



Vol. 10 (2016), pp. 411–457
<http://nflrc.hawaii.edu/lc>
<http://hdl.handle.net/10125/24714>
Revised Version Received: 19 April 2016

Series: Emergent Use and Conceptualization of Language Archives
Michael Alvarez Shepard, Gary Holton & Ryan Henke (eds.)

A Brief History of Archiving in Language Documentation, with an Annotated Bibliography

Ryan E. Henke

University of Hawai'i at Mānoa

Andrea L. Berez-Kroeker

University of Hawai'i at Mānoa

We survey the history of practices, theories, and trends in archiving for the purposes of language documentation and endangered language conservation. We identify four major periods in the history of such archiving. First, a period from before the time of Boas and Sapir until the early 1990s, in which analog materials were collected and deposited into physical repositories that were not easily accessible to many researchers or speaker communities. A second period began in the 1990s, when increased attention to language endangerment and the development of modern documentary linguistics engendered a renewed and redefined focus on archiving and an embrace of digital technology. A third period took shape in the early twenty-first century, where technological advancements and efforts to develop standards of practice met with important critiques. Finally, in the current period, conversations have arisen toward participatory models for archiving, which break traditional boundaries to expand the audiences and uses for archives while involving speaker communities directly in the archival process. Following the article, we provide an annotated bibliography of 85 publications from the literature surrounding archiving in documentary linguistics. This bibliography contains cornerstone contributions to theory and practice, and it also includes pieces that embody conversations representative of particular historical periods.

1. Introduction It is difficult to imagine a contemporary practice of language documentation that does not consider among its top priorities the digital preservation of endangered language materials. Nearly all handbooks on documentation contain chapters on it; conferences hold panels on it; funding agencies provide money for it; and even this special issue evinces the central role of archiving in endangered language work. In fact, archiving language data now stands as a regular and normal part of the field linguistics workflow (e.g., Thieberger & Berez 2011).

This state of affairs has not always been the norm. Moreover, the idea of archiving as an ongoing process instead of something to be done at the end of one's career is a relatively new development. This paper is a historical exploration of the chain of

events that have led us to this state, beginning in the late eighteenth century and continuing through to the present day.

Traditionally, archived resources consisted of physical objects (e.g., books, tools, photographs, artwork, and clay tablets), and because of the value of such objects, archives restricted access to them to varying degrees (Austin 2011, Nordhoff & Hammarström 2014, Trilsbeek & Wittenburg 2006). Typical homes for archived materials have long included museums, libraries, universities, and, of course, dedicated archival institutions (Linn 2014). In terms of access, this traditional model of archiving has entailed a ‘one-way’ street: Depositors put material into archives managed by archivists, and only people with the requisite permission and ability can find and access archived resources (Nathan 2014). In a nutshell, this was more or less the model for archiving from the beginning of modern linguistic work.

In order to provide a foundation for assessing how conceptualizations of archiving have changed dramatically, especially over the last twenty-five years, it is helpful to define what we mean by *endangered language archive*. We take *archive* to mean “a trusted repository created and maintained by an institution with a demonstrated commitment to permanence and the long-term preservation of archived resources” (Johnson 2004:143). Furthermore, this history is concerned primarily with archives designed to preserve materials related to small, endangered, and/or Indigenous languages.

We have identified four major periods in the development of endangered language archiving, each of which is discussed in the sections below:

- An early period, lasting from before the time of Boas and Sapir until the early 1990s, in which analog materials—everything from paper documents and wax cylinders to magnetic audio tapes—were collected and deposited by researchers into physical repositories that were not easily accessible to other researchers or speaker communities (§2);
- A second period, beginning in the 1990s, in which increased attention to language endangerment and language documentation brought about a redefined focus on the preservation of languages and language data (§3);
- A third period, starting in the early twenty-first century, in which technological advancements, concerted efforts to develop standards of practice, and large-scale financial support of language documentation projects made archiving a core component of the documentation workflow (§4);
- The current period, in which conversations have arisen toward expanding audiences for archives and breaking traditional boundaries between depositors, users, and archivists. (§5).

In §6, we present some critical review of the current state of archiving. This includes assessing how archiving has actually permeated the workflow of documentary linguists as well as how our field acknowledges and rewards scholarly and professional contributions in archiving.

2. Early linguistic archiving: Late 19th century–1991 For Americanist pioneers like Franz Boas and Edward Sapir in the late nineteenth and early twentieth centuries, archiving was an essential component of the work to document Indigenous languages (Johnson 2004). Documentation during this period consisted mostly of textual materials such as fieldnotes, translations, elicitation data, lexical compilations, and grammatical descriptions (Golla 2005, Johnson 2004). Throughout this period, linguists deposited their records in archives, universities, and museums; even monographs from such institutions as well as publications like the *International Journal of American Linguistics* served as “archiving mechanisms” for texts, grammars, and dictionaries from Indigenous languages, inasmuch as they became part of the published record (Woodbury 2011:163). However, with the exception of publications, such collections were available only to researchers with the inclination and capabilities to travel to archives and access the materials (Johnson 2004).

This conceptualization of archiving as the protection of physical items behind a brick-and-mortar wall remained relatively stable for many decades, and several notable archival institutions arose during this period. Among the most significant are the following:

1. Since its founding in 1743, the American Philosophical Society (APS)¹ collected Native American manuscripts, including a famous and extensive collection from Thomas Jefferson (Golla 1995). With its 1945 acquisition of the Franz Boas Collection of American Indian Linguistics from the American Council of Learned Societies, the APS became the “primary repository for the records of twentieth-century American Indian linguistics” (1995:148).
2. The University of California, Berkeley has been involved with archiving linguistic data since the early twentieth century,² beginning with the work of A. L. Kroeber, Pliny Earle Goddard, T. T. Waterman, Edward Sapir, and E. W. Gifford (Golla 1995). The Survey of California Indian Languages was officially founded at Berkeley in 1953 and renamed The Survey of California and Other Indian Languages in 1965. The leadership of Murray Emeneau and Mary Haas yielded a particularly important period: Under their direction, Berkeley housed “a veritable factory of graduate students who produced Boasian grammar-dictionary-text trilogies published by the *University of California Publications in Linguistics*. These texts were linked to audio-recordings which, along with field notes and slip-files, were archived with the Survey of California Indian Languages” (Woodbury 2011:166).
3. The National Anthropological Archives (NAA)³ was created in 1965 from a merger between the Department of Anthropology at the Museum of Natural History of the Smithsonian Institution and the Bureau of American Ethnology (BAE). The latter was the “most active sponsor of linguistic research on American Indian languages” during the late nineteenth and early twentieth centuries

¹<https://www.amphilsoc.org/>

²<http://linguistics.berkeley.edu/~survey/about-us/history.php>

³<http://anthropology.si.edu/naa/index.htm>

(Golla 1995:148). Along with many other linguists, the BAE employed John Peabody “J. P.” Harrington from 1915 until 1954, and he produced a massive amount of documentary linguistic work (Golla 1995, Macri & Sarmiento 2010).

4. In 1972, Michael Krauss founded the Alaska Native Language Center (ANLC), later renamed the Alaska Native Language Archive (ANLA), at the University of Alaska, Fairbanks.⁴ ANLA’s archival library contains an unparalleled collection of print and audio materials from and about Alaska’s 20 Indigenous languages (Krauss 1974, Woodbury 2010, Holton 2012, Holton 2014).

From Boas’ time onward, technological developments changed linguistic fieldwork as well as the types of materials stored in archives. As noted, text (whether handwritten or created via typewriter) had always served as a cornerstone of linguistic archives, but the beginning of the twentieth century also brought about the capacity to archive audio materials. Linguists captured and archived sound data using a progression of technology, employing wax cylinders (used to collect, for example, recordings of Native American music and language for the BAE) until the arrival of the phonograph in the 1930s (used by linguists like Melville Jacobs and J. P. Harrington), which was then replaced by tape recording technology in the 1950s before video recording technology became widely available in the 1980s (Golla 1995, Johnson 2004, Thieberger & Musgrave 2007). Of course, these analog methods gave way to the rise of digital technology in the latter half of the twentieth century: The digital archiving of language materials finds its origins in the use of computers for social science research in the early 1960s (Austin 2011, Doorn & Tjalsma 2007). The Oxford Text Archive,⁵ founded in 1976 by Lou Burnard, represents one of the earliest text archives in use by linguistic communities (Doorn & Tjalsma 2007), and the Linguistic Data Consortium was formed at the University of Pennsylvania in 1992 to address data shortages by serving as a repository and distributor for language resources.⁶

This progression to digital technology brought increasing efficiency and ease for data collection, but not enough attention went toward devising bigger and better ways to archive linguistic material systematically and sustainably. For example, Indiana University began the Archives of the Languages of the World in the mid-1950s to store vast volumes of tape records, but a lack of technical support forced the abandonment of the project (Golla 1995).⁷ At least part of the problem stemmed from the fact that traditional archives were not equipped to handle the massive amounts of data being produced, whether in terms of providing long-term storage or managing access by researchers or communities (Johnson 2004). Untold masses of text materials and thousands of hours of recordings, which had been accumulating for decades in the possession of linguists and anthropologists around the world, sat idle—only a fraction of linguistic data managed to make it into dedicated archives (Johnson 2004, Trilsbeek & Wittenburg 2006). This state of affairs did not change much until the

⁴<https://www.uaf.edu/anla/about/>

⁵<http://ota.ox.ac.uk/>

⁶<https://www ldc.upenn.edu/about>

⁷This collection has been subsumed into the Indiana University Archives of Traditional Music: <http://www.indiana.edu/libarchm/index.php/atm-collections.html>.

1990s, which saw the rise of documentary linguistics and a renewed and redefined focus on archiving.

3. Documentary linguistics and a new approach to archiving: 1991–2006 In the early 1990s, a growing number of linguists turned their attention to the problem of mass language endangerment and death (e.g., Hale et al. 1992). These scholars perceived an unprecedented crisis in the field, and the conversation began toward finding solutions: “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” (Krauss 1992:10). Soon after, this concern helped fuel Himmelmann’s (1998) refinement of *documentary linguistics* (or *language documentation*) as a distinct subfield of linguistics, although some say this was simply a homecoming back to the discipline’s roots as a fieldwork-based research enterprise, as mainstream linguistics had become increasingly more theoretical since the generative revolution of the 1950s and 1960s (Conathan 2011, Himmelmann 2006, Thieberger & Musgrave 2007, Woodbury 2003).

But what makes documentary linguistics different from descriptive linguistics? Traditionally, *descriptive* linguistics revolves around the Boasian trilogy of texts, dictionaries, and grammars based on in-depth analyses of primary data from a given language (Himmelmann 1998, Himmelmann 2006, Woodbury 2003, Woodbury 2011). *Documentary* linguistics is much broader and more ambitious in scope. As Himmelmann himself defined it, a language documentation is a “record of the linguistic practices and traditions of a speech community” (1998:166). Woodbury (2003:46–48) usefully elaborated upon this definition by proposing some widely agreed-upon values for proper documentation: A good documentation is diverse, large, ongoing, distributed, and opportunistic with material that is transparent, preservable, ethically created, and portable. Broadly speaking, a documentation provides a sizeable record of a language in use across a range of discourse, furnishing a copious amount of transcribed and annotated audio/video materials accompanied by contextual metadata (Austin 2013, Austin & Grenoble 2007, Johnson 2004). This creates “a lasting, multipurpose record of a language” (Himmelmann 2006:1), which can be employed not only to address language endangerment but also to provide data for linguistics and other disciplines, improve scientific accountability, and maximize the economy of research resources. As such, another element distinguishing modern documentation efforts from those of the past is “concern for long-term storage and preservation of primary data” (Himmelmann 2006:15). We return to this point later.

A handful of major factors enabled the rise of documentary linguistics during this time period (Austin 2012, Austin 2014, Austin & Grenoble 2007, Woodbury 2003). First, of course, was the increased attention to language endangerment. A second factor was the increase in funds for documentary projects, primarily from three major sources: Germany’s Volkswagen Foundation, which began the Dokumentation bedrohter Sprachen⁸ (DOBES) program in 2000; the Arcadia Trust⁹ in

⁸<http://dobes.mpi.nl/dobesprogramme/>

⁹<http://www.arcadiafund.org.uk/about-arcadia/about-arcadia.aspx>

the United Kingdom, which started the Endangered Languages Documentation Programme¹⁰ (ELDP) in 2003; and the National Science Foundation and the National Endowment of the Humanities, which together initiated the Documenting Endangered Languages^{11,12} (DEL) program in 2005 (Austin 2012, Austin 2014, Woodbury 2003). Other notable funders emerged in this period as well, such as the Community-University Research Alliance¹³ and the Aboriginal Research Programme¹⁴ of the Social Sciences and Humanities Research Council of Canada (SSHRC), the Foundation for Endangered Languages¹⁵ (FEL) in the United Kingdom, and the Endangered Language Fund¹⁶ (ELF) in the United States (Woodbury 2011). Finally, modern documentary linguistics was able to emerge due to monumental developments in digital information technology, which enabled more efficient and higher quality recording of audio and video; processing, analysis, and storage of such materials; and the widespread distribution of such information through the internet—all to extents that were previously impossible (Austin 2013, Austin & Grenoble 2007, Bird & Simons 2003, Evans & Dench 2006, Johnson 2004, Woodbury 2003). In 1991 the Australian Institute of Aboriginal and Torres Strait Islander Studies¹⁷ (AIATSIS) created what might be the first digital archive dealing with endangered languages, the Aboriginal Studies Electronic Data Archive¹⁸ (Thieberger 1994).

Along with this new conceptualization of documentary linguistics came a renewed and redefined focus on archiving. From the beginning, archiving occupied one of the four steps laid out in Himmelmann's model of documentation: "presentation for public consumption/publicly accessible storage (archiving)" (1998:171). A host of scholars agreed that archiving is a cornerstone of documentation, (e.g., Austin & Grenoble 2007, Johnson 2004, Rehg 2007, and Woodbury 2003). The reason for this is simple: If we are going to dedicate immense amounts of time, money, and energy to preserve endangered languages, then all of our efforts would be futile without a plan for that information to be put to use safely and sustainably by future generations for a variety of purposes—including facilitating studies in a range of scientific disciplines, enabling verification of data analyses, and producing language teaching materials (Austin 2014, Evans & Dench 2006, Himmelmann 2006, Nathan 2014, Thieberger & Musgrave 2007). This perspective on archiving is considered by some to be another factor distinguishing documentation from description (Himmelmann 2006, Nathan & Austin 2014). With this new outlook on archiving, it was not long before many came to see an inseparable relationship between language documentation and the archive: "All documentation projects should be conceived with an eye toward the ultimate deposit of the recorded data and analysis in an archive" (Austin & Grenoble

¹⁰<http://www.eldp.net/>

¹¹https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12816

¹²<http://www.neh.gov/grants/preservation/documenting-endangered-languages>

¹³<http://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/cura-aruc-eng.aspx>

¹⁴http://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/priority_areas-domaines_prioritaires/aboriginal_research-recherche_autochtone-eng.aspx

¹⁵<http://www.ogmios.org/index.php>

¹⁶<http://www.endangeredlanguagefund.org/>

¹⁷<http://aiatsis.gov.au/>

¹⁸<http://aseda.aiatsis.gov.au/asedaDisclaimer.php>

2007:19). Importantly, it was not just linguists who came to regard archiving as an integral part of language documentation—so did a lot of the people with the money: Organizations like DOBES, EDLP, DEL, and the ELF have come to mandate archiving as part of their documentation project requirements (Austin 2014).

Finally, along with this new view of language documentation, linguists increasingly acknowledged the importance of archiving to Indigenous language revitalization efforts (e.g., Gerds 2010 and Johnson 2004), which had been gaining steam particularly in the United States since the late 1960s (Gehr 2013). As Hinton (2001) explained, revitalization efforts often begin with a search for existing documentation, which may be housed in large national archives like the Smithsonian or in small, local archives. Moreover, when a strong reliance on native speakers is not possible, the development of pedagogical materials for revitalization efforts, such as dictionaries or language lessons, is often based on archived linguistic documents (2001). The oft-cited case of the Mutsun language represents a famous case for the value of archiving: Records from the nineteenth and early twentieth centuries enabled the production of a grammar in 1977—more than 40 years after the death of the last speaker—as well as subsequent revitalization endeavors (Conathan 2011, Macri & Sarmiento 2010). In 1996, one of the most significant American revitalization efforts began when the Advocates for Indigenous California Language Survival¹⁹ held its first Breath of Life Workshop,²⁰ bringing Indigenous community members to the Berkeley archives to teach them linguistic fundamentals and show them how to use archived materials to facilitate language restoration. Another example of a revitalization program began around 2000 in Canada, when Peter Brand and SENĆOŦEN speaker and teacher John Elliott, Sr. began using the internet to “support Aboriginal people engaged in language archiving, language teaching, and culture revitalization” through the FirstVoices²¹ project (Czaykowska-Higgins 2009:31).

By the early 2000s, documentary linguistics had arrived, and it brought a new conceptualization of the power and necessity of archiving. Now linguists faced the question: *How* should we archive?

4. How should we archive?: 2000–2010 Documentary linguists recognized the benefits conferred by *digital* archives. For one, digital information is not susceptible to the same problems of physical deterioration that plague wax cylinders, vinyl records, paper documents, magnetic tapes, and other analog materials—whether housed in traditional archives or sitting idle on researchers’ shelves (Bird & Simons 2003, Chang 2010, Johnson 2004, Nathan 2011). Some noticed this particular advantage early on, drawing attention to the need to digitally curate such legacy materials: “One of the major tasks of linguistic anthropology in the decades ahead will be to exercise appropriate stewardship over the archival record of American Indian languages” (Golla 1995:152). Other advantages of digital archives include providing much greater capacity for long-term preservation and storage of multimedia data

¹⁹<http://www.aicls.org/>

²⁰<http://www.aicls.org/breath-of-life>

²¹<http://www.firstvoices.com/>

(Nathan 2011), and enabling easier access to and retrieval of information (Trilsbeek & Wittenburg 2006). These capacities shattered conceptions of limitations on both the scope of a given documentary corpus as well as the ability of researchers to fact-check claims directly by going to the data. The following passage embodies this sentiment:

Digital audio and video recording, portable storage, and the development of software enabling the tagging, management and analysis of collected data raises the stakes for corpus collections. Our traditional published text collection consisted of a few hundred pages of narrative text with interlinear glosses, free translation and explanatory notes, but the modern published corpus may potentially consist of digital audio recordings of data collection sessions, some with accompanying video, and linked to a range of transcriptions representing different kinds and levels of analysis. Where the published text collection once served as the grounding evidence for a linguistic analysis, the digital archive will come increasingly to fill that role. (Evans & Dench 2006:24)

With this recognition of the possibilities granted by digital archives, many documentary linguists seemed mostly unaware that archivists outside of linguistics had already been working for a while to figure out *best practices*²² for digital archiving (Woodbury 2011). For instance, the Task Force on Archiving of Digital Information was created in 1994 by the Commission on Preservation and Access and the Research Libraries Group, and the task force reported in 1996 the need for trustworthy digital archiving organizations (Chang 2010). Moreover, between 1995 and 2002, the NASA Consultative Committee for Space Data Systems²³ developed the Reference Model for an Open Archival Information System (OAIS), which aimed at requirements for long-term preservation of digital information, including navigating issues with changing user communities and technologies (2010). Years later, “the OAIS Reference Model continues to have wide acceptance in the digital library community, and has become the authoritative model for best practices in digital archiving” (2010:61). Despite an ostensible lack of interdisciplinary communication in this regard, documentary linguists in the early and mid-2000s (e.g., Bird & Simons 2003, Evans & Sasse 2004, and Himmelmann 2006) were becoming increasingly interested in figuring out the best ways to carry out digital archiving of language documentation.

Bird and Simons (2003) took one of the earliest and most important steps toward best digital archiving practices. They called attention to some of the biggest issues facing documentary linguists looking to make data as long-lasting and usable as possible. For example, Bird and Simons noted that “a substantial fraction of the resources being created can only be reused on the same software/hardware platform, within the same scholarly community, for the same purpose, and then only for a period of

²²According to E-MELD (Electronic Metastructure for Endangered Language Data), best practices for digital archiving of linguistic work are “practices which are intended to make digital language documentation optimally longlasting, accessible, and re-usable by other linguists and speakers” (<http://emeld.org/school/what.html>).

²³<http://public.ccsds.org/default.aspx>

a few years.” (2003:579). To fix this problem, they called for a sea change in both technologies and attitudes. In their words: “We need nothing short of an open source revolution, leading to new open source tools based on agreed data models for all of the basic linguistic types, connected to portable data formats, with all data housed in a network of interoperating digital archives” (2003:579).

Another topic in best-practices conversation focused on approaches toward metadata. Metadata, often described as data *about* data, accompanies primary data to provide valuable context and meaning (e.g., speaker identification, date of recording, and genre of text), and is especially useful in determining how data can be located in an archive and how it can and should be used (Austin 2013, Innes 2010, Thieberger & Berez 2011). An important metadata development came in December 2000 with an NSF-funded workshop, Web-Based Language Documentation and Description, held in Philadelphia (Bird & Simons 2003). This workshop gave rise to the founding of the the Open Language Archives Community²⁴ (OLAC), which is devoted to “(i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resource.” (Bird & Simons 2003:572-573). Among OLAC’s contributions are the OLAC Metadata standard²⁵ and the OLAC Repositories standard,²⁶ a protocol for harvesting metadata (2003). Another metadata standard arose during this time, too:

The International Standards for Language Engineering Metadata Initiative²⁷ (IMDI), developed by DOBES, which “is a more comprehensive metadata system that can be used to manage several archival functions, including not only description but also preservation and access” (Conathan 2011:246). Both the OLAC and IMDI schemas have come to be endorsed and adopted by many documentary linguists (Johnson 2004, Himmelmann 2006, Thieberger & Berez 2011).

Other best-practice discussions centered on the collection and management of primary data. For example, Austin (2006) covered ways to manage various forms of data involved in a language documentation, including how to select and use recording equipment, choose data formats (e.g., XML, WAV, or MPEG2), transfer analogue materials to digital form, and process data with software tools like Shoebox. Gippert (2006) discussed the history of and best practices for digitally encoding text (e.g., problems with ASCII and the power of Unicode), including managing structural elements like phrases and clauses. Robinson (2006) talked about the importance of archiving directly from the field to enhance the safety of collected data in conditions that are often inhospitable to electronics. Schroeter and Thieberger (2006) explored the need to have standard data structures, provided to linguists through templates and workflow directives, that can apply across various tools for transcribing and annotating linguistic data. Thieberger (2010) further dealt with data management, location and citation, formation, storage, reuse, and interoperability—while stress-

²⁴<http://www.language-archives.org/>

²⁵<http://www.language-archives.org/OLAC/metadata.html>

²⁶<http://www.language-archives.org/OLAC/repositories.html>

²⁷<http://www.mpi.nl/imdi/>

ing the need for training other linguists in best practices like employing consistent file naming, using OLAC metadata standards, and making data searchable by others. Following Bird & Simons (2003), one of the most important best practices to emerge during this period was the insistence on the use of open-source and uncompressed data formats for collecting and structuring linguistic data (e.g., Good 2011 and Thieberger 2010), which together help stave off obsolescence and make information as rich, long-lasting, and accessible as possible. By 2010, the discussion about *how* to archive even resulted in at least one MA thesis providing a checklist intended to help language documenters choose the proper archive for their deposits (Chang 2010).

Concomitant with these discussions in the literature came the development of organizations and initiatives devoted to implementing and disseminating best practices for archiving language documentation. Established in 2001 after a one-year pilot project, the DOBES program at the Max Planck Institute in the Netherlands mandated that its funded projects adopt “specifications for archival formats, recommendations about recording and analysis formats, and the development of new software tools to assist with audio and video annotation (such as ELAN), and the creation and management of metadata (various IMDI tools)” (Austin 2014:61).²⁸ From 2001 to 2006, the National Science Foundation funded the Electronic Metastructure for Endangered Language Data²⁹ (E-MELD) project, which aimed at creating consensus and sharing information on best practices in documentation, including data markup, labels for interlinear glossing, and metadata creation (Austin 2014, Boynton et al. 2006). E-MELD has particular importance because it represented the first time linguists came together to create a significant set of digital standards for documentation. As part of the task of creating stronger networks within the archiving community, the Digital Endangered Languages and Musics Archives Network³⁰ (DELAN) came about in 2003 as an international umbrella body dedicated to creating stronger networks within the archiving community. The push for best practices even resulted in a newsletter that ran from 2004 to 2007, the *Language Archives Newsletter*,³¹ which was specifically devoted to issues in archiving (Woodbury 2010).

Furthermore, established archival projects were increasingly going digital (Trilsbeek & König 2014), and new archives emerged with a focus on digital formats and best practices. The ANLC became a founding member of OLAC in 2000, creating an electronic catalog database as well as a digital archive for the Dena’ina Qenaga language (Holton 2014, Holton et al. 2006). The Archive of the Indigenous Languages of Latin America³² was founded in 2000 at the University of Texas at Austin. Three years later, linguists and musicologists established the Pacific and Regional Archive for Digital Sources in Endangered Cultures³³ (PARADISEC) to digitize and curate field recordings compiled since the 1960s by Australian researchers (Thieberger &

²⁸<http://tla.mpi.nl/tools/tla-tools/elan/>

²⁹<http://emeld.org/>

³⁰<http://www.delaman.org/>

³¹<http://www.mpi.nl/LAN/>

³²<http://www.ailla.utexas.org/site/welcome.html>

³³<http://paradisec.org.au/>

Barwick 2012, Thieberger 2013, Thieberger et al. 2015a). After a year of development, the Hans Rausing Endangered Language Project at the School of Oriental and African Studies opened the Endangered Language Archive³⁴ (ELAR) in 2005 (Nathan 2010, 2014). Modeled on PARADISEC, yet another digital archive opened in 2008: *Kaipuleohone*, the University of Hawai'i Digital Language Archive,³⁵ which aims to make extant research more discoverable and to preserve language documentation materials (Albarillo & Thieberger 2009, Berez 2013, Berez 2015, Rehg 2007).

From around 2000 to 2010, it appears that documentary linguists had largely succeeded in establishing a general set of (or at least a very rich dialogue around) best technological practices along with initiatives and organizations for digitally archiving language documentation data. However, throughout this period we also see a recognition of various limitations and problems associated with digital archiving. This includes challenges to the idea that a single, comprehensive set of 'best practices' makes sense, given the wide spectrum of language documentation situations. This critical response has also been observed and discussed at length by Austin (2014:62–65). Austin (2013:4) summarized the situation well:

Some researchers have emphasised standardization of data/metadata and analysis and “best practices” (e.g., E-MELD, OLAC) while others have argued for a diversity of approaches which recognize the unique and particular social, cultural and linguistic contexts within which individual languages are used.

For example, Bowden & Hajek (2006) pointed out that seemingly 'best' practices are not always relevant or possible to carry out, given varying circumstances in the field: Perhaps there is no electricity; team members may be spread out over wide distances, which inhibits workflow; or local community members may be completely unfamiliar with digital technology. In the face of challenges such as diverging goals and cumbersome workflows, Berez & Holton (2006) noted the difficulties of getting speaker communities—and even other linguists—on board to adopt best practices for long-term data preservation.

On the other hand, arguments also critiqued the limited vision of existing best-practice concepts. For instance, Johnson (2004) and Nathan & Austin (2004) called for richer contextual information to be added to metadata, claiming that existing metadata standards and archival protocols do not go far enough in adding value to data. Nathan (2009) cited the need for an 'epistemology' for audio recording in language documentation, one that goes beyond existing discussions limited to formats and resolution to deal with recording spatial and configuration information as well as controlling signal and noise. Still others pointed out that archived materials will have uses beyond the original purposes for which they were collected and archived: “It is imperative for linguists to understand both the possibilities and the limitations of current archival practices so they can prepare for and advocate for the best possible management of the records they create, and of legacy archival collections” (Conathan

³⁴<http://elar.soas.ac.uk/>

³⁵<http://kaipuleohone.org>

2011:236). Ironically, plenty of time, attention, and resources had been spent developing and promoting best practices regarding documentary linguistic data, but linguists still had not conceived a system to test the effectiveness and longevity of language archives themselves (Chang 2010).

Perhaps the biggest reaction to best-practices conversations has concerned variegated issues of ethics and access (e.g., Dwyer 2006, Green et al. 2011, and Innes & Debenport 2010). Although the digital nature of archives can allow for easier, increased access of archived materials, this is not always a simple matter. For instance, O’Meara and Good (2010) raised issues relating to defining a ‘community’; establishing rights to access archived material retroactively; establishing rights and access to “orphan” works that do not have an identifiable copyright holder; and assessing and dealing with sensitivities related to the content of archived materials. Garrett and Conathan (2009) described problems resulting from failures of planning by linguists and archives, which are compounded when parties—whether linguists, speakers, or heritage communities—seek restrictions to access for materials. Although Garrett and Conathan suggested having a consistent, comprehensive, and clear strategy for archiving and developing access restrictions in consultation with heritage communities, this cannot solve every problem. We see this, for example, with informed consent (e.g., Thieberger & Musgrave 2007). Given the fact that linguists and speakers cannot exhaustively anticipate future technological developments and new uses for language documentation data, Thieberger and Musgrave wondered “how the data collector can fully inform the speakers about the nature of the activities to be undertaken” (2007:31). Other ethical dilemmas involve increased public access to sensitive materials, where community members may regard archived data (e.g., narratives, songs, and stories) as sacred, embarrassing, or even dangerous to others (Innes 2010,³⁶ Macri & Sarmiento 2010, Thieberger & Barwick 2012). In such cases, linguists may have an ethical responsibility of “providing as rich a system of ethnographic information as possible,” such as ideological statements and behavioral descriptions, in order to ameliorate future problems with the reinstatement or reproduction of archived texts and discourse (Innes 2010:202). Finally, ethical concerns arise from the fact that “the rules of intellectual property, although set by international standards, often conflict with customs of traditional indigenous groups” (Macri & Sarmiento 2010:195).

This has certainly not been an exhaustive account of all the reactions to the “best practices” conversation during this period. However, they do illustrate the broader progression of history: By around 2010, documentary linguists had developed a healthy discourse around both 1) establishing sustainable digital archives that last a long time, permit access to various parties, and provide utility to scientists and speech communities; and 2) grappling with the problems and limitations of trying to squeeze a one-size-fits-all archival approach upon the varied, idiosyncratic contexts of the field. Archiving in language documentation had come a long way in a very short

³⁶In the case of her Mvskoke language work, Innes simply chose to stop working with some sensitive materials: “Here, I find that I cannot continue to work on these narratives as this causes my consultants real difficulty and concern” (2010:202).

amount of time, and new discussions soon began around further reconceptualizing the model and role of the archive.

5. Redefining archiving through participatory models: 2010–present Throughout the transition from traditional analog repositories to the power and potential of digital archives, we see the persistence of a “one-way” model of archiving: “providers lodge their materials with the archive and users can (if permissions allow) find and access them” (Nathan 2014:193). This model entails limits on the interaction between depositors and users and between users and archived material (Trilsbeek & Wittenburg 2006). Throughout, the archivist is at the center of the archiving process. In the last few years, however, this situation has changed dramatically with the development of *participatory* archiving models in linguistics. Specifically, one definition of a *participatory archive* is “an organization, site or collection in which people other than the archives professionals contribute knowledge or resources resulting in increased understanding about archival materials, usually in an online environment” (Theimer 2011). The rise of such a model in linguistics seems to have been enabled by four primary factors:

1. The development of community-oriented models of linguistic research
2. The increasing empowerment of Indigenous communities in stewarding their own languages
3. The integration of social media models in archiving
4. The development of participatory models in the archival sciences

5.1 Community-oriented research By late in the first decade of the twentieth century, documentary linguists were increasingly turning to models of research that relied upon collaboration with language communities (e.g., Cameron et al.’s 1992 “empowering” model). Of particular significance is the Community-Based Language Research (CBLR) model outlined by Czaykowska-Higgins in 2009 (author’s emphasis):

Research that is on a language, and that is conducted *for, with, and by* the language-speaking community within which the research takes place and which it affects. This kind of research involves a collaborative relationship, a partnership, between researchers and (members of) the community within which the research takes place (24).

The CBLR represents a departure from the traditional model of research in linguistics. For more than a century, research has mostly been carried out by linguists for an audience of linguists, regarding speakers and speaker communities primarily as sources of data—no matter how ethically conscious such engagements might actually be (2009). Although Czaykowska-Higgins was not the first linguist to advocate and practice a collaborative approach to research (e.g., Cameron et al. 1992, Dwyer

2006, and Yamada 2007), she was one of the first to put forth a clear, systematic model for others to follow.

Around this time, there seems to be a shift in the language documentation literature, a stronger acknowledgement of the value of collaborating with communities in linguistic enterprises and producing research that serves the interests of both linguists and speakers (e.g., Good 2011, Dorbin & Holton 2013). This move reconceptualizes the longstanding research paradigm by moving from treating communities as objects of study to “actively including them in the process of documenting their language” (Wilbur 2014:68).

5.2 Empowerment of Indigenous communities Another factor facilitating participatory developments in linguistic archiving has been the fact that Indigenous communities over the last several decades have taken increasing levels of agency and ownership in stewarding their languages through documentation and revitalization (Hinton 2001, Macri & Sarmento 2010). Native communities in the United States, for example, have been stepping up in language scholarship as well as producing materials like phrasebooks, dictionaries, and curricula for revitalization (Hinton 2005). As Indigenous archive activist Allison Boucher Krebs put it (2012:182):

Whereas historically the flow of information about Indian Country has been away from Indian Country and once outside, about Indian Country by scholars, researchers, and non-Indigenous professionals, today information is flowing back to communities and within communities. The scholars, researchers, and professionals are increasingly likely to be Indigenous.

Of course, this also means that Indigenous communities in the United States, Canada, and Australia have been taking much more active roles in archiving their cultural heritage. In 2005, Hinton noted that the archives at Berkeley were “being used far more by Native Americans than by social scientists for purposes of language and cultural maintenance and revitalization” (24–25), and Holton (2014) observed that ANLA has become an increasingly important resource for revitalization activities in Alaska since the late 1990s.

Indigenous communities have also been taking the reins by creating their own archival institutions (which are often locally based), organizations, and initiatives (Ormond-Parker & Sloggett 2012). In the United States, for instance, The Native American Archives Roundtable³⁷ was founded in 2005, and a year later the First Archivist Circle³⁸ issued the Protocols for Native American Archival Materials (Krebs 2012). Moreover, the Administration for Native Americans (ANA) and the Smithsonian National Museum of the American Indian issued a nearly 300-page reference guide for Indigenous communities interested in establishing archives (ANA 2005). The guide covers an extensive range of subjects, including: 1) why it is important to

³⁷<http://www2.archivists.org/groups/native-american-archives-roundtable>

³⁸www.firstarchivistcircle.org/

preserve Native language materials, 2) how to decide what to preserve, 3) what an archive is, 4) how to build an archive infrastructure, 5) how to use existing archives to find language materials, and 6) how to approach archiving costs. As a final example, Alaska's Ahtna community created its own archive, *C'ek'aedi Hwnax*, in 2009 to digitize, curate, and distribute Ahtna language materials—all under OLAC standards and best-practice guidelines undertaken by other archives (Berez et al. 2012, Berez 2013). Such developments exemplify how communities long regarded as objects of study have instead increasingly become leaders in the study and stewardship of their own languages.

5.3 Social networking and archiving A third factor leading to the development of participatory approaches to archiving in linguistics has been a move toward integrating archiving with social networking models (often called “Web 2.0”). Between 2005 and 2010, we saw “the explosive growth of social networking” (Nathan 2011:271), which aims to “link people rather than documents, with a focus on interaction and collaboration instead of passive downloading and viewing of content” (Austin 2014:65). The approach integrating archives and Web 2.0 was pioneered by ELAR in 2010, where “the archive is reconceived as a platform for conducting relationships between information providers (depositors) and information users” (Nathan 2010:111). This integration changes the nature of both access and distribution by allowing parties to negotiate directly with each other—rather than always going through an archivist/archive—which helps address problems such as accessing sensitive materials as well as managing the complexities of growing collections stewarded by small numbers of dedicated staff (Nathan 2010, 2011). This model, of course, shatters traditional boundaries of archiving: The digital archive is not just a place for preserving data; it has been reconceptualized as “a forum for conducting relationships between information providers (usually the depositors) and information users (language speakers, linguists and others)” (Nathan 2011:271). Nathan (2015:53) also discusses the concept of reach, an archive’s “multifaceted capacity to successfully provide language resources to those who can gain value from them.”

5.4 Development in the archival sciences Finally, as noted by Linn (2014), archival scientists had already been talking about “participatory models” in their own circles since at least the late 2000s. Shilton and Srinivasan (2007), for instance, confronted problematic issues of power entailed by traditional archives. In particular, archives have long directed the selection, collection, and curation of cultural materials from Indigenous communities—who are not involved in the archiving process—to represent those communities: “archives have appropriated the histories of marginalized communities, creating archives *about* rather than *of* the communities (authors’ emphasis; 2007:89). To address these problems, Shilton and Srinivasan advocated a Participatory Archiving Model that “encourages community involvement during the appraisal, arrangement, and description phases of creating an archival record” (2007:98). By arising in collaboration with Indigenous communities, a participatory model can help not only to restore power to marginalized people but also to improve the quality

of archives themselves by enhancing their contextual knowledge and value (2007). Huvila (2008:25) built upon this work to formulate the concept of a participatory archive, which has three defining characteristics: 1) *Decentralized curation*, where archivists and participants share curatorial responsibilities; 2) *Radical user orientation*, where the locatability and usability of archived materials takes priority over preservation and the archival process; and 3) *Contextualization of both records and the entire archival process*, which means that archives include knowledge and context provided by others involved in the archiving process, such as a language community. By 2011, participatory models of archiving had become ‘sexy’ within the archival sciences (Theimer 2011).

5.5 Participatory models of archiving in language documentation Given these four factors, the stage was set for a discourse in documentary linguistics around participatory archiving. By 2011, researchers and archivists were asking themselves how they could expand the usage and impact of archives beyond the limitations of their original conceptions. This entailed a recognition that an archive is not a finished, static repository for data—instead, it is an ever-unfinished research product that involves taking in new information, digitizing old materials, and navigating developments in digital infrastructures, formats, and standards (Albarillo & Thieberger 2009, Holton 2012). Aside from the four factors described above, efforts to expand archives were at least in part also motivated by financial realities:

In particular, now that some of the major language documentation funding initiatives are coming to an end, the question arises how maximum advantage can be gained from the archiving infrastructures that have been created, for example by encouraging a wider range of people to engage in documenting languages and to deposit their materials into archives, as well as by drawing more users to the various archives (Trilsbeek & König 2014:51–2).

Part of this process involves figuring out who uses archives and for what purposes. Austin (2011), for instance, ascertained that DOBES and ELAR seem to be used primarily by linguists, while ANLA and the California Language Archive are “essentially used by speaker communities or their descendants to access materials for cultural, historical or language-learning purposes.” Holton (2012) also found that ANLA users tend to be from Native language communities, who are often looking for information that is not necessarily, or at least primarily, linguistic. For example, he cited requests for ethnobotanical information, music, and even a eulogy from the nineteenth century—all for non-linguistic purposes. This usage trend in part reflects changing demographics in Alaska, where speaker numbers are declining and language archives often serve as the only records of languages (2012). At the same time, DOBES was exploring how to broaden the impact of its archived data by making it a more accessible resource for scientists and non-scientists interested in language questions (Schwartz 2012). As part of this effort, DOBES created a new general portal to “attract users

to the archive, facilitate access to the data, and generate new user scenarios and communities” (2012:126).

By 2014, discussions had started exploring the benefits of participatory archiving in documentary linguistics. Green et al. (2011) explained that getting language practitioners involved in both recording their language and making decisions about how to represent it is a good way to encourage not just participation in research but also the long-term availability of data. Furthermore, many linguists (e.g., Gardiner & Thorpe 2014, Garrett 2014, Nathan 2014, Linn 2014, and Woodbury 2014) asserted that participatory archiving models can increase levels of participation in and support for documentary projects among speaker communities, while also maximally engaging audiences and expanding usages for archived material—especially within language communities and other academic disciplines. Simply put, researchers and archivists started to spread the idea that a participatory model might be the best way to get the most out of an archival project.

This has recently led to specific recommendations for participatory models. Woodbury (2014:33) addressed three ways to help archives reach wider audiences “by developing more direct and explicit protocols of communication between documenters and audiences through the medium of language archives.” For language documenters, his proposal centers on a “book model,” which includes furnishing a guide for exploring a given documentary corpus, explaining the design of the corpus, assigning the corpus to a genre, and providing a narrative about how the data was compiled. For archivists, Woodbury has suggested an “art museum model,” based on the fact that such museums curate and provide access to materials. This model includes making the information in archives accessible and discoverable, ensuring that linguists provide adequate descriptions of what they have collected, inviting deposits from people who are not traditional language documenters, holding exhibitions to facilitate public outreach, and getting archives reviewed by both academic and popular outlets to provide public exposure and generate feedback. And for audiences, Woodbury has outlined a ‘critic’ model that consists of various levels of review for a documentary corpus by a variety of stakeholders (e.g., editors, other language documenters, and archivists).

Linn has recommended a Community-Based Language Archive (CBLA) model, where archives are part of the effort to “bring about community-driven social change through maintaining, revitalizing, or renewing language” (2014:56). Specifically, a CBLA is “an archive or collection that is focused on a language, and that cares for and disseminates documentation that is conducted for, with, and by the language-speaking community within which the documentation takes place and which it affects.” (61). Such an archive “actively engages with the relevant community in conducting all levels of documentation, describing and contextualizing, maintenance, and dissemination of information” (61).

In a similar vein, Garrett put forth a model for participant-driven language archiving (PDLA), “an archiving component that assigns role appropriate archiving rights and responsibilities to individuals and communities who participate as ‘human subjects’ of linguistic research” (2014:68). Although archives have traditionally focused

on building relationships with depositors, a “PDLA’s primary objective is to establish direct, web-based, relationships between participants and archives, minimizing the use of depositors as proxies” (69). In the PDLA model, community members become active participants in archiving. They work, for example, to enrich archival resources (e.g., improving or creating metadata) and improve communication between speakers and archives—helping, among other things, to address tricky issues like ongoing informed consent.

Some archives seem already to be moving toward a more participatory model. The Aboriginal and Torres Strait Islander Data Archive, for example, has as an “overarching goal” the “commitment to connect Indigenous Australian communities with research data” (Gardiner & Thorpe 2014:103). In the literature, many of these dedicated discussions of participatory archiving models in documentary linguistics began in 2014, several of which were in the pages of *Language Documentation and Description Volume 12: Special Issue on Language Documentation and Archiving*. This is all quite recent, but it appears that the movement is gaining steam. The next few years will show just where exactly this conversation is going and what its results will be for linguists, other researchers, archivists, and language communities.

6. Conclusion: How are we doing, and where are we going? This overview has divided the history of archiving in language documentation into four general periods:

- Archiving prior to the 1990s, when analog materials were collected and deposited in repositories that were difficult to access by anyone other than a select group of researchers with the requisite dedication, means, and permissions;
- The rise of documentary linguistics in the early 1990s and the subsequent distinction between linguistic *description and documentation*, which engendered both a renewed and redefined focus on archiving and an embrace of digital technology;
- Beginning in the early 2000s, the development of “best practices” for digital archiving and critical reactions addressing the variegated contexts of field situations and ethical issues in language documentation; and
- Since about 2010, developments toward participatory models for linguistic archiving, which break traditional boundaries between depositors, users, and archivists to expand the audiences and uses for archives while involving speaker communities directly in language documentation and archival processes.

Of course, these periods overlap with each other, and the conversations from one period do not—and should not—necessarily end with the beginning of the next.

For example, we are still seeing developments around best practices for digital archiving. Organizations like Innovative Networking in Infrastructure for Endangered Languages (inNET), founded in 2012, are still springing up and seeking better ways to reinforce and extend digital archive networks, facilitate the dissemination of

information to strengthen relationships between archives and the scientific community, promote common archiving standards to help shape archiving policies, and establish relationships between archives and non-scientific communities. Best-practices advocates (e.g., Thieberger 2012) continue to call important attention to the needs for improved methods and tools for language documentation, better metadata and more useful primary data, bigger data storage capacities, and wider promotion of best practices to both linguists and speaker communities.

The critical responses to “best practices” continue as well. Austin (2013:6), for example, says we need to go beyond the normal bounds of best-practice discussions to construct a theory of “meta-documentary linguistics,” which he defines as a “documentation of the documentation research itself” that describes “the methods, tools, and theoretical underpinnings for setting up, carrying out and concluding a documentary linguistics research project.” Linguists will also keep working on situation-specific solutions to problems in the field that present challenges for a one-size-fits-all approach to archiving (e.g., Bow et al. 2015). Dobrin and Holton (2013:140), for instance, have examined how the priorities and interests of a language community can shift over generations, “reactivating the documentary materials and community-researcher relationships in ways that were not anticipated by anyone involved.” Again, Austin (2014:62–65) has more on such critical responses.

The timeline presented here also implies that the development of endangered language archiving since the time of Boas has been an uninterrupted forward trajectory embraced widely by the field. Unfortunately, however, it has not necessarily been the case that linguists—either individually or collectively—have embraced the need for archiving, nor have we agreed upon how to assess the kinds of professional rewards that archiving ought to bring (Thieberger et al. 2015b). Archiving by documentary linguists is still by no means a universal practice, although the number of linguists for whom archiving is a task undertaken at regular intervals—as opposed to waiting until the end of a project or a career—is growing. This has been aided in part by increased awareness of the need to do so, and the falling financial burden of archiving on individuals. Among linguists who do archive regularly, though, most are motivated by personal or professional ideology rather than by discipline-wide expectation or hope of scholarly professional reward.

As an illustration, Gawne et al. (2015) find that very few descriptive linguists are transparent regarding their archiving practice in their publications, including making clear to readers that the primary data is archived, where it is archived, or how to access it. In a survey of more than 100 grammars completed between 2003 and 2012, it was found that only about 10 percent of authors included any reference to the archiving of the primary data upon which the publication was based (2015). This is likely due to the unclear rewards of data management in academia.

In 2010, the Linguistic Society of America passed its *Resolution Recognizing the Scholarly Merit of Language Documentation*,³⁹ in order to provide academic incentive for archiving by encouraging colleges and universities to consider the products

³⁹<http://www.linguisticsociety.org/resource/resolution-recognizing-scholarly-merit-language-documentation>

of documentation to be valid results of research. The resolution specifically supports the recognition of documentary materials such as the following:

[...] archives of primary data, electronic databases, corpora, critical editions of legacy materials, pedagogical works designed for the use of speech communities, software, websites, or other digital media [...] as scholarly contributions to be given weight in the awarding of advanced degrees and in decisions on hiring, tenure, and promotion of faculty. (Linguistic Society of America 2010)

The significance of the resolution is two-fold. First, the resolution acknowledges the value of scholarly work done in the service of increasing linguistic vitality and the inextricability of revitalization efforts from language documentation. Second, it notes that the scholarly products of language documentation go beyond the traditional peer-reviewed journal articles and into the realm of digital products, including archived corpora. Although the resolution is laudable in calling for recognition for archiving practices, it falls short in providing methods to do so. As of yet there is no discipline-wide metric for appraising the quality of preserved linguistic data sets, nor do we know of any departments of linguistics that have made their internal rating system widely available. The number of tenure and promotion cases in which archived collections of annotated data have been given the same weight as journal articles is likely very low. Without the promise of academic attribution, individual linguists have been slow to adopt an archiving workflow or cite primary data in publications.

The value of the historical overview presented here is to point out important trends that have developed within documentary linguistic archiving over the years—especially since the 1990s. At this point, it is also natural to wonder where things may be heading. It seems likely that the next several years will bring further developments in participatory models of archiving. For example, Trilsbeek and König (2014) suggest archives will likely continue to seek expanded audiences (especially in other academic disciplines) and increased community involvement by facilitating the documentation and depositing of archival materials with a range of tools such as smartphones apps. We may also see further development in large-scale, existing e-infrastructure projects (e.g., CLARIN and DARIAH) that will help researchers better share and integrate their work (2014). Moreover, we will also see more critical reactions to participatory models in archiving. What does it mean, for instance, if a community of speakers has no concept of ideas like “digital” and “access” (Robinson 2010, Stenzel 2014)? Importantly, participatory archiving will be part of the process of finding ways to evaluate “the quality, significance and value of language documentation research so that its position alongside such sub-fields as descriptive linguistics and theoretical linguistics can be assured” (Austin 2014:67).

Wherever we end up going, it will surely entail novel and exciting reconceptualizations of archives, expanded audiences, and brand-new uses for language documentation materials.

References

Administration for Native Americans (ANA). 2005. *Native language preservation: A reference guide for establishing archives and repositories*. <http://www.aihec.org/our-stories/docs/NativeLanguagePreservationReferenceGuide.pdf>

This is essentially a ‘how to’ manual for Indigenous communities interested in archiving for the purposes of language documentation and revitalization. As such, it covers a wide range of issues in an informative, practical manner while providing specific, real-world examples. Topics include choosing between (and even building from scratch) a physical or digital archive; concerns of access, copyright, and informed consent; salvaging damaged materials; locating and accessing language materials in existing community, university, government, and private archives; the monetary costs of various aspects of the archival process, including infrastructure maintenance, staffing and labor, and equipment and software; and preserving, copying, and migrating materials.

Albarillo, Emily E. & Nick Thieberger. 2009. Kaipuleohone, the University of Hawai‘i’s digital ethnographic archive. *Language Documentation & Conservation* 3(1). 1–14. <http://hdl.handle.net/10125/4422>.

This article documents the founding and first year of operation of the Kaipuleohone archive in the Department of Linguistics at the University of Hawai‘i at Mānoa. The archive is a response to both calls for institutes of higher education to be involved in the creation and preservation of digital collections, as well as the need for preservation of rare endangered language materials. Topics discussed include the purchase of digitization equipment and development of workflow procedures; preservation of materials in ScholarSpace, the University of Hawai‘i DSpace repository with an OLAC-compliant metadata catalog; and collaboration with other units on campus like the Music Department, the Anthropology Department, and the Charlene Sato Center for Pidgin, Creole, and Dialect Studies.

Austin, Peter K. 2006. Data and language documentation. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics Studies and Monographs 178), 87–112. Berlin: Mouton de Gruyter.

A data workflow for language documentation data is presented, alongside some brief overviews of various tools and file formats that the documenter may encounter along the way. The processes of documentation are recording, metadata creation, and capture (or digitization); these are discussed along with backup and file-naming procedures. Processing documentary materials includes linguistic analysis, archiving, and presentation. Although some of the software tools presented are outdated now, the value of this paper lies in recognizing which open formats have remained in use in today’s documentary workflow. For example, XML has persisted as a method for storing interlinearized glossed texts.

Austin, Peter K. 2011. Who uses digital language archives? <http://www.paradisec.org.au/blog/2011/04/who-uses-digital-language-archives/>.

This is a short, informal blog post, but in it Austin explores pivotal questions by asking the leaders of major language archives about their user bases. Austin shares brief replies from ANLA, DOBES, ELAR, and the Survey of California and Other Indian Languages. These responses describe who uses the archives, numbers of visitors (online and in person, if applicable), and their reasons for using the archives. Austin reports important differences: Regional archives are used more by language communities for “cultural, historical or language-learning purposes,” but the other archives are used primarily by researchers.

Austin, Peter K. 2013. Language documentation and meta-documentation. In Mari C. Jones & Sarah Ogilvie (eds.), *Keeping languages alive: Documentation, pedagogy and revitalization*, 3–15. Cambridge: Cambridge University Press.

Going beyond traditional ideas of best practices, this piece argues that documentary linguistics also needs a theory of *meta-documentation* that focuses on the theory, methodology, and tools of language documentation—as Austin describes it, “the documentation of the documentation research itself” (4). Austin suggests three different directions for approaching a theory of meta-documentation: 1) deductive, theorizing principles and then applying them to documentation projects; 2) inductive, extracting principles from actual documentation projects; and 3) comparative, examining the role of documentary linguistic metadata in light of what is done in related fields like anthropology and archaeology.

Austin, Peter K. 2014. Language documentation in the 21st century. *JournalLIPP* 3. 57–71.

The author takes a look at the defining characteristics and rise of language documentation, and he discusses changes in the field since 1995. This includes a review of developments in best practices in documentary linguistics, focusing on the efforts of DOBES and the E-MELD project. Importantly, Austin also reflects at length upon critical responses to the emphasis on best practices, which question whether there really is one ideal model for documentary linguistic research. Finally, the author considers developments in archiving, which includes the integration of social networking models and the reconfiguration of relationships between depositors, archives, and users. This article makes a great follow-up companion to Austin and Grenoble’s 2007 piece.

Austin, Peter K. & Lenore Grenoble. 2007. Current trends in language documentation. In Peter K. Austin (ed.), *Language Documentation and Description, Volume 4*, 12–25. London: SOAS.

Writing about 15 years after Hale et al.’s seminal 1992 call to action, Austin and Grenoble evaluate the then-current state of language documen-

tation. This includes a review of the theoretical underpinnings and goals of documentary linguistics, discussion of the kinds of projects language documentation can facilitate—especially linguistic research and language revitalization—as well as comments on issues of best practices and access rights. The authors also discuss the factors behind the emergence of documentary linguistics in the late twentieth century (e.g., technological advancements and the development of digital archives). The piece concludes with reflection upon important theoretical issues, including delineating the boundary between documentary and descriptive linguistics as well as defining a “comprehensive” documentation of a language.

Berez, Andrea L. 2013. The digital archiving of endangered language oral traditions: Kaipuleohone at the University of Hawai‘i and C’ek’aedi Hwnax in Alaska. *Oral Tradition* 28(2). 261–270.

This article compares two small-scale digital language archives—Kaipuleohone at the University of Hawai‘i, and C’ek’aedi Hwnax, which serves the Ahtna Alaska Native community of south central Alaska—in terms of their relevance to oral history research. The former was developed primarily to fulfil the language data preservation needs of an academic department that is known for its linguistic fieldwork in the Asia-Pacific region, while the latter was developed in response to community concerns for the preservation of and access to records of their own linguistic heritage. Both were built according to best practices for digital endangered language preservation and both are members of OLAC, although the audiences they serve are quite different.

Berez, Andrea L., Taña Finnesand & Karen Linnell. 2012. C’ek’aedi Hwnax, the Ahtna Regional Linguistic and Ethnographic Archive. *Language Documentation & Conservation* 6. 237–252. <http://hdl.handle.net/10125/4538>.

This article details the development of C’ek’aedi Hwnax, the Ahtna Regional Linguistic and Ethnographic Archive in Copper Center, Alaska. C’ek’aedi Hwnax, founded in 2010, was the first OLAC-compliant, Indigenously administered digital language archive in North America. Discussed here are the history of Native Language archiving in the state of Alaska; the identification of the need within the Ahtna community to collect, preserve, and disseminate records of Ahtna language; and the establishment of the archive under the Ahtna Heritage Foundation, including funding, staffing, purchasing equipment, training, digitization, and policy development.

Berez, Andrea L. 2015. Reproducible research in descriptive linguistics: Integrating archiving and citation into the postgraduate curriculum at the University of Hawai‘i at Manoa. In Amanda Harris, Nick Thieberger & Linda Barwick (eds.), *Research, records and responsibility: Ten years of PARADISEC*, 39–51. Sydney: Sydney University Press.

The notion of reproducible research, in which researchers provide the dataset upon which scientific claims are based, is explored in the context of linguistics. As in other fieldwork-based sciences, true replicability is often not possible for linguistics, but reproducibility is often possible. The author discusses an initiative in the linguistics department at the University of Hawai'i to increase reproducibility by requiring PhD students to the archive primary data sets upon which dissertations are based, and then to cite back to that data in the text of the dissertation.

- Berez, Andrea & Gary Holton. 2006. Finding the locus of best practice: Technology training in an Alaskan language community. In Linda Barwick & Nicholas Thieberger (eds.), *Sustainable data from digital fieldwork*, 69–86. Sydney: University of Sydney Press.

The training component of the NSF-sponsored *Dena'ina Archiving, Training and Access* project included two types of training: 1) A three-week class during the summer of 2005 in basic language technology at the Dena'ina Language Institute in Soldotna, Alaska, which was designed for young members of the Dena'ina community; and 2) Four semesters of training in advanced multimedia technology applications to linguistics graduate students. While it had been expected that both learner groups would adapt easily to best practices for language data sustainability, it later became apparent that this expectation ignored community member expectations and interests for the role of technology in language revitalization.

- Bird, Steven & Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3). 57–582.

This landmark paper discusses seven problem areas, or *dimensions*, that potentially affect the portability of digital data in language documentation and description. These are content, format, discovery, access, citation, preservation, and rights. The authors propose value statements for the field of linguistics with regard to each of these dimensions in order to encourage discussion among linguists toward the development of best practices.

- Bow, Catherine, Michael Christie & Brian Devlin. 2015. Shoehorning complex metadata in the Living Archive of Aboriginal Languages. In Amanda Harris, Nick Thieberger & Linda Barwick (eds.), *Research, records and responsibility: Ten years of PARADISEC*, 115–131. Sydney: Sydney University Press.

The authors present an interesting case study that highlights complications with implementing best-practice approaches in archiving. Specifically, Bow et. al examine challenges involved when attempting to “shoehorn” complex and varied types of data into the standardized approach of an accessible digital archive. For example, the authors discuss conflicts between scientific nomenclature standards and the terms actually used

in language communities; problems trying to fit data into strict categorization protocols, such as when controlled vocabularies oversimplify the complexities of particular Aboriginal language materials; and difficulties determining which materials to include or exclude.

Bowden, John & John Hajek. 2006. When best practice isn't necessarily the best thing to do: Dealing with capacity limits in a developing country. In Linda Barwick & Nicholas Thieberger (eds.), *Sustainable data from digital fieldwork*, 45–56. Sydney: University of Sydney Press.

This one of many papers from the mid-to-late 2000s that questions the relevance of 'best practices' when working with endangered languages in developing countries. The authors examine the success of digital documentation workflows in the Waima'a speaking community of East Timor. The project trained and employed a local assistant in the full digital workflow, to great success, but the authors determined that in the end the archival resources are ultimately of little value to the Waima'a community, which favors instead traditional paper publications.

Boynton, Jessica, Steven Moran, Anthony Aristar & Helen Aristar-Dry. 2006. E-MELD and the School of Best Practices: An ongoing community effort. In Linda Barwick & Nicholas Thieberger (eds.), *Sustainable data from digital fieldwork*, 87–98. Sydney: University of Sydney Press.

This article outlines the development of the Electronic Metastructure for Endangered Languages (E-MELD) project in general, and the School of Best Practice website developed under E-MELD in particular. 'The School' was one component of the five year E-MELD project which was designed to instruct field linguists and anyone in possession of analog endangered language materials in the digitization and care of those items. The article discusses the various stages of development of The School, including identifying the need for such a resource; reaching the appropriate audience; and designing various instructional components like a showroom of case studies and a 'classroom' area with short articles on various topics.

Cameron, Deborah, Elizabeth Frazer, Penelope Harvey, M. B. H. Rampton, & Kay Richardson (eds.). 1992. *Researching language: Issues of power and method*. London: Routledge.

This book presents some of foundational work underlying participatory approaches to archiving. Cameron et al. define and delineate a model of "empowering research," which they describe as research undertaken *on, for, and with* language communities. This model contrasts with 'ethical' and 'advocate' research, both of which fail to incorporate fully interactive methods, the agendas of the people being researched, and a commitment to sharing the knowledge generated through research. In light of the conceptualization of an empowerment model, the editors present four case

studies from their own work to furnish comparative material for reflection upon power and methodology in linguistic research.

Chang, Debbie. 2010. *TAPS: Checklist for responsible archiving of digital language resources*. Dallas: Graduate Institute of Applied Linguistics MA thesis.

The TAPS (target, access, preservation and sustainability) checklist is developed as a metric to assist depositors in assessing the quality of archival practices when selecting a repository for digital endangered language materials. The checklist is then tested at nine digital archives. TAPS was developed for use by nonspecialists by selecting and comparing relevant components from other tools already in existence for assessing digital repositories. These tools are also discussed, although they are not necessarily geared to language repositories, and the author also reflects on the need to develop more formal tools for assessing language archives.

Conathan, Lisa. 2011. Archiving and language documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 235–254. Cambridge: Cambridge University Press.

Most linguists who regularly deposit their materials in an archive are only familiar with some aspects of the archiving workflow. This article presents the entire archiving process from the point of view of archival science, but with special attention to the needs of endangered language records. The stages in the workflow are appraisal and accession (assessing whether a collection is of enough value to warrant archiving, and the legal process by which an archive acquires materials for deposit), arrangement and description (the hierarchical grouping of materials and the use of metadata to provide information about the records for later finding), preservation (the long-term commitment to care for the physical form and intellectual content of the materials), and access and use (the mobilization of materials for educational and other purposes).

Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation & Conservation* 3(1). 15–50. <http://hdl.handle.net/10125/4423>.

This paper proposes a model for ethical linguistic fieldwork based on the author's experiences working in Canadian First Nations communities. The model, termed *community-based language research*, or CBLR, calls for research projects to be designed *for, with, and by* members of an endangered language community. In this model, linguists are full collaborative partners in the research, but they are not the primary agents of the research. The paper discusses other models of linguist-focused research and reflects on why one might choose to adopt the CBLR approach when working in Indigenous communities. The author also considers challenges that may arise in collaborative research programs.

Dobrin, Lise M. & Gary Holton. 2013. The documentation lives a life of its own: The temporal transformation of two endangered language archive projects. *Museum Anthropology Review* 7. 140–154.

Dobrin and Holton addresses a critical issue related to archiving, ethics, and access: The viewpoints and interests of a language community can change throughout the life of a project. Case studies explore Dobrin's Arapesh research in Papua New Guinea and Holton's work with Dena'ina in Alaska. In both cases, Indigenous communities became increasingly interested in documenting their own languages and interacting with extant collections of linguistic material held in digital archives. As such, the authors advise that documentary linguistics and archiving be approached as works in progress that are attuned to the wishes of language communities.

Doorn, Peter & Heiko Tjalsma. 2007. Introduction: Archiving research data. *Archival Science* 7(1). 1–20.

Coming from the discipline of archival science, this article introduces the concept of archiving research data (as opposed to archiving public records). Doorn and Tjalsma provide very useful information concerning the historical development of archives for research data as well as the advent and challenges of preserving digital information. In the latter half of the article, the authors survey the main issues and contemporary trends regarding demands on data archiving. This includes discussion of organizational infrastructures for data facilities, data strategies at national and international levels, issues of open access and data availability, and more.

Dwyer, Arienne M. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics Studies and Monographs 178), 31–66. Berlin: Mouton de Gruyter.

The first half of this chapter introduce basic ethical concepts related to language documentation (e.g., rights and responsibilities of fieldworkers and informed consent), and also legal aspects of data ownership and copyright. The second half is much more practical in nature, and offers a framework for ethical language documentation under the aegis of 'the five Cs': criteria, contacts, cold calls, community, and compensation. The value of this chapter is its clarity of presentation for those new to fieldwork and language documentation.

Evans, Nicholas & Hans-Jurgen Sasse. 2004. Searching for meaning in the Library of Babel: Field semantics and problems of digital archiving. In Linda Barwick, Allan Marett, Jane Simpson & Amanda Harris (eds.), *Researchers, communities, institutions and sound recordings*, 1–31. Sydney: University of Sydney.

The authors contribute to best-practice discussions by exploring challenges involving the archiving of semantic documentation. Evans and

Sasse assert that technological advancements have greatly expanded our abilities to collect and store sound recordings, but this has not necessarily been accompanied by parallel developments in capturing and conveying the *meaning* of these recording (e.g., explaining gestures, cultural context, or language-specific semantic relationships). The authors present case studies to illustrate the problem, and they advocate developing appropriate archiving technology—such as multi-layered annotations created over time and involving contributions from a variety of relevant parties—to facilitate the documentation of meaning.

Evans, Nicholas & Alan Dench. 2006. Introduction: Catching language. In Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing* (Trends in Linguistics Studies and Monographs 167), 1–39. Berlin: Mouton de Gruyter.

This is, first and foremost, the introduction to a volume about writing descriptive grammars, but Evans and Dench nonetheless engage ideas very relevant to archiving in documentary linguistics. For example, they discuss the progression of technology that has changed not only the kinds of linguistic data we collect but also how we interact with, store, and preserve this information. This includes the expectation that digital archives will be used increasingly for purposes such as testing linguistic analyses, but this entails significant implications for questions of access and data-stewardship best practices.

Gardiner, Gabrielle & Kirsten Thorpe. 2014. The Aboriginal and Torres Strait Islander Data Archive: Connecting communities and research data. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 103–119. London: SOAS.

Gardiner and Thorpe overview ATSIDA, a part of the Australian Data Archive that places an emphasis on collaboration and relationship building with researchers and language communities. The authors discuss the development, structure, and stakeholders of ATSIDA. They describe the archive's operations and furnish a look into the particulars of data curation and preservation as well as protocols designed to connect language communities with linguistic, cultural, and historical research data. Gardiner and Thorpe also explore the challenges and opportunities that have arisen during the establishment of ATSIDA, which should be valuable for anyone interested in participatory archiving.

Garrett, Edward. 2014. Participant-driven language archiving. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 68–84. London: SOAS.

In this article pertaining to participatory models of archiving, Garrett outlines the motivations and preliminary requirements for implementing what he calls *participant-driven language archiving* (PDLA). He claims

that existing archives have focused too much on building relationships solely with depositors, ignoring opportunities to involve the people who are the ‘human subjects’ of documentary linguistic research. In particular, Garrett explains that participants can enrich archived resources and address challenges of informed consent. The author explores some of the potentials and challenges of the PDLA model, including negotiating access, repatriating resources, and facilitating payment for language consultants.

Garrett, Andrew & Lisa Conathan. 2009. Archives, communities, and linguists: Negotiating access to language documentation. *Linguistic Society of America Annual Meeting*. http://www.ailla.utexas.org/site/lsa_olaco9/conathan-garrett_lsa_olaco9.pdf

Garrett & Conathan present several case studies from their own experiences to illustrate conflicts involving access to archived materials related to languages of California and the western United States. Such problems have hindered collaboration between archives, linguists, and heritage communities. Examples include failures to create access protocols, attempts by linguists or language communities to restrict access, and “turf disputes” between parties with stakes in archived materials. Garrett & Conathan review some archival protocols designed to help facilitate collaboration with communities while advocating for their rights, and they discuss lessons learned from these case studies.

Gehr, Susan. 2013. *Breath of Life: Revitalizing California’s native languages through archives*. San Jose: San Jose State University MA thesis.

This thesis is an oral history of the *Breath of Life* workshops held biennially since 1996 by the Advocates for Indigenous California Language Survival at the University of California, Berkeley. Gehr begins by surveying the history of Native American language revitalization efforts since the mid-twentieth century, with special focus on the role of archives and archived/archival material. She interviews participants, linguists, and archivists involved in the workshop and presents thoughts about future revitalization efforts.

Gerds, Donna. 2010. Beyond expertise: The role of the linguist in language revitalization programs. In Lenore A. Grenoble & N. Louanna Furbee (eds.), *Language Documentation: Practice and Values*, 173–192. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Based on her own experiences with the Halkomelem language, the author addresses the tension that can sometimes arise between members of an endangered language community and linguists in the context of language revitalization. She discusses the kinds of skills that linguists can bring to a revitalization project, and potential misunderstandings about linguists’ roles and abilities. She also presents her experiences of what Na-

tive language communities tend to want an academic linguist to provide, and what the needs of revitalization programs are.

- Gippert, Jost. 2006. Linguistic documentation and the encoding of textual materials. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics Studies and Monographs 178), 337–361. Berlin: Mouton de Gruyter.

The first half of this chapter discusses issues of character encoding, especially as it applies to presenting non-English (rather, non-ASCII) characters in textual materials. 8-bit to 32-bit encoding and Unicode are presented, along with some recommendations for avoiding character encoding problems (much of the discussion will be useful today, if one is in possession of older digital materials). The second half of the chapter discusses content-driven markup of textual structure, and proposes HTML as a potential way to get the benefits of true markup—XML—without too much trouble. XML is also discussed briefly.

- Golla, Victor. 1995. The records of American Indian linguistics. In Sydel Silverman & Nancy J. Parezo (eds.), *Preserving the anthropological record*, 143–157. New York: Wenner-Gren Foundation for Anthropological Research.

Golla's chapter summarizes vital information about the history of linguistic anthropology in North America, primarily since the late nineteenth century. He discusses the various types of records that have been created and collected by scholars, which includes lexical compilations, texts, file slips, sound and video recordings, and digital files. Golla also describes the history and collections of some of the most important archives preserving Native American linguistic material. The chapter concludes with a look at the challenges of preserving these records while properly training future generation of scholars to steward and study them.

- Good, Jeff. 2011. Data and language documentation. In Peter K. Austin & Julia Salabank (eds.), *The Cambridge handbook of endangered languages*, 212–234. Cambridge: Cambridge University Press.

Good discusses conceptual issues surrounding the nature of data in language documentation, which includes primary data as comprised of direct recordings of speech events and the transcriptions, or written representations, of those events. Primary data are contrasted with descriptive resources like texts, dictionaries, and grammars. The author also discusses the differences between data structure on the one hand, and implementation or presentation on the other. Also presented are the notions of proprietary versus open formats; markup; archival, working, and presentation formats; and metadata.

- Green, Jennifer, Gail Woods & Ben Foley. 2011. Looking at language: Appropriate design for sign language resources in remote Australian Indigenous communities. In

Nick Thieberger, Linda Barwick, Rosey Billington & Jill Vaughan (eds.), *Sustainable data from digital research: Humanities perspectives on digital scholarship*, 66–89. Melbourne: Custom Book Centre.

Sign languages are common in Arandic communities in Central Australia. These endangered languages are generally used by people who also use spoken language, and are culturally valued for use in certain rituals, and in situations like hunting and at times when audibility is disadvantageous. The authors describe a project to document, preserve, and promote Arandic sign through digital resource development. The project was designed to maintain respect for the dignity and desires of the communities by recording video in natural bush settings, by eliciting in local languages, and through careful editing. The authors also describe their data storage, annotation, and web publication procedures.

Hale, Ken, Michael Krauss, Lucille J. Watahomigie, Akira Y. Yamamoto, Colette Craig, LaVerne Masayesva Jeanne & Nora C. England. 1992. Endangered languages. *Language* 68(1). 1–42.

This collection of six essays appeared as a collection in the journal *Language* following a symposium at the 1991 Linguistic Society of America annual meeting. Hale's first essay introduces the collection and touches on language endangerment as the potential loss of cultural and intellectual diversity. Krauss's celebrated essay, described more fully below, is a call to arms for linguists to organize against language endangerment. Watahomigie and Yamamoto discuss reactions to language loss in Native America with particular emphasis on Hualapai in reference to both the American Indian Languages Development Institute and the Native American Languages Act. Craig discusses legislation from the 1980s in Nicaragua known as the Autonomy project under which several language planning projects were implemented for the Indigenous languages there; Craig focuses on the Rama Language Project and its successes. Jeanne proposes a Native American Language Center, which would be dedicated to a range of support and research activities for Native American languages, and staffed by and serving the concerns of speakers of Native American languages. England reflects on the role of Mayan language scholarship in Guatemala. Hale's second essay considers more deeply the value of linguistic diversity to humanity.

Himmelman, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–95.

In this, the definitive article now commonly cited as launching the subfield of language documentation as distinct from descriptive linguistics, the author describes the activities of language documentation as the creation of “a record of the linguistic practices and traditions of a speech community” (166). Practical and theoretical considerations are presented for the

four steps of language documentation: 1) decisions about which data to collect; 2) recording the data; 3) annotation, the transcription and translation of the data with commentary; and 4) preservation and presentation. Also discussed are ethical and privacy considerations, as well as guidelines for collecting a documentation that is varied in genre and spontaneity.

Himmelman, Nikolaus. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelman & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics Studies and Monographs 178), 1–30. Berlin: Mouton de Gruyter.

This is the introductory chapter to the first edited volume on language documentation proper. Eight years after the publication of Himmelman 1998, the author further refines this field of linguistic inquiry, and defines a language documentation as “a lasting, multipurpose record of a language” (1). He also discusses the value of language documentation to other disciplines both inside and outside of linguistics, and presents a format for a documentation. This format includes records of observable linguistic behavior; indications of metalinguistic knowledge including paradigms, usage scenarios, and other generalizations; lexical databases; and the apparatus. The *apparatus* is defined as the set of information that is used to interpret and understand the rest of the documentation, including metadata, transcriptions, translations, ethnographic sketches, glossing conventions, and the like.

Hinton, Leanne. 2001. Language revitalization: An overview. In Leanne Hinton & Kenneth Hale (eds.), *The green book of language revitalization in practice*, 3–18. San Diego: Academic Press.

In this first chapter of a guide to language revitalization, Hinton surveys language shift and endangerment as well as various approaches to revitalization. This includes discussion of the role of archives in revitalization. For instance, archives play a vital part at the starting point of revitalization efforts, when communities seek out existing material on their languages. Archived materials also serve as critical resources for the creation of language-teaching materials, such as reference grammars and language lessons. Accordingly, Hinton discusses programs like Breath of Life, which aim to increase access to archives for Indigenous communities.

Hinton, Leanne. 2005. What to preserve: A viewpoint from linguistics. In Administration for Native Americans (ed.), *Native language preservation: A reference guide for establishing archives and repositories*, 24–26. Washington, D.C.

This is a very brief selection from a guidebook for Indigenous communities about archival matters related to their languages (see ANA 2005 above). Nonetheless, Hinton touches upon several important themes and issues: Indigenous communities are increasingly enlisting archives in the

service of language maintenance and revitalization, particularly in the creation of dictionaries, curricula, and the like; archived language materials often lack crucial metadata, such as detailed annotations and transcriptions; and speakers and collectors must determine together the access conditions for their archived data.

Holton, Gary. 2012. Language archives: They're not just for linguists any more. In Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts & Paul Trilsbeek (eds.), *Language Documentation & Conservation Special Publication No. 3, Potentials of Language Documentation: Methods, Analyses, and Utilization*, 111–117. Honolulu: University of Hawai'i Press. <https://scholarpace.manoa.hawaii.edu/handle/10125/4523>.

In this short chapter, Holton provides an insightful look at how language archives are actually used. He draws upon his experience at ANLA to present examples demonstrating that the audiences and uses of an archive can go far beyond the founding aims of linguists simply preserving language data. Holton describes, for example, an ethnoastronomy project relying upon ANLA's archived sources. He also discusses community efforts to revitalize Eyak, where ANLA is the only surviving source of information about the language. Thus, Holton advises archives to facilitate non-linguistic uses for their materials and to position linguistic data to create derived products in the service of language revitalization.

Holton, Gary. 2014. Mediating language documentation. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 37–52. London: SOAS.

A recurring thread in best-practice discussions concerns negotiating and facilitating access to archived materials, but Holton calls attention to a critical point: Providing access alone is not enough to ensure that such materials are actually used. This problem is particularly significant when language maintenance and revitalization efforts are involved. As such, this article proposes that archives must *mediate* between collections and users. Using his experiences at ANLA as a case study, Holton suggests how archives can make their materials more accessible and more relevant to language communities, which requires that archives work closely with the people they aim to serve.

Holton, Gary, Andrea L. Berez, & Sadie Williams. 2006. Building the Dena'ina language archive. In Laurel Evelyn Dyson, Max Hendricks, & Stephen Grant (eds.), *Information technology and indigenous people*, 205–209. Hershey: Idea Group.

This paper discusses the development of the Dena'ina Language Archive, a digital archiving project created under the aegis of the NSF-sponsored Dena'ina Archiving, Training, and Access project. Dena'ina is an Athabascan language spoken in south central Alaska, and under this project the

Dena'ina language materials in ANLA were digitized and made available online. Metadata were made discoverable through OLAC and were embedded in a value-added online portal known as qenaga.org (*qenaga* means 'language' in Dena'ina). The project represented an early digital collaboration between linguists, language technologists, and community members in an Alaska Native language.

Huvila, Isto. 2008. Participatory archive: Towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Archival Science* 8(1). 15–36.

Building upon the groundwork laid by Shilton and Srinivasan (2007), Huvila explicitly formulates the concept of a “participatory archive.” He describes the development of this idea through a case study of two projects building digital historical archives in Finland. The three defining characteristics of a participatory archive are: 1) *decentralized curation*, 2) *radical user orientation*, and 3) *contextualization of both records and the entire archival process*. This model radically reconfigures the responsibilities of and interactions between archivists, depositors, and users throughout the archival process.

Innes, Pamela. 2010. Ethical problems in archival research: Beyond accessibility. *Language & Communication* 30(3). 198–203.

Innes offers a brief-but-significant exploration of ethical considerations in archiving. This article relates her experiences working to prepare for publication Mary Haas' archived notes on Mvskoke. Innes encounters a major problem: Some members of the language community felt that particular narratives were inappropriate for certain audiences, and that other texts were even dangerous. This case study raises critical issues of obtaining and documenting informed consent, managing access to archived materials, and navigating tensions between the language ideologies of a community and those of scholars who expect data to be open and available.

Innes, Pamela & Erin Debenport. 2010. Editors' introduction. *Language & Communication* 30(3). 159–161.

Although this is but a short introduction to an entire journal issue devoted to ethics and language documentation, it is worth reading to hear from the editors themselves about what motivated the production of such a volume: Documentary linguistics had spent plenty of time and resources developing “best practices” for many of the technological and archival aspects of documentation, but the same dedication had not been committed to exploring the ethical implications of these aspects.

Johnson, Heidi. 2004. Language documentation and archiving, or how to build a better corpus. In Peter K. Austin (ed.), *Language Documentation and Description Volume 2*, 140–153. London: SOAS.

Johnson's article is a must-read primer for understanding the relationship between archiving and language documentation. She offers an informative review of the role of archiving in early and modern documentary linguistics, along with a description of the progress of technology used in such endeavors. For anyone looking for a quick guide on where archiving fits into documentary linguistics, Johnson provides a breakdown explaining "who should archive, and where, why, when, and how one should archive" (3). The bulk of this article covers the ethos and best-practice methodology of archiving language documentation, spanning topics such as data formats, access permissions, item labelling, and metadata.

Krauss, Michael E. 1974. Alaska Native language legislation. *International Journal of American Linguistics* 40(2). 150–152.

This brief describes the 1972 passing of four bills in the Alaska State Legislature concerning Alaska Native Languages. Senate Bill 421 authorized mandatory bilingual education in state schools where students speak a Native language; Senate Bill 422 authorized the establishment of the Alaska Native Language Center at the University of Alaska; Senate Bills 424 and 423 appropriated funds to the other two bills respectively. The text of all four bills are presented.

Krauss, Michael. 1992. The world's languages in crisis. *Language* 68. 4–10.

The most-cited of the essays edited by Hale and appearing together in *Language* (1992), this piece starts by citing some sobering figures about language vitality in North America and beyond. Krauss proposes a cline of statuses for vitality including "endangered," "moribund," and "safe." Endangered languages are compared to endangered species, and the author draws parallels about the expected reaction of the scientific community in face of endangerment. The essay ends with the admonishment that linguistics not "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" (10).

Krebs, Allison Boucher. 2012. Native America's twenty-first-century right to know. *Archival Science* 12. 173–190.

This article provides valuable historical and cultural context related to the increasing self-empowerment of Indigenous people in the United States over the course of the last several decades. Krebs evaluates two initiatives supporting the development of libraries, archives, and information centers for Indigenous communities: 1) the Institute of Museum and Library Services' Grants to Indian Tribes, and 2) the Fourth Museum of the National Museum of the American Indian. Of particular value here is the overview of activist Vine Deloria Jr.'s advocacy for an Indigenous 'right to know,' along with Krebs' timeline, which breaks down relevant developments re-

garding the relevant interplay between federal, citizen, and professional organizations.

Linn, Mary S. 2014. Living archives: A community-based language archive model. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 53–67. London: SOAS.

Linn outlines a proposal for a Community-Based Language Archive (CBLA), a radical departure from traditional models of archiving. In a CBLA, the archive engages with a language community throughout every component of the archiving process. Along with explaining the concept, Linn provides a case study of her experiences integrating the CBLA model while transforming collections and building new ones at the Sam Noble Oklahoma Museum of Natural History. This article also includes a useful overview of literature exploring participatory and community-based approaches to archiving and language research.

Linguistic Society of America. 2010. *Resolution recognizing the scholarly merit of language documentation*. <http://www.linguisticsociety.org/resource/resolution-recognizing-scholarly-merit-language-documentation>

This resolution, passed in 2010 by ‘a sense of majority’ within the Linguistic Society of America, declares the outputs of language documentation for scholarly and community use—including dictionaries, grammars, text collections, digital data sets, web products, and more—to be considered academic output for the purposes of hiring, tenure, and promotion.

Macri, Martha & James Sarmiento. 2010. Respecting privacy: Ethical and pragmatic considerations. *Language & Communication* 30(3). 192–197.

Macri & Sarmiento provide a helpful, brief case study that illustrates ethical problems involved in archiving sensitive materials. This article details issues encountered by researchers transcribing and coding notes in the J. P. Harrington Database Project, which aims to create resources for use by a variety of academic and non-academic audiences. In particular, notes have involved gossip and hearsay, sensitive customs, sacred sites, and even potentially physically dangerous knowledge. Macri and Sarmiento raise important questions about conflicts between international standards and Indigenous communities, and deciding who—if anyone—can speak for a community.

Nathan, David. 2009. The soundness of documentation: Towards an epistemology for audio in documentary linguistics. *Journal of the International Association of Sound Archives* 33. 50–63.

A critique of so-called ‘best practices’ in language documentation that encourage the use of ever-advancing technologies without truly understanding the goals and impacts of audio recording, this article encourages *critical listening* when making recordings. One aspect of this includes giving

serious consideration to signal-to-noise ratio: Determining what counts as signal and what counts as noise should be guided by the aims of the documentation project. Another aspect is the consideration of psychoacoustic effects of capturing spatial information through advanced stereo techniques like ORTF. It is argued that critical listening will produce better documentation than carelessly adopting the latest advancements in media like video.

Nathan, David. 2010. Archives 2.0 for endangered languages: From disk space to MySpace. *International Journal of Humanities and Arts Computing* 4(1–2). 111–124.

Nathan describes how ELAR has attempted to implement the properties of Web 2.0 (e.g., social networking and interaction online) in order to restructure and enhance the experiences of its depositors and users. This moves the archive beyond a traditional role as a data repository. Instead, ELAR now aims to facilitate relationships between parties involved in archiving. Nathan argues that this approach is better equipped for managing issues of access (especially sensitivities and restrictions) as well as the diversity of resources held by ELAR.

Nathan, David. 2011. Digital archiving. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 255–273. Cambridge: Cambridge University Press.

In some sense this handbook chapter is a companion to Conathan 2011, in that it addresses specifically the digital aspects of archiving within the larger framework of archive curation. The author discusses the nature of digital data and digital encoding; several sections are dedicated to describing extant digital archives, their services, and their policies; and the author ends by touching on data migration, the archiving of video, and archive assessment.

Nathan, David. 2014. Access and accessibility at ELAR, an archive for endangered languages documentation. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 187–208. London: SOAS.

This article illustrates a shift *in practice* toward a participatory model for ELAR, one of the most important archives involved in documentary linguistics. Nathan describes how ELAR has integrated a social networking approach to reconfigure the way the archive interacts with—and facilitates interactions between—its depositors and users. This, of course, is a departure from the traditional ‘one-way street’ model of archiving. He walks the reader through the ELAR protocol for navigating resources as well as searching and browsing, and he explains how this approach enhances access for various types of users.

Nathan, David. 2015. On the reach of digital language archives. In Amanda Harris, Nick Thieberger & Linda Barwick (eds.), *Research, records and responsibility: Ten years of PARADISEC*, 53–79. Sydney: Sydney University Press.

The author discusses the concept of *reach* as a measurement of the capacity of an archive to provide materials to the appropriate audience. Ten facets of reach are defined: acquisition, audiences, discovery, delivery, access management, information accessibility, promotion, communication ecology, feedback channels, and temporal reach.

Nathan, David & Peter K. Austin. 2004. Reconceiving metadata: Language documentation through thick and thin. In Peter K. Austin (ed.), *Language Documentation and Description, Volume 2*, 179–187. London: SOAS.

In this critical addition to best-practices discussions, Nathan and Austin put forth a distinction between ‘thin’ and ‘thick’ metadata. They argue that most attention in documentary linguistics goes toward the former, which does not provide enough value for linguists and speech communities interested in working with language materials. *Thin* metadata is primarily for cataloguing, mostly aimed at facilitating resource discovery. On the other hand, *thick* metadata involves more context—such as transcriptions, commentary, and time-aligned annotations—and is intended to enhance the access and use of archived materials.

Nathan, David & Peter K. Austin. 2014. Editors’ introduction. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 4–16. London: SOAS.

This is the introduction to “the first journal publication symmetrically targeted at both language documentation and archiving” (6). As such, it presents a helpful overview of the papers inside the publication. However, this chapter also offers value in its own right. In particular, Nathan and Austin furnish a useful glance at the relationship between archiving and language documentation. They also point out issues that recur throughout their volume: community curation, the promotion of archived language resources, the contextualization of archived materials, the ‘form’ of documented material (e.g., structure and granularity), and the conceptualization of archiving as a publishing.

Nordhoff, Sebastian & Harald Hammarström. 2014. Archiving grammatical descriptions. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 164–186. London: SOAS.

Much of the best-practices talk in archiving has revolved around primary data, and so Nordhoff and Hammarström call attention to the need for a methodology of archiving *grammatical descriptions*. Grammatical descriptions are based on primary data but entail different information types

and structures, and their users have specific needs for retrieving information at certain levels of granularity. Given these differences, the authors recommend a semantic-markup architecture based upon the Text Encoding Initiative (TEI). They present a systematic appraisal of existing TEI schema as well as special TEI elements, which could facilitate the archiving and access of grammatical descriptions.

O'Meara, Carolyn & Jeff Good. 2010. Ethical issues in legacy language resources. *Language & Communication* 30(3). 162–170.

This article offers a critical contribution to best-practice recommendations in archiving. O'Meara and Good examine the pilot phase of the Northeastern North American Indigenous Languages Archive to probe vital ethical issues surrounding the establishment of rights and access to archived language resources. In particular, the authors raise questions related to four areas: 1) the notion of 'community,' 2) establishing rights and access retroactively, 3) establishing rights and access to resources without an identifiable copyright holder, and 4) navigating concerns associated with sensitive materials.

Ormond-Parker, Lyndon & Robyn Sloggett. 2012. Local archives and community collecting in the digital age. *Archival Science* 12. 191–212.

Ormond-Parker and Sloggett focus on Aboriginal communities in Australia to take an important look at the increasing self-empowerment of Indigenous people in archiving. This, of course, has been fueled in part by the proliferation of digital tools and technology. The authors identify the benefits of such developments for these communities, which include economic development, community empowerment, and the creation of opportunities for young people. At the same time, however, Ormond-Parker and Sloggett argue that community-driven efforts are often not equipped to handle the various threats inherent to digital archiving. As a solution, the authors recommend a national framework to support community-controlled archives.

Rehg, Kenneth L. 2007. The Language Documentation and Conservation Initiative at the University of Hawai'i at Mānoa. In D. Victoria Rau & Margaret Florey (eds.), *Language Documentation and Conservation, Special Publication No. 1, Documenting and Revitalizing Austronesian Languages*, 13–24. Honolulu: University of Hawaii Press. <http://hdl.handle.net/10125/135>.

Although it focuses on one initiative at a single university, Rehg's piece is a useful treatment about putting into practice some of the most crucial themes from the history of archiving in linguistics. This includes best-practices training for linguists in the theory, methods, and ethics of language documentation. Rehg also describes efforts to create collaborative research models that benefit linguists and non-linguists alike. As such, he outlines then-developing plans to create a digital archive at the University

of Hawai‘i, one that safely stores data in accordance with the desires of speech communities. This archive, named Kaipuleohone, opened in 2008.

Robinson, Laura. 2006. Archiving directly from the field. In Linda Barwick & Nicholas Thieberger (eds.), *Sustainable data from digital fieldwork*, 23–32. Sydney: University of Sydney Press.

Depositing materials into an archive on a regular basis has not always been part of the linguist’s workflow, so this author discusses her own procedures for developing a regular archiving practice while on a year-long fieldwork trip to the Philippines. She describes her solar power configuration, her digitization workflow, and her metadata documentation workflow. She sent her data regularly to PARADISEC via the postal service during this period. Although archiving from the field has become *de rigeur* since this article was written, it is important to remember that this was not always common practice.

Robinson, Laura. 2010. Informed consent among analog people in a digital world. *Language & Communication* 30. 186–191.

The ethical bind that comes with obtaining informed consent about digital dissemination of language data from people with no knowledge of the internet is discussed in the context of the author’s fieldwork with a remote community of Agta speakers in the Philippines. Institutional review boards will often allow oral, as opposed to written, consent in cases of non-literate consultants, but the author argues that because researchers have a moral obligation for informed consent, consultants with no knowledge of the internet could be considered a vulnerable class when the researcher wants to disseminate data online. The two solutions available—nondissemination of that data versus assuming speakers would want their data to be disseminated online “if they only understood”—are presented as equally paternalistic.

Schroeter, Ronald & Nick Thieberger. 2006. EOPAS, The EthnoER online representation of interlinear text. In Linda Barwick & Nicholas Thieberger (eds.), *Sustainable data from digital fieldwork*, 99–124. Sydney: University of Sydney Press.

The authors describe the initial development phase of EOPAS, a tool designed to convert the normal outputs of a digital language documentation workflow into presentation formats suitable for online viewing. The tool primarily works with time aligned transcripts (e.g., those from ELAN and Transcriber) and interlinear text (e.g., Toolbox). EOPAS transforms the validated XML output of those other tools into EOPAS XML via stylesheets. The resultant file is then stored alongside the original media file for display; at the time, a tool known as Annodex was being explored as a streaming delivery option, and other HTML displays were also developed.

Schwartz, Gabriele. 2012. Online presentation and accessibility of endangered languages data: The general portal to the DOBES Archive. In Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts & Paul Trilsbeek (eds.), *Language Documentation & Conservation Special Publication No. 3, Potentials of Language Documentation: Methods, Analyses, and Utilization*, 126–128. Honolulu: University of Hawaii Press. <http://hdl.handle.net/10125/4526>.

This very brief chapter belongs to conversations about expanding the audiences and uses of archives. As one of the primary funders of endangered language documentation work, DOBES maintains a large archival collection of data from its projects. In order to expand the archive's user base and increase access to materials, DOBES launched a general web portal in March 2013. With a bare-bones approach, Schwartz walks through the structure and features of the portal, describing how it aims to serve researchers, depositors, language communities, and the general public.

Shilton, Katie & Ramesh Srinivasan. 2007. Participatory appraisal and arrangement for multicultural archival collections. *Archivaria* 63. 87–101.

Shilton & Srinivasan offer perhaps the first contribution to the discussion around participatory models in archival sciences. As institutions creating collective memory, archives often fail to include different ethnic and cultural communities in the foundational archival practices of appraisal, arrangement, and description. This contributes to imbalances in power and representation for historically marginalized people. As such, Shilton and Srinivasan recommend 'rearticulating' appraisal and arrangement as community-driven, participatory processes. In doing so, a participatory model can improve the quality of archives, preserve more local knowledge and context, and help empower people traditionally left out of the archiving process.

Stenzel, Kristine. 2014. The pleasures and pitfalls of a "participatory" documentation project: An experience in northwestern Amazonia. *Language Documentation & Conservation* 8. 287–306. <http://hdl.handle.net/10125/24608>.

Stenzel presents her experiences documenting languages in the Amazon, providing a critical response in the ongoing discourse around collaborative and participatory research models in documentary linguistics. The piece is primarily a narrative history of Stenzel's four-year project, with perhaps the most valuable contribution coming from her discussion of the various 'pitfalls' she encountered. This includes a host of "logistical, technical, cultural, and philosophical" challenges, which all have a bearing on important issues like project sustainability, accountability, and the complex human relationships that provide the underpinnings for collaborative projects.

Theimer, Kate. 2011. Exploring the participatory archives: What, who, where, and why. *Annual Meetings of the Society of American Archivists*. <http://www.slideshare.net/ktheimer/theimer-participatory-archives-saa-2011>.

Although this is a brief conference presentation, Theimer's contribution is another good example of conversations in archival sciences about participatory models of archiving, which had been taking place for several years before penetrating the field of linguistics. Theimer helpfully introduces her concept and definition of 'participatory archiving,' which entails contributing knowledge and resources in a (typically) online environment. Moreover, she outlines a distinction between engagement and participation. This is a slideshow rather than an article, so this piece is best considered together with a paper like Shilton and Srinivasan 2007 or Huvila 2008.

Thieberger, Nicholas. 1994. Report on the AIATSIS visiting research fellowship, Aboriginal Studies Electronic Data Archive: A report to AIATSIS Council on the conclusion of the Visiting Research Fellowship. <http://trove.nla.gov.au/work/33785959?q&versionId=41559386>.

This report contains a summary of the structure and operations of the Aboriginal Studies Electronic Data Archive, which was established in 1991 and is now integrated with AIATSIS. This piece also describes various projects undertaken by the archive, including the AIATSIS Aboriginal Dictionaries Project, a workshop on copyright, and more. The value of this report primarily lies in its historical information and thorough accounting of the activities of what might be the first digital archive dedicated to endangered languages.

Thieberger, Nicholas. 2010. Anxious respect for linguistic data: The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) and the Resource Network for Linguistic Diversity (RNLD). In Margaret Florey (ed.), *Endangered Languages of Austronesia*, 141–158. Oxford: Oxford University Press.

This chapter is a prime example of best-practice discussions in linguistic archiving: Thieberger presents a thorough walkthrough of recommended methods for creating and storing language documentation data. He draws upon his own experience documenting the Oceanic language South Efate and working with PARADISEC to provide specific advice for proper data management and workflows, making data locatable and citable, choosing file formats and software tools, and more. Additionally, this chapter discusses the operations of PARADISEC and stresses the importance of training academics and speaker communities to employ best-practice methods in the documentation of endangered languages.

Thieberger, Nicholas. 2012. Using language documentation data in a broader context. In Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna

Margetts, & Paul Trilsbeek (eds.), *Language Documentation & Conservation Special Publication No. 3, Potentials of Language Documentation: Methods, Analyses, and Utilization*, 129–134. Honolulu: University of Hawaii Press. <http://hdl.handle.net/10125/4527>.

In this short chapter, Thieberger provides critical commentary related to making language documentation data as long-lasting, accessible, and useful as possible. Topics include creating data that can be reused and migrated to different formats and media to survive for generations; providing proper methods training in documentation and data management for academic and speech communities; encouraging repositories to conform to accepted data management and curation standards; meeting the evolving needs of users in an increasingly social media-oriented environment; and, of course, creating incentives for parties involved to follow best practices.

Thieberger, Nicholas. 2013. Curation of oral tradition from legacy recordings: An Australian example. *Oral Tradition* 28(2). 253–260.

This piece is a brief introduction to PARADISEC, aimed at an interdisciplinary audience interested in the world's oral traditions. Thieberger summarizes the mission, history, and operations of PARADISEC. Discussion includes the technical features of the archive, annotations and transcriptions, and trainings offered by PARADISEC. Thieberger also describes how interested researchers can use the archive to access online recordings and their accompanying analyses.

Thieberger, Nicholas & Linda Barwick. 2012. Keeping records of language diversity in Melanesia: The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). In Nicholas Evans & Marian Klamer (eds.), *Language Documentation & Conservation Special Publication No. 5, Melanesian Languages on the Edge of Asia: Challenges for the 21st Century*, 239–253. Honolulu: University of Hawaii Press. <http://hdl.handle.net/10125/4567>.

Thieberger & Barwick present an overview of the context behind the creation of PARADISEC and a summary of how the archive operates. PARADISEC is a cutting-edge digital repository for recordings primarily from the region around Australia (but open to materials from around the world), and aims to make such materials available to researchers and communities. Founded in 2003, the archive has long been a best-practices leader, being designed specifically to interoperate with researcher workflows, accommodate the domains and standards of different disciplines, and consider ongoing ethical and technological developments.

Thieberger, Nicholas & Andrea L. Berez. 2011. Linguistic data management. In Nicholas Thieberger (ed.), *The Oxford handbook of linguistic fieldwork*, 90–118. Oxford: Oxford University Press.

This article is a guide to managing digital workflows for language documentation both in and out of the fieldwork setting. Good data management in a documentation project is likened to building a house: When the foundation is solid, the house is long-lasting and extensible. The article discusses a wide range of topics of interest to the documentary linguist who is preparing to develop procedures for managing digital data, including the difference between data and metadata; the distinction between form and content (e.g., form-driven markup versus content-driven markup); and a workflow for well-formed linguistic data from field to archive to presentation. The authors offer suggestions for planning for data management well in advance of fieldwork, including planning for archiving and developing procedures for consistent file naming and data backup. Finally, the paper discusses the principles behind a relational metadata database, the value of regular expressions in data manipulation, and creating well-structured time-aligned interlinear glossed texts.

Thieberger, Nicholas, & Simon Musgrave. 2007. Documentary linguistics and ethical issues. In Peter K. Austin (ed.), *Language Documentation and Description, Volume 4*, 26–37. London: SOAS.

This article discusses vital ethical concerns that have arisen in linguistics due to developments in technology and modern language documentation. Thieberger and Musgrave focus primarily on informed consent and data ownership and rights. For example, researchers must grapple with the fact that language documentation is more intrusive than traditional descriptive data collection, and documentary linguists cannot predict all future uses for their data. Moreover, archives have become central to language documentation, which introduces a third party that must be taken into account when constructing consent. The authors also address issues regarding the ownership of language data and the products derived from them.

Thieberger, Nick, Amanda Harris, & Linda Barwick. 2015a. PARADISEC: Its history and future. In Amanda Harris, Nick Thieberger & Linda Barwick (eds.), *Research, records and responsibility: Ten years of PARADISEC*, 1–15. Sydney: Sydney University Press.

The introductory chapter in a volume to commemorate the tenth anniversary of the founding of PARADISEC, this piece describes the founding of the archive in 2002 and reflects on its evolution over the following decade. At the time of writing, the archive houses some 94,500 files on 860 distinct languages worldwide. Technical specifications are described, including the development of Nabu, the archive's catalog software. The authors also provide examples of academic and community uses of PARADISEC collections over the years. PARADISEC now rates five stars on the Open Language Archive Community metric and holds the European Data Seal of Approval.

Thieberger, Nick, Anna Margetts, Stephen Morey, & Simon Musgrave. 2015b. Assessing annotated corpora as research output. *Australian Journal of Linguistics* 36. 1–21.

This paper represents an important step in the valuation of documentary linguistics corpora as scholarly output. The authors explore options for valuing corpora in the Australian research context, although they note that these discussions can and should take place in other countries as well. Options considered include publishing corpus reviews, which would be similar to book reviews; and a publication or ‘journal’ model, in which corpora are ‘published’ in a serial publication. The authors propose a peer review process for corpora that is similar to the peer review process of traditional publications, under the auspices of the Australian Linguistics Society, and they include discussion of parameters for assessing the accessibility and quality of corpora.

Trilsbeek, Paul & Alexander König. 2014. Increasing the future usage of endangered language archives. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 151–163. London: SOAS.

Trilsbeek & König approach crucial issues of using existing infrastructures to expand the usage and audiences of digital archives that preserve endangered language materials. This includes discussion of acquiring additional materials by facilitating and increasing contributions from language communities; integrating with existing large-scale e-infrastructures to furnish users with access to more data and research tools; and making endangered language data more available to researchers in disciplines other than linguistics by finding means to enrich metadata and provide useful annotations, transcriptions, and translations.

Trilsbeek, Paul & Peter Wittenburg. 2006. Archiving challenges. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics Studies and Monographs 178), 311–336. Berlin: Mouton de Gruyter.

This article surveys the challenges of digital archiving by assessing the ‘three key players’ involved: depositors, users, and archivists. Each places different demands upon the archive, and a given key player has motivations, goals, and preferences that differ from those of the others. Trilsbeek and Wittenburg review these demands and the conflicts they create, and they discuss interactions between an archive’s key players. The article also examines conflicts generated by an archive’s need to preserve data for the long term while meeting the short-term needs of various user groups. Finally, it offers a valuable look at legal and ethical issues of access and managing access to archived materials.

Wilbur, Joshua. 2014. Archiving for the community: Engaging local archives in language documentation projects. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 85–101. London: SOAS.

Wilbur describes his experiences with the Pite Saami Documentation Project working with local archival institutions to improve access to language materials for speech communities. Modern archiving of language documentation materials is primarily digital, online, and aimed at a global audience. However, Wilbur notes that this can create barriers for many communities interested in accessing information about their own language and culture. Such barriers include a lack of requisite technological infrastructure or computer and language skills. Wilbur presents a case study to illustrate the benefits and challenges of working with national, regional, and municipal institutions to overcome these barriers.

Woodbury, Tony. 2003. Defining documentary linguistics. In Peter Austin (ed.), *Language Documentation and Description Volume 1*, 35–51. London: SOAS.

In this edited version of a plenary address from the 2003 annual meeting of the Linguistic Society of America, Woodbury provides an overview of the relatively new field of language documentation. The motivations for documentation include changes in technology, an increased interest in linguistic and social diversity, and, of course, the language endangerment crisis. The author notes that one of the defining characteristics of the field as distinct from other areas of inquiry is the discourse-centered approach of documentation, wherein attention to naturally occurring speech takes a place of importance alongside more traditional endeavors like language description. The author also addresses the need for a theorization of language documentation, and he discusses specific projects in Alaska and Peru.

Woodbury, Anthony. 2011. Language documentation. In Peter K. Austin & Julia Salabank (eds.), *The Cambridge handbook of endangered languages*, 159–211. Cambridge: Cambridge University Press.

Woodbury's chapter is dedicated to defining language documentation in a handbook on endangered languages more generally. He traces the development of the field as having its roots in the Americanist tradition, especially the ethnographically rich fieldwork of Franz Boas. Boas' practices and values then transferred via his student Sapir to structural era scholars including Emeneau and Haas, then to Krauss, and even to Gumperz in the 'ethnography of speaking.' The author also discusses the relationship between documentation and community-based language work and values, making the point that good documentation can be widely useful in practical and emblematic ways in language revitalization programs.

Woodbury, Anthony C. 2014. Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. In David Nathan & Peter K. Austin (eds.), *Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving*, 19–36. London: SOAS.

The author provides advice to language documenters, archivists, and audiences for improving the frequency and purpose of usage of archival collections. Documentary linguists can make their collections more valuable by creating corpus guides, including good descriptions of the documentation project activities, and sharing fieldwork journals. Archivists can increase usage by making collections easily discoverable and accessible; asking depositors to create collection guides (or creating one when the depositor is no longer available); and following practices undertaken by art museums, including guest curators and ‘exhibits.’ Audiences (e.g., journal editors) can increase the value of collections by encouraging reviews of archival collections.

Yamada, Racquel-María. 2007. Collaborative linguistic fieldwork: Practical application of the empowerment model. *Language Documentation & Conservation* 1(2). 257–282. <http://hdl.handle.net/10125/1717>.

Yamada presents a case study of linguistic fieldwork designed to meet the needs of both academic and speech communities. Linguists working to document endangered languages can struggle to achieve their own professional and academic goals while balancing the needs and desires of the communities with which they work. Yamada provides examples from her own work with speakers of the Cariban language Kari’ña to illustrate a model of collaborative, community-based linguistic research. She describes several projects, including the creation of pedagogical materials, collaborative linguistic analysis, and the repatriation of previous language recordings.

Ryan E. Henke
rhenke@hawaii.edu

Andrea L. Berez-Kroeker
andrea.berez@hawaii.edu