private sources. DoBeS recently lost its funding (which had come from the Volkswagen Foundation) and is now being reconstituted by Felix Rau's group at the University of Cologne. Since it is in flux, we will mainly describe ELAR. Mandana Seyfeddinipur, head of ELAR, explained that ELAR is closely connected to a funding source that provides grants for language documentation. ELAR receives deposits from all grantees since the deposits are a condition of their funding. Seyfeddinipur identified the following user groups for ELAR:

- Depositors
- Linguists interested in conducting research
- Language communities
- Researchers from fields like history, anthropology, political science, etc.
- Artists
- The general public

ELAR funds are largely earmarked for documentation, and this was also the case for DoBeS. Therefore, it might make sense for Type 1 language archives to partner with other organizations to prepare materials for language revitalization in cases where that would be useful. Speaking as an archive manager, Seyfeddinipur pointed out that it would be useful to have a list of what kinds of linguistic materials were necessary to collect for revitalization purposes. Such a list would help her train linguists who would like to contribute to revitalization efforts.

A challenge that pertains especially to Type 1s is that language communities may not speak the language of the archive's interface. For instance, Seyfeddinipur pointed out that ELAR is only available in English, but the archive includes materials of language communities whose members may not speak English.

3.2. Type 2: Large Language Archives with Regional Collections

Type 2 language archives are similar to Type 1, but with a regional focus to their collections. There are quite a few Type 2s, and they may be funded privately, by government grants, or by public sector institutions such as universities. Like Type 1s, their primary mission and funding tend to be focused on language documentation, so it might make sense for them to partner with other organizations that could develop materials for language revitalization.

3.3. Type 3: Language Archives Embedded in Larger Digital Repositories

Although the majority of language archives are associated with an institution to some degree, Type 3s are fully subservient to a parent repository that belongs to an institution such as a university, historical society, state archive, or national archive. Their high level of embeddedness can lead to challenges in customizing Type 3 language archives for the characteristics of linguistic materials or the needs of users. Andrea Berez-Kroeker said that the user interface of Kaipuleohone was determined by the University of Hawai'i library system: "As for the front end, I have no real control over what kind of information gets displayed, or how things can be searched. It's really geared towards traditional library publications, not media."

The parent repositories for these language archives often have a regional focus, but not always. Type 3s range from well-established language archives to small, emergent collections. As described in Section 4, their level of collaboration with language communities can vary.

3.4. Type 4: Single Language Community Archives

The language archives created by and for specific language communities display significant differences from Types 1-3. The primary mission of Type 4 archives is to serve their community in whatever ways are relevant for that community. Collections typically focus on a single language, or on a few languages if the community encompasses more than one. The activities of Type 4s commonly include language revitalization. Furthermore, Type 4s are often not *just* language archives; they may combine linguistic collections with cultural and historical materials to provide a more holistic set of resources for the community. The archives may be part of an information center that includes functions such as a library, a museum, and classes. These venues are typically located on tribal lands.

Daryl Baldwin noted that creating single language community archives was important for tribes because "it allows for the development of those archives in such a way that larger institutions either don't have the time, staff or money to develop for community use. For instance, all of our language documents were/are unpublished, untranscribed, and in some cases untranslated. With our own digital copies we are able to organize based on our need, and create tools that process them in such a way that we are able to gain access to the information without having to work through another institution's protocols and procedures." At the same time, he considered it likely that such language archives will be affiliated with a larger tribal archive, or another institution such as a university.

Type 4 language archives often seek to bring together all the materials on their language that they can locate around the world. Unlike Types 1-3, therefore, a significant part of their contents may be copies of collections from other archives. An example from the Miami-Illinois Digital Archive is illustrative: in 1999, researcher Michael McCafferty discovered a Jesuit manuscript by Pierre-Francois Pinet about the Miami-Illinois language, dating from about 1700, in the Archives De La Compagnie De Jesus in Quebec, Canada. After a process of negotiation, the Miami Tribe of Oklahoma obtained a digital copy of the manuscript in 2003.

Three Type 4 language archives were represented at the workshop, and each was different from the others in significant ways, displaying the range of variation that can be found across Type 4s.

- **Dinjii Zhuh K'yaa** (Gwich'in Language Archive and Language Revitalization Center) emerged from discussions at a Gwich'in gathering in 2010, so it is relatively new. It is located in Fort Yukon, Alaska, a town primarily inhabited by Gwich'in people. Its collections focus on language materials, but it also functions more broadly as a community center. Because the Internet is problematic in Fort Yukon, visiting the physical site of this language archive is currently the main way for community members to access its materials. Dinjii Zhuh K'yaa does have a public Facebook page at <https://www.facebook.com/dinjiizhuhkyaa/>.
- For the Miami Tribe of Oklahoma, language and cultural preservation and revitalization functions are spread across multiple, interconnected entities.
 - The **Myaamia Center** at Miami University in Ohio advances the Miami Tribe of Oklahoma's language and cultural revitalization efforts. Created in 2001, the center offers a wide range of educational experiences and programs for students who are members of the Miami Tribe. One of the goals of the Myaamia Center is to collect high-resolution copies of all Miami-Illinois language documents.
 - The **Miami-Illinois Digital Archive** (MIDA) is an ongoing research project that is now managed by the Myaamia Center (<http://www.ilaatawaakani.org>). MIDA was founded in 1999 to transcribe, translate and make accessible manuscripts written by Jesuit missionaries about the Miami-Illinois language. It is intended as a resource for researchers.
 - The online **Myaamia Dictionary** is intended as a resource for language learners (<https://myaamiadictionary.org/dictionary2015/>). It is also managed by the Myaamia Center. The Myaamia Dictionary is linked to MIDA, so materials are shared across the two sites.
 - Finally, there is also a more **general tribal archive**, which collects and shares a range of historical and cultural materials pertaining to the Miami Tribe. In summer 2016, the Tribe started to make materials from this archive available online, using the Mukurtu platform.

Daryl Baldwin, director of the Myaamia Center, noted that because MIDA is a language archive that focuses specifically on in-depth analyses of historical documents, the documents it makes available are not in themselves particularly useful for community members who want to engage in language learning, unless they already have some prior training. The Myaamia Dictionary, by contrast, is intended as a tool for language learners. In addition, language and cultural revitalization for the Miami Tribe is advanced by the many educational activities of the Myaamia Center.

- **FirstVoices** was launched in 2003 as an online language archive for the First Nations of British Columbia (<http://www.firstvoices.com>). So it is different from the other two examples in that it caters to multiple language communities. It includes 34 languages from British Columbia (61 dialects), and has also added a few languages from the rest of Canada and from California. FirstVoices has an especially strong pedagogical focus in its online offerings. It provides not only an alphabet, dictionary, and grammar for each language, but also extensive educational materials and games for both children and adult learners. While its offices are physically located in Victoria, British Columbia, FirstVoices is primarily an online entity.

In the U.S., the government does not provide ongoing financial support for Type 4 language archives. Members of language communities articulated nuanced positions regarding the pros and cons of state funding. On the one hand, they noted the U.S. government's failure to

recognize the importance of funding archives. FirstVoices, based in Canada, formed a contrast since it does receive government funding. On the other hand, members of language communities expressed a preference for being self-sufficient. As Daryl Baldwin put it, “you have to feed yourself.” They were wary of accepting money from outside sources due to the strings attached. Yet they also recognized that some tribes had more financial resources than other tribes, and that those with few resources were in a difficult situation.

For Type 4s associated with small language communities that have limited resources, it can make sense to partner with a Type 2 language archive. The Type 2 can provide secure and stable long-term storage of materials, while the Type 4 can make materials accessible to local community members. Such a partnership was developed between Dinjii Zhuh K’yaa (Gwich’in Language Center and Archive) and ANLA. Sometimes this is referred to as a “hub and spokes” model. Workshop participants seemed to think this might be a future growth area.

Type 4 language archives may also be more likely than other types to use cultural heritage platforms such as Mukurtu. Mukurtu is gaining in popularity due to its sensitivity toward cultural protocols and access restrictions, its free availability, and its flexibility in terms of hub-spoke operations.

3.5. Type 5: Elder Archives

Workshop participants also identified a fifth type of linguistic repository that is neither an online repository nor a repository in a building. Instead, it is a person who holds knowledge – a living person. Such people are referred to as elders. They are usually senior in age but not always. Their knowledge may include linguistic, cultural, historical, and genealogical dimensions. It may include knowledge about who in the community is the keeper of different kinds of information, and how information is transmitted across generations. This is living knowledge; it may not be written down. While Type 5 is technically not a formal archive, the concept can provide useful lessons and implications for the UCD of language archives.

4. Stakeholder Groups and Their Perspectives

Another key outcome from the workshop was mapping the diverse perspectives of different stakeholder groups concerning language archives. These groups included users as well as other stakeholders. For purposes of the workshop, we identified five main stakeholder groups, of whom the first two were user groups:

1. Language communities
2. Linguists
3. Archivists
4. User-centered design practitioners
5. Representatives of funding agencies

All workshop participants were able to represent at least one of these groups; many participants belonged to several groups.

In order to map the perspectives of each group, we took the following steps:

- Prior to the workshop, we divided the participants into the five groups listed above. We gave members of each group a set of questions about their perspective as a representative of that group. We asked them to prepare a short presentation for the workshop answering those questions.
- At the workshop, participants delivered presentations filled with rich insights.
- The entire workshop was video recorded.
- A detailed transcript summary of the workshop was prepared based on the video recordings.
- We coded the transcript using the qualitative analysis software Dedoose to capture all statements by participants that revealed the perspectives of each stakeholder group. In this way we not only examined participants' presentations but also considered relevant remarks made at other times during the workshop.
- By grouping the statements coded for each stakeholder group, we were able to identify common themes and patterns in the perspectives of each group.

The results of our analysis are summarized below.

4.1. Language Communities

The workshop participants who were asked to represent the perspective of language communities in their presentations were Loriene Roy, Daryl Baldwin, Edward Alexander, Wesley Leonard, and Michael Shepard. The first four are members of tribal communities; Michael is a non-indigenous anthropologist who has worked extensively with several tribal communities on archiving issues. In addition, Jennifer O'Neal contributed insights as a member of a tribal community. In planning this workshop, the organizers decided to limit language communities to those of Native North America in order to create more coherence in the discussions. The discussion in this section overlaps somewhat with Section 3.4, above, where Type 4 language archives were described.

Both Daryl Baldwin and Edward Alexander had experience developing Type 4 single language community archives. Baldwin played an instrumental role at the Myaamia Center, which was founded in 2001 at Miami University in Ohio. Alexander guided the development of Dinjii Zhuh K'yaa (Gwich'in Language Archive and Language Revitalization Center) in Fort Yukon, Alaska. The idea for the language archive emerged from a Gwich'in gathering in 2010.

Language community members made it clear that language was fundamental to the identity of tribal groups and their members. Wesley Leonard shared this quote from L. Frank:

“Language is the basket that holds all of our culture... in order to understand why this oak tree sitting on a hilltop is *so critical* to my afterlife, the language is the only thing that explains that and carries that and is that... I need the language to understand. I don't want to take a wrong road when I get to the edge of the land of the dead... so language is pretty much everything” (quoted in Leonard, 2016).

The history of U.S. colonial policies and practices (including the boarding schools) created significant hurdles for language use in the majority of Native American communities. The after-effects of colonialism undergirded many workshop discussions. For instance, Edward Alexander commented, “right now it seems like archives, and this kind of information, it still feels like the BIA days. Because we don't have self-determination over the information.” He compared the poor quality of services under BIA management with the improved quality of services after the Indian Self-Determination and Education Assistance Act, when tribes obtained more control. Loriene Roy said, “we forgive you for taking our words. We forgive you for recording them sometimes without our permission.”

Michael Shepard pointed out that archives are sites of struggle for power, control, access, and ownership. Language communities wish to exercise sovereignty over their cultural and linguistic heritage. Some forms of knowledge may not be appropriate to share. Some kinds of information can play an important role in supporting a community's legal rights, including land claims.

At an ideological level, Wesley Leonard argued that there can be mismatches between the values and categories that guide archiving practices, and those that guide language reclamation efforts. “A key for language reclamation is appropriate integration of cultural beliefs, including community notions of language – so not a top-down ‘this is what language is’ as defined in this archive, but from the bottom up, from the community.” A number of workshop participants suggested that language communities needed to step away from traditional Western linguistic and archiving frameworks in order to conceptualize Type 4 language archives in the ways that would be most useful to the communities.

4.1.1. Perspective on Language Archives of Types 1-3

The representatives of language communities spoke of encountering numerous challenges in their efforts to work with languages archives of Types 1-3. First of all, it is often very difficult to locate all the materials pertaining to their languages. These materials can be scattered across archives all over the world. Archives may use spellings for language names that users are not familiar with, or completely different names. Materials in large-scale repositories, like federal archives, are typically easiest to find.

Another major challenge is the difficulty of understanding and interpreting many of the materials found in language archives. As Edward Alexander said, “you need a key to be able to understand what you’re looking at. Not only is it in another language but there’s another language of linguistics on top of it.” Susan Kung pointed out that for older documents, the linguistics “language” of 200 years ago could be indecipherable to today’s linguists as well. So there is a need to translate academic and archaic language into accessible terms. Daryl Baldwin talked about the work he has done to transform source documents into a form that members of his community can easily use: “that’s a huge process... There’s no easy way to do that.”

With respect to the legacies of colonialism, sometimes Type 1-3 language archives included materials taken in earlier times without the permission or understanding of tribal members, making information publicly available that the language community would prefer not to share openly. Sometimes access to a Type 1-3 language archive was limited for members of a tribal community by the expense of traveling (to a non-online archive), copying materials, and so forth.

4.1.2. Perspective on Type 4 Language Archives

As described in Section 3, the primary mission of Type 4 language archives is to serve their communities. Michael Shepard noted that Type 4s can be thought of as a reservoir that a language community draws on for a variety of purposes, from land claims to education. Each community shapes its language archive according to local needs and interests. For instance, Edward Alexander described Dinjii Zhuh K’yaa (Gwich’in Language Archive and Language Revitalization Center) as a place to share, a source of innovation where community members make new things from the past. It emerged from Gwich’in concerns that their language, and therefore their very existence, were under threat. Alexander worked with ANLA to bring copies of materials to Fort Yukon. Many of these materials had not previously been digitized or placed online, so they had not been accessible locally. Gwich’in community members benefit from Dinjii Zhuh K’yaa by gaining access to previously unknown materials. Alexander himself discovered a recording of his grandfather that he hadn’t known existed. One of the innovative things Dinjii Zhuh K’yaa has done is to record knowledge and then imprint it on the land using augmented reality. For instance, the building that houses the archive is scannable; visitors can pull up information about it on their phone. This helps share information between elders and the younger generation.

Language Revitalization/Reclamation. Preparing language learning materials from the linguistic materials stored in language archives was accorded great importance by workshop participants. As a result of American colonial policies, including the infamous boarding schools, many members of Native American language communities no longer speak their tribal languages in the home. Yet tribal communities understand their languages to be closely linked to cultural knowledge systems and values. Community members often have great interest in language learning activities that can promote “revitalization” or increased use of such languages. To highlight tribal sovereignty, Wesley Leonard has adopted the term “language reclamation” to describe the right of community members to claim, learn and speak their language, and to define what success means for their language learning efforts (Leonard, 2011: 141). During workshop discussion, Michael Shepard argued that “there’s a real role for a pedagogical specialist to be involved” in the development of language archives, since the skill set needed to prepare educational materials is not widely found among language community

members, linguists, or archivists. FirstVoices is a great example of how language learning materials can be developed from language archives.

Create Engagement. Workshop participants pointed out that most current language archives are not very engaging for most language community members, outside of programs designed specifically for language learning. Daryl Baldwin shared his experience of education trips or NAGPRA-related trips with both elders and young people. He contrasted their engagement with physical objects to their lack of connection with language documents:

“In most cases, when community members step into the archive, they’re looking for objects... in every case it’s been a very powerful moment for them to engage with or interact with physical objects that are directly not only tied to their culture, but in many cases tied to their genealogies. So these are objects that came from their great-grandparents or whatever it may be. What makes the experience rich really hinges on what they know about the object. In some cases they don’t know a lot about the object. So if there’s somebody there from the community that can serve as an historian or cultural expert they can contextualize their experience. And when that is happening, their interaction with the object is lengthened.”

Baldwin has found that while the average community member may find looking at original language documents cool and interesting, such documents often don’t have the same impact as a cultural object (like a basket), and the person may quickly lose interest. However, students who are more engaged in the language learning process find materials from language archives highly meaningful. For instance, the Myaamia Center sent three tribal students to work with the Smithsonian’s Recovering Voices program, doing research in preparation for the 2015 National Breath of Life Archival Institute for Indigenous Languages (<http://nmnh.typepad.com/recoveringvoices/2015/01/>). These students had been taking classes on Myaamia language and culture. Seeing Myaamia language documents helped them contextualize their experiences in the larger framework of indigenous language revitalization.

Workshop participants brainstormed ideas for how to make language archives more engaging. Suggestions included the use of audio and video recordings, and making abstract information more tangible. Lorie Roy gave an example of working with a group of 8-10 year olds to narrate a virtual tour of an online exhibit, using their language. One boy chose to talk about a piece of pottery from Maria Martinez. He started his narration by saying, “my auntie made this plate. And first she went to the river and she gathered the clay and she sifted everything,” and continued from there. Having children tell stories that would be viewed by other children seemed like a great engagement technique.

Let Users Annotate Language Archives. Language community representatives expressed a wish for users to be able to annotate language archives. For instance, Edward Alexander said there were recordings in Gwich’in at ANLA that have only minimal descriptions. There is no way for him to add more detailed information that might be helpful for other users. Likewise, Daryl Baldwin expressed a wish to work with archives to:

- Update their descriptions and keywords for Miami-Illinois (the broader language of which Myaamia is a dialect)
- Share the transcription and analysis work done by linguists and tribal researchers, and connect this work with the archival documents

- Create a list of ongoing research associated with certain collections so that others who visit the archives become aware of who is doing what

Workshop participants explored ideas for enabling community members to annotate materials themselves, and even to debate interpretations with each other online.

Let Users Be Depositors. Edward Alexander suggested that it would be helpful if users could be depositors – in other words, the language archive could let them upload materials themselves. He pointed out that recording devices are now ubiquitous, unlike earlier days when they were rare and expensive. Alex Wadsworth indicated that FirstVoices already has a self-deposit tool that allows users to upload materials. Alexander also pointed out that users needed to be thoughtful about possible future uses of recordings they upload. “You don’t know how people are going to use information.” When his grandfather was recorded, he probably just regarded it as a novelty, without realizing that years later his grandson would find this information on a computer.

Constraints: Literacy, Costs, Technical Support. Workshop participants identified a number of constraints that Type 4 language archives need to negotiate. Lorie Roy noted that literacy might be an issue; sometimes recordings may be more useful than written materials. Many workshop participants talked about the financial challenges of Type 4 archives. As noted in Section 3.4, the U.S. government does not currently provide funding for the language archives of tribal communities. Similarly, some communities have limited Internet access or face other constraints in their technological infrastructure.

4.2. Linguists

The workshop participants we asked to represent a linguist perspective were Andrea Berez-Kroeker and Justin Spence. In addition, participants Shobhana Chelliah, Mandana Seyfeddinipur, Susan Kung, Felix Rau, Daryl Baldwin, Wesley Leonard, and Gary Holton are also linguists, and at times their remarks were shaped by this perspective. One of the goals of the workshop organizers was to treat linguists the same as any other stakeholder group. Because many online language archives have been designed by linguists, there often seems to be a default assumption that they must be working well for linguists. We did not want to make this assumption. Rather, we wanted to ask linguists what their experience was like. And we recognized that linguists might have different experiences depending on whether they were interacting with a language archive as a depositor or as a researcher.

Some aspects of a linguist perspective on language archives were described in the first half of Section 2, which summarized Holton’s presentation at the start of the workshop. The current section focuses on additional contributions from Berez-Kroeker’s and Spence’s presentations, and comments made by linguists in later discussions.

4.2.1. Linguists as Researchers

One significant finding was that the linguists at the workshop did not seem to be using language archives much as researchers. The most common way that linguists engaged with language archives was to deposit their materials. Aside from Daryl Baldwin’s and Edward Alexander’s

engagement with language archives to develop their own community's archive and to bring more knowledge about their language to their community, Justin Spence was the only workshop participant who used language archives extensively for linguistic research. While many authors have noted the great potential of language archives to contribute to linguistic research (cf. Seifart et al., 2012), we suspect that low usage of language archives by linguists may be a generalizable finding; it would be interesting to conduct more systematic research on this question.

Justin Spence acknowledged that he was willing to “slog through sometimes hours of recordings just because I kind of enjoy it!” His careful approach to navigating materials that could be “very haphazard” almost sounded like a form of detective work. The old field notes, recordings, and transcripts he examined were useful in recovering variation, examining language change, and understanding people's lived experiences of the languages they knew while processes of language shift were in motion.

A second, related finding was that linguists as a user group had a great deal of difficulty using language archives effectively for research purposes. In some ways, this was surprising, since most online language archives were created by linguists. However, Shobhana Chelliah noted that “even as an academic I feel like I'm shut out a lot.” One problem she identified was that linguists may not be ready to share their data by placing them in an archive because of concerns about intellectual property and preserving a publishing advantage. Mandana Seyfeddinipur pointed out that there isn't much of a reward for linguists to deposit their data.

Even when linguistic data have been deposited, they may be hard to use. Workshop participants said that the major difficulty in using existing deposits for linguistic research was that they often lack sufficient annotations. Annotations are needed both to find relevant materials during the search process, and to interpret the materials once they are found. Such annotations could range from a text paragraph summarizing the contents, to specific metadata to facilitate language comparisons, to a gloss for every single morpheme in a text. Different linguists have different annotation needs, depending on their specific research topics. Voice recordings are most accessible when they are accompanied by transcripts, because transcripts can be searched in a way that recordings currently cannot be.

4.2.2. Linguists as Depositors

Linguists also encounter challenges in their role as depositors. One challenge is their lack of training in how to prepare deposits in a way that will be useful to other users. As Chelliah put it:

“In my education as a linguist I was really never taught how to fashion my data so that community members could access it and use it. And we're still really lacking in our education, as linguists, how to do that properly. So we can't expect archivists to know that for us. I'm called on as a linguist to do two things. One is to produce theoretically or descriptively exciting information that pushes forward the science of what we know about language and cognition. And secondly, for the people who invited me into their communities to come and help them with the creation of a history of their language, to work on that information with them... I don't have the training to know how to use my limited time to do both of those things.”

Another challenge is that language archives may not have a clear submission process for deposits. Berez-Kroeker noted that she had encountered opaque submission procedures, especially for metadata creation, and poor communication in the “black box” period between submission and ingestion.

A third issue is that developing a collection for a language archive is a lot of work, and linguists often get little credit for such efforts in academic merit reviews. This issue is being addressed by a group of linguists that includes some workshop participants; they are seeking to gain recognition for deposits as a form of publication.

4.3. Archivists

The workshop participants we asked to represent an archivist perspective were Mandana Seyfeddinipur, Felix Rau, Susan Kung, Jennifer O’Neal, and Alex Wadsworth; Table 4 provides more information about them. Andrea Berez-Kroeker (Kaipuleohone), Daryl Baldwin (Myaamia Center), Edward Alexander (Dinjii Zhuh K’yaa/Gwich’in Language Center and Archive), and Gary Holton (ANLA) also brought experience in managing archives to the table. One thing to keep in mind is that language archives are often managed by people who don’t, in fact, have formal training in archiving. Among workshop participants, only Jennifer O’Neal and Loriene Roy had degrees in library and information science.

In their presentations and subsequent discussions, the archivists painted a rich portrait of each of the archives they worked with and presented a nuanced view of the complexities of managing a language archive. It was useful to have five examples, because each one was different in significant ways. In this section we describe patterns that emerged across the examples. Most of these patterns applied to some but not all of the five language archives under discussion.

Table 4. Workshop Participants Asked to Represent Archivist Perspective

Archivist	Language Archive	Language Archive Type
Mandana Seyfeddinipur	ELAR	Type 1
Felix Rau	Language Archive Cologne	Type 1
Susan Kung	AILLA	Type 2
Jennifer O’Neal	University of Oregon	Type 3
Alex Wadsworth	FirstVoices	Type 4

4.3.1. Feeling Disconnected from Users (Other Than Depositors)

Seyfeddinipur, Kung and Rau all talked about feeling disconnected from their users, except for some of their depositors. These three archivists manage Type 1 and Type 2 language archives. Kung said:

“As someone who runs an archive, the biggest issue I have, since it’s a digital archive, is knowing who’s using the archive. When somebody contacts me by email or phone... then I know who they are and what they’re trying to do, but otherwise I can track the downloads, I know people are logging in every day and downloading materials, but I have no idea who they are or what they’re using these materials for, what their agendas

are, what they're researching. So I feel like I'm just totally disconnected from most of my users, I have no insight into their needs or their wants."

While ELAR collects more information about users when they create an account, Seyfeddinipur still expressed frustration about her limited knowledge of users. "We are trying our best with the knowledge that we have, and we need more information." She also expressed concern that ELAR is underutilized. Specifically, she noted that the restricted collections, which mainly belong to Native American and Australian Aboriginal communities, are not getting any use at all. "User statistics are actually zero, which is heartbreaking." For Seyfeddinipur, not only do language archives need to become more engaging for users, but they also need to do a better job of alerting people to their existence.

At the same time, Seyfeddinipur pointed out that sometimes there is no longer any community associated with a language whose materials have been deposited in ELAR. The few remaining speakers may be dispersed, or there may be none left.

The one set of users that did regularly communicate with these archivists was the depositors. Rau noted that at the Language Archive Cologne, the majority of users are, in fact, depositors: "it's bordering to the ridiculous sometimes how the whole thing is focused on the producer side." ELAR also works closely with depositors; ELAR gives grants for language documentation that require the recipients to deposit their data with ELAR. Since ELAR provides training to the grant recipients in data management, there is a fair amount of contact with them.

It was notable that Seyfeddinipur, Kung and Rau all manage archives of Types 1 and 2, which have a primary mission of language documentation and preservation rather than community engagement. For instance, Kung said that while AILLA has an open collection policy, it does not actively seek out collections to acquire. "We don't hunt it down, it comes to us."

4.3.2. Collaboration with Language Communities

By contrast, Alex Wadsworth and Jennifer O'Neal collaborate closely with language communities. FirstVoices, where Wadsworth works, was specifically created to support the Indigenous languages of British Columbia. It is a conglomeration of Type 4 language archives. The University Archives of the University of Oregon, where O'Neal works, has a much broader mission; language archives constitute only a small part of the holdings. We labeled language archives that are embedded within a parent repository like this Type 3. O'Neal is currently working on two projects related to endangered languages in the Northwest.

It seemed like O'Neal's collaboration with language communities was, at least to some extent, a personal choice. She spoke of how important it was to her to be an advocate for tribal communities; she herself is a member of one. She described the ongoing challenge of persuading her Dean to dedicate resources to language archives rather than other urgent university priorities.

O'Neal's case shows that an archivist's collaboration with language communities is not only shaped by institutional structures (such as our typology), but that it can also be shaped by the agency and personal commitment of the archivist. There may be Type 3 archivists whose experience more closely resembles that of the Type 1 and 2 archivists described above.

4.3.3. Scarce Resources

A common theme in the comments of archivists was the challenge of finding adequate resources to maintain and improve their archives. Most language archives have few employees and limited financial resources. They are already doing a lot with a little. Several archivists expressed concern that their funding sources could dry up or become totally inadequate – a terrible scenario for an archive, whose most fundamental mission is after all to preserve collections for hundreds and even thousands of years. They noted that a language archive’s need for guaranteed, long-term sustainability is a poor fit with the dominant funding model of short-term grants for specific projects. A related concern was the need to compete with other institutional priorities; O’Neal noted that many of her colleagues had to “fight” for support because language archives are often not a high priority at their institutions. In their concluding remarks at the workshop, archivists made comments such as:

- “Things are changing faster than we can keep up with”
- “I know the problems users have... but at the same time, with very few resources, and feeling very over-extended, this wasn’t really something that I wanted to address”
- “We’re trying to do the best job we can with the small resources and staffing that we have”
- “I’ll echo what has been said about the strain on all of us that we’re dealing with”

Any UCD efforts will need to take this context into account, and strive to ease the workloads of those who are involved with language archives, rather than making things more burdensome for them.

4.3.4. Problems Encountered by Users of Language Archives

In their presentations, the archivists provided a rich list of problems that might be encountered by users of language archives. The most common items they mentioned were:

- A lack of contextual information at the deposit level, or metadata
- Incomplete materials – missing annotation, missing translations
- Inadequate search/browse functions
- Problems with the interface/information display
- Users may be frustrated when they don’t have access to data; it may be hard for the archivist to get hold of a collection owner to request access for a user
- Technology issues – outdated, broken scripts, Flash/Java problems, etc.
- Interface language(s) may not include a language spoken by would-be users

4.4. UCD Practitioners

The UCD practitioners invited to the workshop were Crysta Metcalf, a design anthropologist, and Santosh Basapur, a designer. Both worked at the Motorola Mobility Applied Research Center until 2013, when the center was sold and disbanded. In addition, Christina Wasson, workshop co-organizer, and Heather Roth, research assistant, brought a UCD perspective to the conversations. The UCD group was different from the other stakeholder groups in that it was the only one whose members did *not* have long experience with language archives. And it was the only stakeholder group that *did* have deep experience with UCD.

Many aspects of a UCD perspective on language archives were described in Section 2 under User Centered Design, which summarized Wasson's presentation at the start of the workshop. The current section focuses on additional contributions from Metcalf and Basapur, articulated during their presentations and in later discussions.

Throughout the workshop, all UCD practitioners emphasized the bottom-up nature of the UCD approach. As Metcalf stated on one of her presentation slides:

“I think that the focus on the users in User-Centered Design is the key. In UCD we don't assume needs or goals or workflows, we don't design what WE want or would use – we talk to the people who will be using the technology, and engage in a dialectical design process, so the result should meet the needs of all of the users in an intuitive way for each group.”

A consequence of this bottom-up approach was that none of the UCD practitioners was ready to say much about specific design solutions for language archives at the workshop. Such design solutions will only emerge after in-depth user research has been conducted.

4.4.1. Explaining UCD to Other Stakeholder Groups

During their presentations, Metcalf and Basapur sought to explain the UCD process to an audience unfamiliar with this approach. Both used visual aids with circular arrows, emphasizing the cyclical and iterative nature of the UCD process. The steps they listed can be summarized as:

- Develop initial understanding of problem
- Conduct research with users in the context of use
- Identify research findings
- Identify design concepts based on research findings
- Create design prototypes
- Circle back with the user groups for their assessment of the design
- Iterate on the design with continued input from user groups

Metcalf and especially Basapur provided examples of previous projects to illustrate the UCD process. Basapur reviewed projects on:

- Social TV, which allows geographically distant friends and family to watch TV together and socialize around it
- A location-based mobile app that allowed people to access videos about particular locations when they approached those spots; the videos were recorded by the parents and grandparents of the users, so this app promoted intergenerational communication as well as knowledge of the urban landscape
- A hospital's operations theater, where researchers discovered that nurses were the group most lacking in support
- A city's bike sharing service, specifically how it could be made more attractive to residents of marginalized communities

The goal of explaining UCD to the other stakeholder groups was to generate shared knowledge and frameworks that all workshop participants could build on during the workshop. Since UCD practitioners regularly engage in interdisciplinary collaboration, they were familiar with both the joys and the challenges involved.

4.4.2. Mapping the Terrain of Language Archives

For the UCD practitioners, their most urgent task at the workshop was to map out the key features of the (to them) new and unknown world of language archives, in ways that would help them plan further research and design activities. Two dimensions emerged as key features of this map:

- Identifying the stakeholder groups and their perspectives
- Identifying different types of language archives.

The first dimension is addressed in the present section of this report; the second dimension was addressed in the previous section.

4.4.3. Moving Forward

While the UCD practitioners were not prepared to identify design solutions during the workshop, they shared ideas concerning research methods and other elements of a UCD approach for language archives, as those topics became relevant during various discussions. For instance, when they learned that language archives don't always know much about their users, they brainstormed methods for identifying current language archive users and collecting information about them. Santosh Basapur suggested that when we start to conduct research, we can place intercepts on a language archive website, i.e. pop-up questions that users can respond to. Crysta Metcalf pointed out that information about language archive users could be aggregated across multiple language archives.

UCD practitioners noted the potential value of partnerships between language archives and other organizations as a way of accomplishing goals such as language revitalization activities. They recognized the need to address funding challenges. They let workshop participants know how useful metaphors such as "opening the door" (described in Section 5) were – such analogies can help inspire design solutions.

Finally, the UCD practitioners stressed the importance of collaboration across stakeholder groups. In a humorous way, Basapur said he had seen linguists feeling sorry for archivists, archivists feeling sorry for language community members, and language community members feeling sorry for the technical people. He pointed out that this kind of empathy is central to the UCD process.

"Everybody has to have that empathy... Usually design gets a reputation for new ideas and fresh ideas, but actually what we usually do is bring people on the same page and then you do something with it. By the time they all come onto the same page they bring their own ideas and all the ideas start compiling and [then we can] say what will really

make sense. So developing empathy and then having the empathy distributed among all the stakeholders is very important.”

4.5. Funding Agency Representatives

The workshop participant representing U.S. funding agencies was Shobhana Chelliah, who recently completed two terms as program officer for the NSF Documenting Endangered Languages program (DEL), 2012-2015. It was also useful to hear Mandana Seyfeddinipur’s perspective as a funder/archive manager outside the U.S. Comments from other workshop participants illuminated the ways in which U.S. federal grants are embedded in a larger system of institutions, archives, language communities, and other groups whose needs and activities are intertwined in complex ways.

4.5.1. DEL Data Management Plan

Chelliah explained that one mission of DEL is to create resources for science and humanitarian purposes. This mission can only be fulfilled if grant recipients make their documentation results available to the public, which usually means depositing them in a language archive. In 2012, when she became DEL program officer, Chelliah started to realize that many grant recipients were not in fact making their materials available. She talked to language archive managers, and they discovered that “we really didn’t know what was being archived, how much was being archived, and what quality those deposits were.”

As a result, the DEL program added a requirement to the existing NSF Data Management Plan (DMP). The DEL-specific DMP required applicants for DEL standard research and dissertation development research grants to include a plan for placing their materials in an archive at the conclusion of the funding period. They had to budget time and money to accomplish those tasks, and they had to get a letter from an archivist approving this plan.

Here is the relevant text from the DEL Program Solicitation (NSF 15-567):

The DMP should provide evidence that the applicant has contacted a trusted repository to arrange for long-term archiving of documentation generated by the DEL project. The language archive selected by a DEL project must have a long-term institutional commitment to data preservation and access. While the DEL Program does not sponsor or have an official arrangement with any language archive, these services are provided by DELAMAN member archives (<http://www.delaman.org>) and by institutions holding the Data Seal of Approval (<http://www.datasealofapproval.org/en/>). Regular data backup should be an integral part of the DMP, but this is not to be equated with archiving in a trusted repository. Backing up data on hard drives, servers, optical media, and cloud based services does not constitute archiving.

The DMP should include a time-line for completion of archiving activities. It is expected that archiving should be completed prior to the submission of the final project report.

Applicants should include a letter of support from the archive indicating their willingness to archive project materials and outlining any specific arrangements which have been made. This statement must be uploaded under “Other Supplementary Documents”.

Language documentation is of little value if it cannot be accessed. To that extent the DEL Program expects that the vast majority of data generated by the DEL project will be publically accessible with minimal restrictions for non-commercial, educational purposes. (Restrictions on commercial use are acceptable.) The DMP should indicate how archived materials will be accessible to the public. Any restrictions to be placed on access should be clearly indicated. If the applicant expects access to some materials to be restricted to certain user groups, the DMP should indicate the criteria delineating such user groups and provide an estimate of the percentage of materials which will be so restricted. If time limits are to be placed on access to materials, the DMP should indicate the period of time after which access restrictions will be removed.

While the new DMP was an important step forward in ensuring that language materials would be placed in an archive, it also had unintended consequences. Some archivists are now feeling overwhelmed by requests to take language materials, while the funding they receive from their institutions has not increased. Chelliah noted that DEL allows grant applicants to add a line item to their budget to offset costs associated with archiving, typically 8% of direct costs, as suggested by the DELAMAN group. Such funds could help alleviate stresses to archives in accepting new materials. Susan Kung said that she had indeed started to ask for 8% of direct costs from NSF grant recipients who want to put their collections in AILLA. However, she has yet to receive any of the funds that have supposedly been budgeted. The grant applicants are not even letting her know whether they have received the grants. Chelliah commented that protocol and communication between archivists and grant applicants needed further development.

4.5.2. Limitations of U.S. Funding Model

Workshop participants identified several limitations of the U.S. approach to funding the development of language archives. One issue was that funding takes the form of “soft money”, meaning short-term grants, usually for three years. Gary Holton pointed out that “within the U.S., at least, this is the way we fund science... this *is* the model.” Yet the concept of preservation is central to the notion of a language archive, and for preservation, the temporal horizon is not three years but hundreds or thousands of years. Mandana Seyfeddinipur argued that “the problem [with soft money] is that’s not sustainable, this is not something that will save the archive. This is something that gives you money for a certain amount of time.” She noted that one of the major language archives with a global collection, DoBeS, was funded on soft money, and when its money was pulled the archive was no longer sustainable.

A second and related point is that the U.S. lacks a central archive to deposit materials funded by the NSF. Seyfeddinipur suggested the Smithsonian. Such an archive would avoid the problems of soft money and lack of sustainability, especially for small communities with limited funds. Jennifer O’Neal also stated that many tribal communities don’t have the funds to create their own language archive. She said all of her archivist colleagues in the Northwest are dealing with the same issue: language communities are coming to them for assistance, but the additional funding needed to help them is not the highest priority for their universities.

However, Chelliah pointed out that the U.S. Congress may not be interested in funding a national language archive. Their current priorities are national security, health, and economics. She also noted that some depositors may want to preserve the freedom to choose where they deposit materials, and that a centralized archive would constrain the format of deposits to a single set of guidelines.

A third limitation to the U.S. funding approach is that DEL draws a line between language documentation and access/use/revitalization. DEL only funds documentation, not the other aspects. However, Chelliah pointed out that it was possible to partner with other funding sources for education/revitalization activities. She mentioned that the director of the NSF Tribal Colleges and Universities Program was very interested in language and linguistics, and in funding the education of tribal university students.

To receive DEL funding, Chelliah suggested the following approaches. She noted that she was only speaking about NSF funding priorities so her advice might sound “cold.” At a general level, she said that you could say you are developing ideas on how language works which would contribute to our understanding of all languages, including languages of interest to national security. More specific suggestions for finding funding for the creation of language archives were:

- Apply for funding to create a computationally sophisticated, annotated archive that facilitates cross-language comparisons. NSF programs in Linguistics and Documenting Endangered Languages (DEL) would be interested in funding that kind of archive.
- Apply for funding to develop an archive to help advance Natural Language Processing or Forced Alignment and Transcription. NSF Computer and Information Science and Engineering (CISE) programs would be interesting in funding that kind of archive.
- Apply for funding to the Social, Behavioral and Economic Science Directorate’s special program called RIDIR, Resource Implementations for Data Intensive Research. This program “seeks to develop user-friendly large-scale next-generation data resources and relevant analytic techniques to advance fundamental research” (NSF Website).

5. Access: Opening the Door

5.1. The Twin Missions of Preservation and Access

In the world of archiving, the two main missions of archives are framed as “preservation” and “access” (Hunter, 2003). According to workshop participants, the language archives we termed Types 1-3 are mainly funded to engage in preservation work, so that tends to be their primary focus, but they are also interested in promoting the use of their materials. For Type 4, preservation remains important, but there is likely to be a greater emphasis on access.

For preservation, the application of UCD is a somewhat abstract concept. Preservation is accorded a high value in linguistics and archiving regardless of the current needs of particular users – in this sense, preservation is not a *user-centered* activity, although it may be end up being extremely *useful* to many people. In the broadest sense, “users” who may benefit from preservation include potential users who have not yet been born, as well as non-users such as members of a language community who don’t themselves use the language archive, but value its existence. The application of UCD to facilitate access to language archives is more obvious.

We should also keep in mind that uses of a language archive may be emergent and unexpected. For instance, the Dinjii Zhuh K’yaa (Gwich’in Language Center and Archive) ended up getting National Institutes of Health funds because it became a community center, a place for people to have meetings about managing health issues.

5.2. Opening the Door

During the workshop, Edward Alexander initiated the idea of access as a “door” to language archives; the metaphor was quickly adopted and extended by the group. Alexander initially remarked:

“Academics are building these archives... and so you build it for people like yourself. So the door is an academic door, right? So other academics walk along and say, ‘oh! I know how to open this door. And it’s for me! And everything in there is for me!’ And for other people who are not academics, they look at these archives and they’re like looking at tools from some foreign thing... the door isn’t *made* for them.”

As other workshop participants built on this notion, it became clear that *none* of the workshop participants, not even the linguists, felt that the doors to language archives were easy to open. The UCD practitioners found the door metaphor very helpful as a way to think about designing for access. The design of a language archive could create doors in the first place, help users open a door and enter, and escort users inside the language archive to help them make sense of the offerings and find what they are looking for.

5.3. Escort

Alexander also suggested that it would be helpful for users to have an escort once they entered the language archive, to help them find and interpret items of interest. When he first visited

ANLA, he was guided by an archivist. Libraries have reference librarians. The Siri on iPhones can help people find their way around a city. During his presentation, Daryl Baldwin also expressed the view that while online search engines have increased access, nothing replaces the knowledge of archivists. This notion of an escort found resonance with workshop participants.

5.4. Collection Guides

The workshop participants with archiving expertise pointed out that the creation of finding aids such as collection guides is a normal part of the archiving process. Usually a curator writes a guide for each collection, including a description of the contents, a list of inventory, who has access rights, and how to navigate the collection. Even when a collection is not digital, the collection guide is usually available online; this is what allows people to find collections that have not been digitized. Jennifer O'Neal noted that collection guides may be placed in federal databases to make the materials more findable.

Online language archives are unusual in the world of archives in that collection guides are often missing. The reason is probably that many of these language archives were created by linguists, who may have lacked expertise in archiving practices. At this point, there is also a shortage of time and staff to prepare detailed finding aids for all collections. Susan Kung noted that she has a few guides on AILLA. But they are dependent on depositors writing the guides for their material.

Workshop participants discussed whether such collection guides would resolve some of the access challenges for language communities and linguists. They concluded that it would depend on how much information was included in a collection guide, and whether that information was targeted to the search needs and interpretation needs of particular users.

6. Organizations to Partner with for UCD Guidelines

The final phase of our planned research trajectory will be the dissemination of guidelines for UCD of language archives. We expect to start this final phase in three to five years. It may be regarded as a culture change process or a paradigm shift. We envision it as similar to the paradigm shift that occurred 20 years ago or so in the business world, when UCD went from being a novelty to becoming the normal, taken-for-granted practice.

We asked workshop participants what organizations we should partner with when we are ready to disseminate UCD guidelines. They gave us many valuable suggestions. The overarching insight was that we should partner with as many organizations as we could. This was a valuable “aha” because we had thought we should select just a few organizations. But workshop participants pointed out that the dissemination process would benefit from the broadest possible number of collaborators, and that contact with each organization would not be too onerous of a time commitment.

We also received names of organizations that could be venues for us to present on our work, as part of our dissemination activities.

Table 5 provides a list of organizations that workshop participants suggested. Some of these might be partners for the development of guidelines adopted by the organization; others might just be venues for us to disseminate our findings.

Table 5. Organizations That Could Be Dissemination Partners and/or Venues

Name	Website	Short Description
CELP – Community on Endangered Languages and their Preservation, Linguistic Society of America (LSA)	http://www.linguisticsociety.org/about/who-we-are/committees/endangered-languages-and-their-preservation-celp	Calls attention to the rapid loss of language diversity worldwide and encourages the documentation and study of endangered languages
DELAMAN – Digital Endangered Languages and Musics Archives Network	http://www.delaman.org	DELAMAN is an international network of archives of data on linguistic and cultural diversity, in particular on small languages and cultures under pressure.
ATALM – Association of Tribal Archives, Libraries, and Museums	http://www.atalm.org	The world's largest gathering of people dedicated to sustaining and advancing Indigenous cultures; a forum for issues important to archives, libraries, museums, language programs, and historical preservation offices.
UNESCO – United Nations Educational, Scientific and Cultural Organization, Endangered Languages Programme	http://www.unesco.org/new/en/culture/themes/endangered-languages/	Its aim is to support communities, experts and governments by producing, coordinating and disseminating tools for monitoring, advocacy, and assessment of trends in linguistic diversity, as well as services such as policy advice, technical expertise and training, good practices and a platform for exchange and transfer of skills.
ICLDC – International Conference on Language Documentation and Conservation	http://www.icldc-hawaii.org	A biennial forum that brings together linguists, students, and community activists to share resources and research and discuss issues of importance in documenting and revitalizing the world's endangered languages.
Indigenous Language Institute (ILI)	http://www.ilinative.org	Provides vital language related services to Native communities so that their individual identities, traditional wisdom and values are passed on to future generations in their original languages.

Society for American Archivists	http://www2.archivists.org	North America's oldest and largest national professional association dedicated to the needs and interests of archives and archivists.
Association for Recorded Sound Collections	http://www.arsc-audio.org/index.php	Dedicated to the preservation and study of sound recordings—in all genres of music and speech, in all formats, and from all periods.
American Library Association	http://www.ala.org	The mission of ALA is “to provide leadership for the development, promotion and improvement of library and information services and the profession of librarianship in order to enhance learning and ensure access to information for all.”
CoLang – The Institute on Collaborative Language Research (LSA)	http://www.linguisticsociety.org/content/colan-g-institute-collaborative-research	A biennial summer school that offers an opportunity for practicing linguists, students, and Indigenous language community members to develop and refine skills and approaches to language documentation and revitalization.
LSA Institute	http://www.linguisticsociety.org/content/future-linguistic-institutes	A biennial summer school that provides advanced training in linguistics not readily available through the regular course offerings of any single academic institution.
American Indian Higher Education Consortium	http://www.aihec.org/who-we-are/index.htm	Loriene Roy stated that tribal colleges have their own invitation-only event every year in mid-June. “They work in settings where it’s not just a library, it’s the tribal records office, language centers, etc.”
For dissemination to Type 4 language archives, we should visit individual language communities – it’s about building personal relationships.	N/A	

We also received further advice concerning the annual meeting of the Linguistic Society of America (LSA). Workshop participants suggested that we put together an organized panel session when we are ready. The call for presentations is in May. A panel could be used to push a resolution at LSA. The meaning of a “panel” at LSA is quite different from the anthropology conferences Wasson is familiar with. At LSA, panels are different from ordinary sessions. There are only a handful each year, and they are endorsed by the executive committee. If we obtained an endorsement from LSA, we would have “a strong footing.”

7. Next Steps

The next stage in our research trajectory will be the development of UCD guidelines for language archives through the (re-)design of two or more archives. We would like to include both a Type 1 and a Type 4 language archive, in order to gain experience working with both ends of the spectrum. Once we have obtained funding, we expect to:

- Conduct research to identify current users of those language archives
- Conduct ethnographic research with Type 1 stakeholder groups and participatory research with Type 4 stakeholder groups, to identify cultural practices of use, as well as the political, economic, environmental, technological, and cultural contexts in which the language archives are situated
- Work with collaborators to translate research findings into design concepts for the language archive
- Develop a (re-)design of the language archive based on the design concepts; test the new design with users and iteratively improve it until users are satisfied
- Write up UCD guidelines for language archives based on our learnings

We expect the guidelines to describe a collaborative process among language archive stakeholder groups, including user researchers, designers, and whatever other specialists may be needed. The guidelines will recognize that the design of each archive needs to be customized to the concerns of its language communities. They will seek to accommodate the potentially conflicting needs of different user groups.

The third and final stage of the long-term research trajectory will be a set of activities to encourage a paradigm shift in the field of language archives, so that the basic principles of UCD become well known and widely adopted. As mentioned in Section 6, we will partner with a number of organizations to develop organization-specific guidelines. We will also present our findings in a variety of different venues, using the archive (re)designs conducted in the previous stage as examples of what a UCD process looks like.

Even after the conclusion of this three-part research trajectory, the UCD practitioners who have gotten involved in this effort (Christina Wasson, Crysta Metcalf, Santosh Basapur) hope to continue their engagement with the issue. It is hard to predict specific activities this far in advance, but we might, for instance, see if we could continue to work with additional language archives. There might also be an opportunity to extend our findings to UCD guidelines for other kinds of archives.

Acknowledgements

We would like to extend our deepest appreciation to the workshop participants for giving so generously of their time, insights, and ideas. Our collaborative discussions were engaging, exciting, and productive. As this white paper documents, the collective knowledge generated was truly impressive. The participants provided a strong foundation on which we can construct the next stage of our research trajectory.

We thank the following workshop participants for reviewing a draft of the white paper: Edward Alexander, Daryl Baldwin, Shobhana Chelliah, Susan Kung, Wesley Leonard, and Michael Shepard. The white paper has been greatly improved by their comments. Any remaining errors are the responsibility of the authors.

Christina and Gary would also like to express our profound gratitude to Heather Roth, Research Assistant extraordinaire. She has been amazingly helpful every step of the way, from event planning to literature reviews to preparation and analysis of workshop transcripts.

Finally, we are grateful to the U.S. National Science Foundation's Documenting Endangered Languages Program, whose support made this workshop possible.

References

- Aiken, J., Schlieder, V., & Wasson, C. (2014). "No more cakes and ale?": Discovering ethical gray areas in a design anthropology class. *Journal of Business Anthropology, Special Issue 1*(38-61).
- Austin, P. (2011, April 29). Who uses digital language archives? Retrieved from <http://www.paradisec.org.au/blog/2011/04/who-uses-digital-language-archives/>
- Bannon, L. (1991). From human factors to human actors: The role of psychology and human-computer interaction studies in system design. In J. Greenbaum & M. Kyng (Eds.), *Design at work: cooperative design of computer systems* (pp. 25-44). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Berez, A. (2013). *Reproducible research in descriptive linguistics: Integrating archiving and citation into the postgraduate curriculum at the University of Hawai'i at Mānoa*. Paper presented at the Conference on Research, Records and Responsibility, Melbourne.
- Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language, 79*, 557-582. Retrieved from <http://www.language-archives.org/documents/portability.pdf>
- Boas, F. (1917). Introduction. *International Journal of American Linguistics, 1*(1), 1-7.
- Dobrin, L. M., & Holton, G. (2013). The documentation lives a life of its own: The temporal transformation of two endangered language archive projects. *Museum Anthropology Review, 7*, 140-154. Retrieved from <http://scholarworks.iu.edu/journals/index.php/mar/issue/view/233>
- Foucault, M. (1982). *The archaeology of knowledge and the discourse on language*. New York: Pantheon.
- Franklin, B. (1743). A proposal for promoting useful knowledge among the British plantations in America. Broadside. Philadelphia.
- Gezelter, D. (2009, December 1). Being scientific: Falsifiability, verifiability, empirical tests, and reproducibility. Retrieved from <http://www.openscience.org/blog/?p=312>
- Green, D. (2003). *The NINCH guide to good practice in the digital representation and management of cultural heritage materials*. New York: Humanities Advanced Technology and Information Institute, University of Glasgow and National Initiative for a Networked Cultural Heritage, New York University.
- Hagan, W. T. (1978). Archival captive – The American Indian. *American Archivist, 41*(2), 135-146.
- Himmelman, N. P. (1998). Documentary and descriptive linguistics. *Linguistics, 36*, 161-195.
- Holton, G. (2011a). The role of information technology in supporting small and endangered languages. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (pp. 371-399). Cambridge: Cambridge University Press.
- Holton, G. (2011b). 'Unknown unknowns' and the retrieval problem in language documentation and archiving. *Language Documentation and Conservation, 5*, 157-168.
- Holton, G. (2012). Language archives: They're not just for linguists any more. In F. Seifart, G. Haig, N. P. Himmelman, D. Jung, A. Margetts, & P. Trilsbeek (Eds.), *Potentials of Language Documentation: Methods, Analyses, and Utilization* (pp. 111-117). Honolulu: University of Hawai'i Press.
- Holton, G. (2014). Mediating language documentation. *Language Documentation and Description, 12*, 37-52. Retrieved from <http://www.elpublishing.org/itempage/136>
- Holton, G., Berez, A., & Williams, S. (2007). Building the Dena'ina language archive. In L. E. Dyson, M. Hendriks, & S. Grant (Eds.), *Information Technology and Indigenous Peoples*

- (pp. 205-209). Hershey, Pennsylvania: Information Science Publishing.
- Hunter, G. (2003). *Developing and maintaining practical archives: A how-to-do-it manual*. Chicago: Neal-Schuman Publishers.
- Huvila, I. (2008). Participatory archive: Towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Archival Science*, 8, 15-36.
- Isaacman, A. F., Lalu, P., & Nygren, T. I. (2005). Digitization, history and the making of a postcolonial archive of Southern African liberation struggles: The Aluka Project. *Africa Today*, 52(2), 55-77.
- Krauss, M. E. (1992). The world's languages in crisis. *Language*, 68(1), 4-10.
- Leonard, W. Y. (2011). Challenging "extinction" through modern Miami language practices. *American Indian Culture and Research Journal*, 35(2), 135-160.
- Leonard, W. Y. (2016). Producing language reclamation by decolonizing "language." To appear in *Language Documentation and Description* 13.
- Linn, M. (2014). Living archives: A community-based language archive model. *Language Documentation and Description*, 12, 53-67. Retrieved from <http://www.e-publishing.org/itpage/137>
- Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.
- Povinelli, E. A. (2011). The woman on the other side of the wall: Archiving the otherwise in postcolonial digital archives. *Differences*, 22(1), 146-171.
- Ritter, F. E., Baxter, G. D., & Churchill, E. F. (2014). Chapter 2: User-centered systems design: A brief history. In *Foundations for designing user-centered systems* (pp. 33-54). London: Springer Verlag.
- Robinson, R. E. (1993). What to do with a human factor: a manifesto of sorts. *Special issue: New human factors. American Center for Design Journal*, 7(1), 63-73.
- Robinson, R. E. (1994). The origin of cool things. In *Design that packs a wallop: Understanding the power of strategic design. American Center for Design conference proceedings* (pp. 5-10). New York: American Center for Design.
- Seifart, F., Drude, S., Franchetto, B., Gasche, J., Golluscio, L., & Manrique, E. (2008). Language documentation and archives in South America. *Language Documentation and Conservation*, 2(1), 130-140.
- Seifart, F., Haig, G., Himmelmann, N. P., Jung, D., Margetts, A., & Trilsbeek, P. (Eds.). (2012). *Potentials of language documentation: Methods, analyses, and utilization*. Honolulu: University of Hawai'i Press.
- Shepard, M. (2015). *"The substance of self-determination": Language, culture, archives and sovereignty*. (Ph.D.), University of British Columbia.
- Shilton, K., & Srinivasan, R. (2007). Counterpoint: Participatory appraisal and arrangement for multicultural archival collections. *Archivaria*, 63, 87-101.
- Squires, S., & Byrne, B. (Eds.). (2002). *Creating breakthrough ideas: The collaboration of anthropologists and designers in the product development industry*. Westport: Bergin & Garvey.
- Stoler, A. L. (2010). *Along the archival grain: Epistemic anxieties and colonial common sense*. Princeton: Princeton University Press.
- Suchman, L. (1995). Making work visible. *Communications of the ACM*, 38(9), 56-65.
- Wasson, C. (2000). Ethnography in the field of design. *Human Organization*, 59(4), 377-388.
- Wasson, C. (2002). Collaborative work: Integrating the roles of ethnographers and designers. In S. Squires & B. Byrne (Eds.), *Creating breakthrough ideas: The collaboration of anthropologists and designers in the product development industry* (pp. 71-90). Westport: Bergin & Garvey.
- Wasson, C. (2005). *Celebrating the cutting edge*. Proceedings of the Ethnographic Praxis in

- Industry Conference (pp. 140-145). Redmond, WA.
- Wasson, C., & Metcalf, C. (2013). Teaching design anthropology through university-industry partnerships. In W. Gunn, T. Otto, & R. C. Smith (Eds.), *Design anthropology: Between theory and practice* (pp. 216-231). London: Berg.
- Wasson, C., & Squires, S. (2012). Localizing the global in technology design. In C. Wasson, M. O. Butler, & J. Copeland-Carson (Eds.), *Applying anthropology in the global village* (pp. 251-284). Walnut Creek: Left Coast Press.
- Wikipedia. (2016). User-centered design. https://en.wikipedia.org/wiki/User-centered_design. Accessed 15 January 2016.
- Woodbury, A. C. (2003). Defining documentary linguistics. In P. Austin (Ed.), *Language documentation and description, Volume 1* (pp. 33-51). London: Hans Rausing Endangered Language Project.
- Woodbury, A. C. (2014). Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. *Language Documentation and Description, 12*, 19-36. Retrieved from <http://www.elpublishing.org/itempage/135>
- Zeitlyn, D. (2012). Anthropology in and of the archives: Possible futures and contingent pasts. Archives as anthropological surrogates. *Annual Review of Anthropology, 41*, 461-480.