



PAM Pokrajinski
arhiv
Maribor

Moderna
arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 3 (2020), št. 2 / Year 3 (2020), No. 2

Maribor, 2020

Pokrajinski arhiv Maribor

Moderna arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 3 (2020), št. 2 / Year 3 (2020), No. 2

Maribor, 2020

Moderna arhivistika

Časopis arhivske teorije in prakse

Journal of Archival Theory and Practice

Letnik 3 (2020), št. 2 / Year 3 (2020), No. 2

ISSN 2591-0884 (online)

ISSN 2591-0876 (CD_ROM)

Izdaja / Published by:

Pokrajinski arhiv Maribor / Regional Archives Maribor

Glavni in odgovorni urednik / Chief and Responsible editor:

*Ivan Fras, prof., Pokrajinski arhiv Maribor, Glavni trg 7, SI-2000 Maribor,
telefon/ Phone: +386 2228 5017; e-pošta/e-mail: ivan.fras@pokarh-mb.si*

Glavna urednica / Editor in chief:

mag. Nina Gostenčnik

Uredniški odbor / editorial board:

- dr. Thomas Aigner, Diözesanarchiv St. Pölten, Avstrija
- dr. Borut Batagelj, Zgodovinski arhiv Celje, Slovenija
- dr. Bojan Cvelfar, Arhiv Republike Slovenije, Slovenija
- mag. Nada Čibej, Pokrajinski arhiv Koper, Slovenija
- Ivan Fras, Pokrajinski arhiv Maribor, Slovenija
- mag. Nina Gostenčnik, Pokrajinski arhiv Maribor, Slovenija
- dr. Joachim Kemper, Institut für Stadtgeschichte Frankfurt am Main, Nemčija
- Leopold Mikec Avberšek, Pokrajinski arhiv Maribor, Slovenija
- dr. Miroslav Novak, Pokrajinski arhiv Maribor, Slovenija
- dr. Rik Opsommer, Stadsarchief Ieper - Universiteit Gent, Belgija
- Darko Rubčić, Državni arhiv u Zagrebu, Hrvaška
- dr. Izet Šabotić, Filozofski fakultet Univerziteta u Tuzli, Bosna in Hercegovina
- mag. Boštjan Zajšek, Pokrajinski arhiv Maribor, Slovenija

Recenziranje / Peer review process:

Prispevki so recenzirani. Za objavo je potrebna pozitivna recenzija. Proces recenziranja je anonimen. / All articles for publication in the conference proceedings are peer-reviewed. A positive review is needed for publication. The review process is anonymous.

Lektoriranje / Proof-reading:

mag. Boštjan Zajšek, mag. Nina Gostenčnik

Prevajanje:

mag. Boštjan Zajšek (slovenščina), mag. Nina Gostenčnik (slovenščina, angleščina)

Oblikovanje in prelom / Design and typesetting:

mag. Nina Gostenčnik

Objavljeni prispevki so prosto dostopni. Vse avtorske pravice ima izdajatelj Pokrajinski arhiv Maribor.

©Pokrajinski arhiv Maribor. Za prijavo in objavo prispevkov ni potrebno plačilo. / The publication offers open access to whole texts of the published articles. ©Pokrajinski arhiv Maribor. All articles are published free of charge.

<http://www.pokarh-mb.si/si/p/3/49/moderna-arhivistika.html>

Prejeto / Received: 09. 07. 2020

1.01 Izvirni znanstveni članek

1.01 Scientific Article

IMPROVING ENDANGERED LANGUAGE ARCHIVES: A COMPARISON STUDY OF MODELS IN USE

Victoria Jesswein

Liverpool University, Centre for Archive Studies, Liverpool, Great Britain

v.j.jesswein@liverpool.ac.uk

Abstract:

Endangered language archives hold important resources of information for students, researchers, and the speech community. However, because their use is often limited to small documentary projects, often these repositories fail to reach all potential researchers, to connect their material to other repositories in the community and throughout the world, and to be involved in a larger body of research on endangered languages and cultures.

This research explores ways of expanding the accessibility, usability, and relevance of (endangered) language archives, with the aim of allowing a larger body of researchers to access materials while supporting the endangered speech community from which the material originates. Different models of linguistic archive practice are compared, each serving as an example of an area for potential archive growth and outreach: (1) the use of the language archive by a larger body of linguistic and non-linguistic researchers, (2) engagement with the speech community for heritage preservation and community-based archives, and (3) outreach to inform the general public. The models are analysed with regard to their rationale and procedures supported by relevant literature concerning best practice in the structure and use of archives. Following the analysis of compared models, recommendations are made to implement best practice for endangered language collections, envisioning the broadest and most accessible presentation of material for resource discovery, use, and preservation.

Key words:

language archives, accessibility, community archives, heritage preservation

Izvleček:

Izboljševanje Arhiva ogroženih jezikov: primerjalna študija obstoječih modelov

Arhiv ogroženih jezikov hrani pomembne vire informacij za študente, raziskovalce in jezikovno skupnost. Ker pa je njihova uporaba omejena na manjše dokumentarne projekte, hranjene informacije ne dosežejo vseh potencialnih raziskovalcev, ki bi gradivo povezali z ostalim v skupnosti in po svetu ter ga uporabljali v večjih raziskavah ogroženih jezikov in kultur.

Pričujoča raziskava raziskuje načine povečanja dostopnosti, uporabe in pomembnosti Arhiva ogroženih jezikov, s ciljem omogočiti dostopnost večji skupini raziskovalcev in ob tem podpirati skupnost jezika od koder gradivo izvira. Avtorica primerja različne modele arhivske prakse, vsak od njih pa predstavlja primer možnega nadaljnjega razvoja in dometa: (1) uporabo arhiva s strani večjega števila raziskovalcev s področja jezikoslovja in drugih področij, (2) sodelovanje z jezikovno skupnost z namenom ohranjanja dediščine in vzpostavitve arhiva, temelječega na skupnosti, in (3) informiranja širše javnosti. Analiza modelov je podkrepljena z literaturo, ki obravnava dobre prakse v strukturi arhivov in uporabi gradiva. Glede na analizo primerjanih modelov, avtorica podaja priporočila za implementacijo dobrih praks na področju zbirk gradiva ogroženih jezikov, s čimer stremi k čim bolj široki in dostopni predstavitvi gradiva za uporabo in hrambo.

Ključne besede:

jezikovni arhiv, dostopnost, arhiv skupnosti, ohranjanje kulturne dediščine

1. INTRODUCTION

Language archives are composed of the data collected and used by linguists in their research. This includes recordings of speakers collected during field research, phonetic and/or literal transcriptions of the recordings, accompanying sociological or anthropological information, and any other data such as field notes. Many language archives arise out of language documentation projects, wherein linguists record speakers to build corpora for future research. There are many endangered languages throughout the world, and with increases in globalization and technology, languages are dying out at rates faster than revitalization efforts and linguist intervention can prevent. (Crystal 2000, p. 3) Therefore, it is these endangered languages that benefit the most from documentation and preservation in an archive. While language death is often not preventable, good documentation of the language while speakers still exist is essential to preserving the memory of the language, and often the cultural and historical aspects related to it as well. This in turn provides the empirical data needed by linguists and others to research a speech community's language, history, and cultural heritage. It also provides a centre for the speech community themselves to connect with their culture and other speakers, and provides material for heritage speaker preservation efforts. (Boas, 2006, pp. 153-74).

Most of the research on linguistic archives has come from the point of view of linguists rather than from archivists. Although most language archives were designed and created for the purpose of preserving data for linguistic research, today language archives have the potential to be used by non-linguists for purposes not originally envisioned by the creators of the archives. While linguists collect data with a focus only on the scope of their project, Lisa Cornathan recognizes that "*archival records have usefulness that extends beyond the purpose for which there were originally created.*" (Cornathan, 2011, p. 238). Gary Holton (2019) recognizes that archives compiled by linguists are not "*just for linguist.*" (p. 105). It has become necessary, therefore, that when "*selecting material and building archives, we should try to imagine the widest range of possible audiences.*" (Woodbury, 2014, p. 21).

Linguistic archives differ from traditional archives in material, arrangement, and intended use, but researchers have suggested that these institutions have similar potential for education, research, and community outreach. Woodbury argues that language documentation can be multi-purpose. (Woodbury, 2011, pp. 235-54) He says "*that at their best, language documenters want their material, however conceived and assembled, to engage diverse audiences.*" (Woodbury, 2014, p. 21).

Although the consensus is that language archives have enormous potential within education, research, and community interaction and outreach, there has been little research into how best to use the material that is unique to language archives. This paper sets out to explore ways of expanding the accessibility, usability, and relevance of endangered language archives, with the aim of allowing a larger body of researchers to access and use materials and supporting the speech community from which the material originates.

2. LITERATURE REVIEW

Material in language archives differs from that of traditional archives. Few researchers have suggested specific recommendations that can be implemented to ensure language archives are successful in cataloguing and the application of meta-data and in technological functionality for discovery and access to overcome the challenges of language archives. Cornathan proposes that material originating from documentary projects is open-ended and ongoing, and is seen as an artificial collection rather than originating from a single fonds (Cornathan, 2011, pp. 235-54). The size and complexity of the records collected during documentary projects “*requires that their creators and collectors take deliberate steps to ensure the long-term usability of the material.*” (Cornathan 2011, p. 239) Because of this, some researchers have suggested that language documenters must serve as archivists of their own data, and Woodbury suggests that they should follow principles of maintaining original order and upholding provenance by organising resources hierarchically with respect to the individual speakers and the time or location of the speech events or documentary sessions. For example, records form series from their location and are divided into individual speaker files, in which each interview or clip makes up the item level. Documenters should “*create metadata based on this hierarchical organization,*” and including ethnographic, geographic, and sociolinguistic settings, so that a clear structure is represented, aiding management and retrieval. The description of a documentary linguistics collection should also give background into the research strategy and goals from which the data arise, and indicate how they might be used. (Woodbury, 2011, p. 24).

Many language archives, particularly more recent archives arising out of documentary linguistics projects are largely digital. Woodbury argues that their digital nature benefits language archives: “*Materials that are ‘born digital’ can be accessioned relatively easily...have huge capacities, so that for text and audio (but not yet video), space is barely a problem, [and]...can be searched quickly, in many more ways than traditional archives can.*” (Woodbury, 2014, p. 20). The digital nature makes the transition from research data to archival material an easy one, and has implications for preservation and ease of access. However, it also raises questions about the best practice for the description and dissemination of the material.

Professional standards such as ISAD-(G), Dublin Core, and OAIS prescribe necessary methods for the description of all types of archival materials. Further field specific recommendations propose a set of recommendations for description, storage and access specific to linguistic material. Bird and Simons set the foundation for the Open Language Archive Community and provide additional fields and codes relevant to language archives. (Bird, Simons, 2003b, pp. 117-28; Bird, Simons. 2003a, pp. 375-88) Anthony Woodbury argues that meta-data of linguistic archives should include the design and goals of the initial language documentation project, and that linguists should make guides to their material as they are most familiar with the methods of collection, arrangement, description, and storage. (Woodbury, 2014, pp. 19-36)

Linguistic records can also be perceived as oral histories, which could appeal to other kinds of researchers beyond linguists, but could require different formatting and access principles. The Oral History Association guide for best practice recommends that “*recordings of the interviews should be stored, processed, refreshed and accessed according to established archival standards designated for the media format used*” and also suggests archivists make guides to their material: “*In order to augment the accessibility of the interview, repositories should make transcriptions, indexes, time tags, detailed descriptions or other written guides to the contents.*” (Oral History Association, 2009). Because researchers using oral history records are interested foremost in the content of the speaker’s interview, rather than the phonological or syntactic structure of

the language, transcriptions play in an important role in the usability of the recordings. Linda Shopes stresses that “[a] *transcript emphasizes the contents of an interview*” and facilitates record use because it is easier and faster to skim a written document and locate relevant material. (Shopes, 2012) Eleanor Mazé emphasizes that good transcription is a labour and time intensive process, and ideally involves:

- a word-for-word rendering in print of the words (and non-verbal utterances) spoken in the interview with minimal editorial intervention;
- review of the resulting document by the interviewer for accuracy and correction as necessary;
- review by the narrator for accuracy and correction, emendation, amplification, and occasionally, redaction or restriction of certain materials;
- revision per narrator changes;
- editing and annotation for sense, context, etc;
- indexing; and
- cataloguing (Mazé, 2006, 237-271)

Each level adds additional ways the material can be found, checks for accuracy in the transcript and catalogue, and protects the rights of the informant. Mazé goes on to say that oral histories “*must be searchable to be useful*” and helping interested users find the materials they need “*relies completely on accurate and thorough content description via metadata.*” (Mazé, 2012, p. 271) Transcriptions, in conjunction with applying relevant metadata, facilitate the use of archival material and could increase the audience who are aware of and interested in the data.

Connecting with the community itself is particularly prudent for endangered language archives. Models express the implications of community participation in documentary linguistics and data creation, archiving, and language maintenance.¹ Anne Gilliland and Sue McKemmish discuss the rights of groups whose records are maintained by organisations and institutions not run by members of the community, and present participatory archives as a way for communities to “*become active, participating agents in recordkeeping and archiving practice relating to all records relating to them.*” (Gilliland, McKemmish, 2014, pp. 78-88). David Nathan argues that in contrast to most traditional archives, the field of documentary linguistics is already concerned with the “*involvement of the speech community*” and “*collaboration with community members not only as consultants but also as co-researchers.*” (Nathan, 2013), p. 2)

Czaykowska-Higgins writes about engagement with endangered language speaking communities by discounting the traditional linguist-focused model of research where the language-speaking community’s participation is limited mostly to being the source of fluent speakers, and the level of engagement in the community by a linguist is relatively small, and introduces the Community-Based Language Research model involving more engaged and collaborative research. (Czaykowska-Higgins, 2008, p. 24). Kirsten Thorpe and Monica Galassi discuss a project to develop a website “*to make the language materials easily accessible to communities*” with the aim of enabling access to

¹ *Most of the literature dealing with community archives and language maintenance is concerned with reconciling situations of colonization and systematic erasure of culture, with communities that have experienced “oppositional stances, incommensurate ontologies, or traumatic experiences” within the documentary record. This paper references participatory methods only as a method of language and heritage maintenance, but community-based participatory methods may have other benefits to language archives of other minority languages. (Gilliland, McKemmish, 2014,p.7)*

documents, establishing a dialogue with the communities, consulting them about appropriate use of material, and increase active participation in research. (Thorpe, Galassi, 2014, pp. 81-100). They also highlight the importance of documentary projects in heritage maintenance and say archives are "*institutions that should be collecting new digital content relating to languages into the future.*" (Thorpe, Galassi, 2014, p 89).

Mary Linn marries the concepts of community-based research and community archives and proposes a Community-Based Language Archive model (CBLA). She advises that archivists should notify communities of relevant collections and make sure finding aids are accessible to the community. She encourages language research for and with the community, where archivists and linguists participate more closely with education and support expanded community use of archival materials to help link communities to many kinds of professionals, which in turn links professionals to more data and new perspectives, and communities to better heritage materials. (Linn, 2014, pp. 53-67).

Archives can involve speech communities not only through the construction of the archive, but also through its continued use. Isto Huvila shows how archives could be "*radically aligned towards integration of and interaction with its actual and potential users and uses*" through participatory archives. (Huvila, 2008, pp. 15-36). Involvement with the community grants greater access to data and the possibility of more documentation, while also improving access to language materials for community members and promoting language and heritage preservation. Some documentary linguists contend that "*preserving cultural heritage through language*" by documenting language and making the materials available "*while protecting the property rights of the community from which the materials originate.*" (Boas, 2010, p. 20)

Beyond interacting and serving with the speech community itself, language archives can also support the community by informing and promoting awareness with the general public, allowing them to learn about their local history and to familiarize themselves with archival materials. Research has shown how exhibitions can promote archives to the general public. Aleksandr Gelfand suggests that "[e]xhibits are an important way for archives to communicate with the public and inform them of the types of records the repository houses, to encourage the public to use these records, and to demonstrate the archive's value and worth." (2013, pp. 49-82). He also notes the value of digital exhibitions in particular because there is virtually no limit on space, no risk of damage to material, and the exhibit can be accessed from anywhere with an internet connection. The non-physical nature of language archives raises difficulties in creating exhibitions of linguistic materials. As many documentary projects exist purely digitally, some language archives have set up exhibitions within other institutions. (Carrie, Drummond, 2019b).

3. SCOPE AND METHODOLOGY

The aim of this research is to compare models of linguistic archiving to understand how to make linguistics archives more useable and to reach a larger audience. It will undertake a review of archives demonstrating different theoretical methods and models to analyse the implementation of the best practice methods as discussed in literature, and assess how applicable these methods would be for all language archives. The research will explore and analyse how language archives function, and support the theoretical rationale behind their policy and procedure decisions with the relevant literature. This research will analyse the procedures, set-up, access principles, and functionality of comparable archival institutions and models to explore the potential of language archives as identified in the literature in 3 broad areas:

1. outreach to a larger body of linguistic and non-linguistic researchers through functionality and access principles,
2. interaction with the archive within a larger museum or archival institution to inform the general public and connect the archive with other material relating to the speech community,
3. heritage preservation and interaction with the speech community of an endangered language.

These institutions and models were chosen as examples of best practice in exploitation of their materials based on a review of the relevant literature concerning these three areas. Other examples of linguistic archival best practice exist, but these examples were chosen because of relevance to the current project and ease of access to their research, rationale, and materials.

The Texas German Dialect Archive (TGDA) is an archive of recordings and data collected by documentary linguistics projects of the Texas German Dialect Project (TGDP), run through the Germanic Studies department at the University of Texas at Austin. The TGDA aims to archive “*irreplaceable materials relating to Texas German that can be stored and maintained in perpetuity, and to make these materials available in a manner that facilitates research while protecting the property rights of the community from which the materials originate.*” (Texas German Dialect Project, nd). They do so through implementation of standards developed by the Open Language Archives Community. OLAC is “*an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by: (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 resources*” (Open Language Archives Community, nd.a). Bird and Simons (2003) set the foundation for OLAC and propose a set of recommendations for description, storage and access specific to language archives.

Manchester Voices serves as an example of an interactive documentary project using the archive of data for exhibitions to inform and involve the community at large. Manchester Voices is an ongoing linguistic research project and an archive of the data the documentary project records. Conducted by linguists at Manchester Metropolitan University, the project involved collecting documentary linguistic data from around the city, recording the accents and vocabularies. The researchers visited all 10 boroughs of Manchester in their research van and “*invited people to climb aboard, talk about their accent and dialect, and think about what it means to be from their particular area of Greater Manchester.*” (Carrie, Drummond, 2019b) The results were displayed in a now permanent exhibit at the Manchester Central Library, and also involves a website where members of the public can interact with the data. The project was selected for this research because it demonstrates an interactive approach with both speech community members and the general public.

The Community-Based Language Research and Community-Based Language Archiving models describe different levels of community involvement within linguistic research and archiving, and put forth ways in which linguists and archivists can improve their projects to better engage and support the communities from which they collect data.

The results of the project will be a report analysing various linguistics archive practices, culminating in a set of recommendations. The recommendations will consider the best practice and most practical, achievable actions to improve and expand the use of the endangered language archives in research, both in the field of linguistics and beyond, the interactive component and the role of a physical archive in public outreach, and interaction with the speech community associated with the language.

4. FUNCTIONALITY AND ACCESS

Bird and Simons describe the need for improving methods of resource discovery. (Bird and Simons, 2003b, pp. 117-28) They describe the creation of the Open Language Archives Community, the aim of which is to make an open access community of language archives easily searchable across domains. The TGDA follows Bird and Simons' detailed best-practice recommendations for the creation of digital language documentation and description concerning technological format and discovery.

While many linguistic annotations in research are done using the International Phonetic Alphabet (IPA), Bird and Simons recommend "transcrib[ing] all recordings in the orthography of the language (if one exists)." (Bird and Simons, 2003b, p. 578). Written Texas German would use a standard German orthography. However, this would not capture the features that make the Texas German dialect unique from other dialects of the language. The TGDA interview annotations use a "*modified German orthography*," one that captures linguistic phenomena of Texas German more than standard German orthography, but is more accessible and less time consuming than producing transcriptions using the IPA. (Boas, 2006, pp. 8-9). Although the annotations use a unique orthography that a user would have to accustom themselves to, this transcription method clearly demonstrates filler sounds, contractions, fast speech, and code-switching better than standard orthography and produces transcriptions that are readable without listening to the accompanying recordings. The annotations demonstrate important linguistic phenomena while remaining useable to those who are not familiar with the IPA.

While the field of documentary linguistics requires researchers to serve as archivists of their data, there are crucial differences between language documentation and the traditional archival concept. Traditionally, archives are generally seen as collections of records preserved in an archival institution for their secondary values, those beyond that of their initial reason for creation and use. However, within documentary linguistics, the collection of the corpus is the primary function of the material, and the researchers, participants, and stakeholders have their own goals and roles in the project, which affect the way data is collected and described, creating an inherent bias towards this initial project of which users outside of the documentary process need to be aware. For the data to be used outside of the project for which they were initially gathered, the material must be put into context, which is largely achieved through the description and application of meta-data.

Language archives provide a vast quantity of diverse resources upon which many varied communities depend. Many different types of users desire to have access to linguistic materials, and users want to have access to all relevant data, which may be spread across multiple repositories, some of which may be unknown to the researcher. (Birds, Simons, 2003a). Digital methods provide the most efficient means of sharing information and connecting individuals and institutions to the language resources they need. Professional meta-data standards such as the Dublin Core prescribe necessary fields and methods for the description of digital archival materials. (DCMI Usage Board, 2011). However, as the material in language archives differs in nature and use, additional means of description and dissemination can increase findability. Bird and Simons of the

Open Language Archives Community promote the use of additional meta-data fields specific to language material, as well as a controlled vocabulary of codes for describing languages uniformly across repositories within the OLAC. (Boas, 2006, p. 11) The OLAC extends the Dublin Core standard to include fields specifically relevant to language material. (Bird, Simons, 2003a, p. 6). The TGDA assigns metadata based on the informants' biographical questionnaires and "*includes the place and date of the recording, the place and date of the informant's birth, the gender, the childhood residence, the current residence, the level of education, the language(s) spoken in parents' home before elementary school, and the language(s) of instruction in elementary school.*" In addition, there are also 38 metadata values based on OLAC schema for language documentation:

"These include (1) general facts information (project, collector, content, participants, resources; (2) content subschema (interaction, explanation, performance, modality, communication context, languages, task, description, keys, register, style); (3) noncontent subschema (ID, type, role, name, language, ethnic group, age, sex, education, origin, occupation); and (4) specific metadata resource schema (resource link, type, size, format, access, quality, recording conditions, position, content encoding, character encoding, software)" (Boas, 2006, p. 11)

In addition to field specific meta-data values, the OLAC also provides a template as a guide to complete description in a standardised format and with controlled labels, permitting users to refine their search to fewer and more specific terms. Controlled vocabularies within the OLAC ensure that all repositories holding language material describe that information using the same fields and codes. The same language can be referred to by many different names, varying by speaker, location, and orthography. The OLAC Language Vocabulary implements a standard based on unique identifiers that can be implemented uniformly and will not change. It improves on previous language code vocabularies which did not account for less common languages and dialects, and employs mechanisms for building additional language codes as necessary. (Bird and Simons, 2003a, p. 7) Many minority languages may still not currently have an assigned code (Texas German being one of them), but after being assigned one, language archives could join the OLAC and integrate their meta-data with other language repositories.

Bird and Simons suggest to "*list all language resources with an OLAC repository*" to facilitate widespread discovery and interconnection of linguistic material. (Bird, Simons, 2003a, p. 576). The OLAC allows cross searching between repositories, referencing related material that can be found in a different archive. The unique identifiers increase searchability and create "*a standard for identifying languages,*" as users can enter the same search term across all repositories to locate a certain type of material, which in turn allows OLAC metadata "*to be mapped to the more general-purpose Dublin Core metadata set and disseminated to the broader community of digital libraries.*" (Bird, Simons, 2003a, p. 576).

Connecting language data to a broader field of linguistic materials could increase the findability of archival material, and field specific meta-data is useful to linguistic researchers who want to search by ethnographic, geographic, temporal, or sociolinguistic data. To expand the use of the archive to a broader field of researchers, however, the needs of non-linguists need to be considered. Other researchers would be using the recordings not as examples of linguistic data, but as oral histories, and would be interested in the content rather than structure of the interviews. Subsequently, oral histories should be accompanied by a transcript, "*a word-for-word rendering in print of the words (and non-verbal utterances) spoken in the interview*" which emphasizes the

content of the recording and makes it easier to locate specific material. A repository of oral history recordings “*must be searchable to be useful*” and the findability of resources relies on content description through the application of metadata. (Maze, 2012)

Further field specific meta-data and vocabulary sets provide additional fields and codes relevant to language archives, and serve to increase findability and uniformity across repositories. However, beyond bare minimum description, meta-data can enrich a user’s view of the archive by providing links to other records, analysing material on different levels, and showing different perspectives to help users create a picture of how a set of records coheres or can be combined with outside material. (Woodbury, 2011). Woodbury argues that corpora arising out of language documentation projects should describe the design and goals of the initial project: “*For any corpus, it is a valuable part of the meta-documentation to describe the appraisal process...[f]or example, in archiving a recorded narrative along with a transcription and translation, one might describe why one chose to include (or exclude) rough drafts of the translation, or an audio or video recording that documents the process of transcription and translation.*” (Woodbury, 2011, p. 25)

Boas asserts the TGDA has an option for “*accessing the contents of the database is via a query system that enables searches based on metadata associated with the [interview] sessions.*” (Boas, 2003a, p. 11) Searches can be conducted based on any combination of nine metadata elements: place of recording, date of recording, date of birth, gender, childhood residence, current residence, languages spoken by parents before elementary school, languages spoken by teacher in elementary school, and level of education, resulting in a list of all files matching the selected criteria. There is also a concordancer function that allows for searches of key terms in transcriptions and translations of the interview. (Speech Islands, nda). The TGDA transcriptions in a modified German orthography facilitate key-word searching more than phonetic transcriptions using the IPA, but users could miss relevant data if their Standard German search term is transcribed differently in the TGDA. One way to mitigate this is to increase indexing and cataloguing to include content information, but this also increases the workload associated with cataloguing the material. (Maze, 2006, pp. 237-71)

5. OUTREACH AND INVOLVING THE GENERAL PUBLIC

Interacting with the general public is an important way of increasing awareness and use of archives, and exhibitions are “*a necessary tool for archival promotion*” (Gelfand, 2013, p. 63). Exhibitions of language archives “*promote linguistic equality and diversity*” and strategic placement of displays within public spaces can benefit the community and serve to “*nurture a sense of social and regional pride.*” (Carrie, Drummond, 2019b) However, the nature of language archives creates challenges in producing exhibits that display linguistic data. One example of a documentary project which has produced exhibitions is Manchester Voices, a series of projects exploring linguistic diversity in the Greater Manchester area of the UK. Following the documentary project, the collected data were put on display in an exhibition at the Manchester Central Library. The exhibition displayed the data through audio-visual clips and interactive displays. Following the end of the main exhibition, and permanent exhibition of the recordings remain available in Archives+, an interactive display of archival material within the library. On the project’s website <<http://manchesterdialectmap.org/>>, visitors can also interact with maps and input their own thoughts and data. Researchers hope that the project will celebrate the region’s “*rich linguistic and cultural heritage*” in order to “*promote linguistic equality and diversity, and to nurture a sense of social and regional pride.*” (Carrie, Drummond, 2019a)

Many language repositories have no permanent or physical exhibits, and do not have the room or resources to display materials in a public space. Boas et al. of the Texas German Dialect Archive has suggested implementing a physical manifestation of their language archive in a local relevant institution, in the form of “*computer terminals in local museums to enable access to the archive. Museum visitors will then have immediate access to the archive and can listen to the stories and learn more about the history, culture, and language of the Texas German community.*” (Boas et al., 2010, p. 20) The computer station would be open to the archive’s website, allowing visitors to read transcripts, look at maps and listen to examples of Texas German. A physical station for the archive would help the resource reach people who would not normally be looking for that sort of information, and allow more connection and interaction with a larger body of users.

Although a display of the archive in museums may catch the eye of visitors who would not otherwise be aware of the archive, the suggestion of a computer terminal display is an outdated one and comes with a variety of problems. Gelfand (2014, p. 65) expresses the downfalls of the use of interactive computer exhibits in the past:

“The visitor had to be physically present in the museum to engage them, and they could only be used by a limited number of visitors at a time. Usually a single person interacted with the terminal while everyone else watched. The terminal was only accessible when the museum was open and none of the information could be saved for later use. In addition, the terminal and its components had to be physically durable if they were to stand up to the stresses of frequent use. This required someone with technical expertise to frequently be onsite to address any issues. Breakdowns were not uncommon”

Moreover, this type of exhibit does not put the archival data into context and offers “*no interpretation to help a regular visitor understand what they were looking at.*” (p. 58) Additionally, in cases where the display is entirely available online, a computer terminal would not give visitors to the museum an experience that would not be possible from a home computer or elsewhere. Still, computer displays have the potential to raise awareness about resources available online and encourage visitors to interact with the archive in their own time.

Other types of exhibitions of language archive material could generate more interest and interaction with the language data. Gelfand (2014) argues that “*[a]nalog exhibit curation allows archivists to put on the shoes of researchers and demonstrate their scholarship, at the same time building relationships with the public at large*” (p. 77). In the cases of Manchester Voices and many other documentary linguistic projects, the archivists are also the researchers, which gives them even better opportunity to display material in a way that represents the research projects out of which the data arise. Furthermore, “*[t]he connections built during these cooperative efforts may someday be put to good use and tend to increase the reputation of both the archivist and the archive*” (p. 77). By forging connections with the public and raising awareness about the work they do, language documenters have the opportunity to gain support for continuing their research, allowing for more and better documentation, research, and dissemination of findings. Exhibitions give the speech community a place within the larger society that they are a part of, and inform members of the general public who may not be familiar with the community. Displays of language archives that interpret and promote interaction with the material rather than simply present raw language data can promote the material and support the community by “*celebrating diversity through language in a variety of artistic forms.*” (Carrie, Drummond, 2019a)

Many language archives lack an “*adequate venue to hold exhibits*” that “*force them to rely on museums and other locales.*” While they must rely on and conform to the larger institution, the archive’s display gains the use of a display space, and there is also the potential for the museum to lend additional materials to the archive’s display, thereby “*substantially improving the content and aesthetic of the exhibit*” (Gelfand, 2013, p. 62). It is advantageous to the Manchester Voices exhibition that it is housed within the Archives+ exhibit, which itself is a part of the large Manchester Central Library. Being housed within a larger institution can increase the number of viewers by catching those who visit the museum/archives for other reasons. Additionally, it allows the relatively small data set of the Manchester Voices project to be linked with a larger collection of materials relating to Manchester’s history, culture and language.

6. HERITAGE PRESERVATION AND SPEECH COMMUNITY INVOLVEMENT

For documentary linguistics projects and endangered language archives, the goal of the work is to ensure vital cultural information is not lost. Czaykowska-Higgins’ Community Based Language Research (CBLR) and Linn’s Community Based Language Archive model (CBLA) advocate for community involvement in research projects and archives to ensure that the community is not only recorded for posterity, but also has a say in the way they are represented in the records, and has access to the data for their own language and heritage maintenance.

The model is a graded ladder of achievement against which an archive can measure progress. The five levels in the model describe stages of involvement of the speech community in the design and use of the repository. In the first level of the CBLR/CBLA model, “linguists are responsible and accountable primarily to themselves and to their academic or scholarly communities” and not to the language community (Czaykowska-Higgins, 2009, p. 22) At the second level, the focus is on the exploitation of data for research purposes, but “*the linguist works to minimize any damage or inconvenience arising from his/her presence or research.*” This is the minimum requirement for conducting ethical research involving human subjects at most US-based research institutions, and works toward protecting the rights and information of their informants, a minimize negative effects of the research: “*to protect anonymity, informants’ names are not included in the web-accessible data, and sections of interviews that could potentially be used to identify the informants are removed.*” (Boas, 2006, p. 5)

Mary Linn (2014) illustrates how the CBLA method can be applied to existing archival collections through her work with a collection of materials on the Osage language, a moribund language with no fluent native speakers. In the project, a non-community member graduate student created a preliminary finding aid based on the original order of the materials, and then staff and members of the Osage Nation Language Program met to discuss collaborating on the materials. The community members decided which materials were most valuable to them, making a list of highest priority items, and thus establishing an order for digitisation. The collaboration between the community and the archivists was able to determine what materials were most important for preservation, and identify the language materials that “*were immediately useful*” to the Osage Nation (Linn, 2014, p. 62).

To better support both the goals of the archive and the community that it serves, language archives should implement further methods of community-based research and archiving methods. At the next level of the CBLR/CBLA model, research begins to advocate for communities, and the barriers between researchers and communities are

broken down. At this level, archivists working with communities interested in heritage preservation should make community members aware of the materials available, and may be asked to provide advice about how best to access, store, and use archival material. Community members could also get involved in the transcription process, as native speakers have the best understanding of the dialect and may be able to produce more accurate transcriptions and translations.

Language revitalisation is a major aspect of heritage maintenance, and one that may attract community members to participate in research and archive projects. In many cases of linguistic documentation, the dialect(s) are moribund and there are no hopes of seeing a revitalized use of the language. However, the data in the archives record important information about the history and culture of the speech community, and serve as vital heritage material for the community even after language death.

By implementing more community-based methods into their design and strategies, language archives can better serve the speech communities and connect them to valuable records concerning their history and culture. Community-based methods could involve displaying and explaining the archive to community members, particularly ones who have participated in interviews. By ensuring data and finding aids “*reside within the community,*” and “*explaining how to use the linguistic material,*” the language archives have the potential to inform community members about their history, language, and culture (Linn, 2014, p. 55). Further levels of CBLR/CBLA may involve training community members on how to use and engage with the linguistic materials, and the engagement of communities by researchers from the outset and for all steps in the development of the archive.

The goal of community-based archiving methods is to enable community-driven social change and identity expression through the maintenance of language and culture. Endangered language archives can advocate for heritage language communities whose records are in their keeping and upon whose data their research depends. Linn proposes that beyond supporting the speech community, the increased relationships between communities and archivists leads to better research by non-community members: “*In supporting expanded community use of archival materials, archives help link communities to many kinds of professionals, and in turn link these professionals to lesser-known knowledge and new perspectives*” (Linn, 2014, p. 65). As more community-based language research and archiving methods evolve, language archives will grow as institutions that advocate for minority communities and support heritage preservation.

7. RECOMMENDATIONS

Following an analysis of compared models, a set of recommendations arises. These recommendations serve as a basis for best practice in the archives examined, and can potentially serve as guidelines for all endangered language archives, taking into account current best practice as discussed in archival theory, the needs of the linguistic data, and the needs of the speech community in term of access, functionality, and usability of records.

1. Language documenters should make corpus specific guides to their material. Linguists involved in documentary projects are the most familiar with their material and best suited to provide guidance to students, teachers, members of the speech community, and the general public.
2. Language archives should be concerned with archival best practice of description and implement corpus specific meta-data. Because the nature of linguistic data differs from material in traditional archives, language archives should implement

specific descriptive fields applicable to their data to assist potential researchers in locating data of which they would otherwise be unaware. Language archives should use controlled vocabularies and meta-data codes to ensure uniformity within their catalogues, as well as increasing cross-searchability between repositories. Using standardised codes for languages eliminates the need to search all possible versions of the language name in order to find resources.

3. To allow their data to be used by a broader field of non-linguistic researchers, language archives should also implement practices that facilitate the use of their recordings as oral histories. Transcriptions, search functions, indexing, and content descriptions could enable a larger body of researchers to locate and use endangered language material.
4. Linguistics archives should provide references to literature discussing the linguistic features found in their material. By forming connections between the raw data and the research that has been conducted by members of the documentary linguistic project or others who have used the data, the archive could help researchers theorise the data. It would also make the archive easier to be used a teaching resource, whereby instructors can link academic articles to real-life examples of linguistic phenomena.
5. Language archives should make use of exhibitions to inform the general public about the work they do, and generate more interest and interaction with their material. They should collaborate with other heritage bodies to ensure that their materials are exhibited alongside other relevant content in places where visitors are already engaged with heritage. By forging connections with the public, language documenters have the opportunity to advocate for their repositories and gain support for their research.
6. Language archive should form connections with other repositories that hold materials related to their archive or which attract researchers working on similar projects. Language archives can benefit from the display spaces and existing customer bases of museums and libraries within the community. By advertising inside a larger institution, language archives promote they work they do and have the potential to interest members of the public who may be previously unaware of the archive.
7. Language archives can use computer terminals to display their web-based archives. However, analogue or other types of exhibitions should also be used to interpret the material and provide context for the viewer. The archive should not be presented without context, and additional information would need to be provided with the web-based data.
8. Language archives should use community-based methods of research and archiving to contribute to the social group's knowledge, control, and use of materials originating from their community. Involving the community enables community-driven social change and identity expression through the maintenance of language and culture.
9. Language archives should involve the community in design and creation of the archive and any research produced from the data. In supporting community involvement, archives can connect communities and researchers, supporting ethical research methods and linking professionals to the internal knowledge and perspectives from the speech community members.

8. CONCLUSION

The goal of this paper was to explore the potential of linguistic archives to increase awareness and use of their materials. The research has argued that language archives such as the TGDA and the entire field of documentary linguistics should be concerned with archival principles to implement best practice for their collections, and envisioning the broadest and most accessible structure to material in order to draw in all potential users. The potential of language archives has been expressed in 3 broad areas: the use of the language archive by a larger body of linguistic and non-linguistic researchers, engagement with the speech community for heritage preservation, and outreach to the general public. This research used different models that each served as an example of one of the 3 potential areas, and applied to practices of these models to the TGDA. Suggestions of changes to TDGA procedures are supported by relevant literature concerning best practice in the structure and use of archives.

Language archives should use relevant models and recommendations to tailor their archive best suit to the nature of their material and serve the needs of the speech community with which they work. By implementing the recommendations presented here, endangered language archives have the potential to grow the current body of users and connect its material to other repositories in the community and throughout the world. They can grow their project by advocating for their research, involving the speech community and gaining the support of the general public. As language archives evolve, archivists, researchers, students, and communities will explore what can be accomplished through increased collaboration and better access to information.

SOURCES AND LITERATURE

- Bird, S., Simons, G. (2003a).** Extending Dublin Core Metadata to Support the Description and Discovery of Language Resources. *Computers and the Humanities*, 37(4), 375-388.
- Bird S., Simons G. (2003b).** The Open Language Archives Community: An Infrastructure for Distributed Archiving of Language Resources. *Literary and Linguistic Computing*, 18(2), pp. 117-128. Oxford: Oxford University Press.
- Boas, H. C. (2006).** From the Field to the Web: implementing best-practice recommendations in documentary linguistics. *Language Resources and Evaluation*, 40. 153-174.
- Boas, H. C. (2009).** *The Life and Death of Texas German*. Durham NC: Duke University Press.
- Boas, H. C. et al. (2010).** The Texas German Dialect Archive: A Multimedia Resource for Research, Teaching, and Outreach. *Journal of Germanic Linguistics*, 22(3), 277-296.
- Boas, H. C., Weilbacher, H. (2011).** Documenting Diaspora Experiences: The Texas German Dialect Archive. in: *Proceedings of the Waterloo Conference on Diaspora Experiences*.
- Brown, K. (2016).** On the Participatory Archive: The Formation of the Eastern Kentucky African American Migration Project. *Southern Cultures*, 22(1), 113-127. Retrieved from <http://www.jstor.org/stable/26221783> on 08 June 2019.
- Carrie, E., Drummond, R. (2019a).** Manchester Voices: Dialect Map. Retrieved from <http://manchesterdialectmap.org/> on 08 June 2020.
- Carrie, E., Drummond, R. (2019b).** Manchester Voices: Documenting and celebrating the accents and dialects of Greater Manchester. Retrieved from <https://ahrc-blog.com/2019/04/25/manchester-voices-documenting-and-celebrating-the-accent-and-dialects-of-greater-manchester/> on 08 June 2020.

- Cornathan, L. (2011).** Archiving and language documentation. In: Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 235-254. Cambridge: Cambridge University Press.
- Crystal, D. (2000).** *Language Death*. Cambridge: Cambridge University Press.
- Czaykowska-Higgins, E. (2009).** Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation and Conservation*, 3(1), 15-50.
- DCMI Usage Board (2011).** Dublin Core Metadata Element Set, Version 1.1: Reference Description. Retrieved from <http://www.dublincore.org/specifications/dublin-core/dces/> -on 06 June 2020.
- Gelfand, A. (2013).** If We Build It (and Promote It) They Will Come: History of Analog and Digital Exhibits in Archival Repositories. *Journal of Archival Organization*, 11:1-2, 49-82. Retrieved from <https://www.tandfonline.com/doi/abs/10.1080/15332748.2013.882160> -on 06 June 2020.
- Gilliland, A., McKemmish, S. (2014).** The role of participatory archives in furthering human rights, reconciliation and recovery. *Atlanti: Review for Modern Archival Theory and Practice*, 24, 78–88.
- Holton, G. (2009).** *Developing relationships between archives and speaker communities*. Open Language Archive Community Tutorial: Working with Speakers to Create Access to Heritage Materials, Linguistic Society of America Annual Meetings, San Francisco, California, January 11, 2009.
- Holton, G. (2012).** Language archives: They're not just for linguists any more. In Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts and Paul Trilsbeek (eds.) *Potentials of Language Documentation: Methods, Analyses, and Utilization*, 105110. *Language Documentation and Conservation Special Publication No 3*.
- Holton, G. (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.
- Huvila, I. (2008).** Participatory archive: Towards decentralized curation, radical user orientation, and broader contextualization of records management. *Archival Science*, 8, 15-36.
- International Organization for Standardization. (2007).** ISO 639-3: Codes for the representation of names of languages. Retrieved from <https://iso639-3.sil.org/code/pdt> - on 06 June 2019.
- Linn, M. (2014).** Living archives: A communitybased language archive model. In David Nathan & Peter K. Austin (eds) *Language Documentation and Description*, vol 12: Special Issue on Language Documentation and Archiving, 53-67. London: SOAS.
- Mazé, E. A. (2006).** The Uneasy Page: Transcribing and Editing Oral History. In Leslie Roy Ballard (ed.) *History of Oral History: Foundations and Methodology*. Lanham, MD: Rowman Altamira. 237-271.
- Mazé, E. A. (2012).** Metadata: best practices for oral history access and preservation. In D. Boyd, S. Cohen, B. Rakerd, & D. Rehberger (Eds.), *Oral history in the digital age*. *Institute of Library and Museum Services*. Retrieved from <http://ohda.matrix.msu.edu/2012/06/metadata/> on 06 June 2020.
- Nathan, D. (2013).** Access and accessibility at ELAR, a social networking archive for endangered languages documentation. In Mark Turin, Claire Wheeler and Eleanor Wilkinson (eds.) *Oral Literature in the Digital Age: Archiving Orality and Connecting with Communities.*, Cambridge: Open Book Publishers.
- Open Language Archives Community. (nd.a).** *OLAC Mission*. Retrieved from <http://www.language-archives.org/> on 04 June 2019.

- Open Language Archives Community (nd.b).** *OLAC resources in and about the Plautdietsch language.* Retrieved from <http://www.language-archives.org/language/pdt> on 15 July 2019.
- Open Language Archives Community (nd.c).** *OLAC resources in and about the Hutterite German language.* Retrieved from <http://www.language-archives.org/language/geh> on 15 June 2020.
- Open Language Archives Community (nd.d).** *OLAC resources in and about the Pennsylvania German language.* Retrieved from <http://www.language-archives.org/language/pdc> on 15 June 2020.
- Oral History Association (2009).** *Principles and Best Practices.* Accessed from <http://www.oralhistory.org/do-oral-history/principles-and-practices/> on 12 June 2020.
- Shopes, L. (2012).** Transcribing oral history in the digital age. In D. Boyd, S. Cohen, B. Rakerd, & D. Rehberger (Eds.), *Oral history in the digital age. Institute of Library and Museum Services.* Retrieved from <http://ohda.matrix.msu.edu/2012/06/transcribing-oral-history-in-the-digital-age/> on -12 June 2020.
- Speech Islands (nd).** *Concordancer.* Retrieved from <https://speechislands.org/concordancer> on 12 June 2020.
- Texas German Dialect Project. (nd).** *About.* Retrieved from <https://tgdp.org/about/#support> -on 08 June 2020.
- Texas German Dialect Project. (nd).** *Dialect Archive Viewer.* Retrieved from <https://tgdp.org/dialect-archive-viewer/#/archive> -on 08 June 2019.
- Thorpe, K., Galassi, M. (2014).** Rediscovering Indigenous Languages: The Role and Impact of Libraries and Archives in Cultural Revitalisation., *Australian Academic & Research Libraries*, 45:2, 81-100.
- Woodbury, A. C. (2011).** Language documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 235-254. Cambridge: Cambridge University Press.
- Woodbury, A. C. (2014).** Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. In David Nathan (ed.), *Proceedings of Workshop on Language Documentation and Archiving.* London: Hans Rausing Endangered Language Project, SOAS.

POVZETEK

IZBOLJŠANJE ARHIVA OGROŽENIH JEZIKOV: PRIMERJALNA ŠTUDIJA OBSTOJEČIH MODELOV

Victoria Jesswein

Univerza v Liverpoolu, Center za arhivske študije, Liverpool, Velika Britanija

v.j.jesswein@liverpool.ac.uk

Cilj prispevka je ugotoviti, kako povečati prepoznavnost in uporabo gradiva, ki ga hranijo jezikovni arhivi. Raziskava je pokazala potrebo, da tovrstni arhivi, kakršen je TGDA (Arhiv teksaško nemškega narečja) kakor tudi celotno področje jezikovne dokumentalistike v svoje delo vključijo arhivska načela ter dajo gradivu široko in dostopno strukturo, ki bo zanimiva za najrazličnejše uporabnike. Potencial jezikovnih arhivov se izraža na treh področjih: v uporabi gradiva s strani jezikovnih in drugih raziskovalcev, sodelovanju z jezikovno skupnostjo za ohranjanje dediščine in doseganju široke javnosti. Pričujoča raziskava je uporabila različne modele, ki so služili kot primer za vsakega od treh področij in ki jih je TGDA uporabil pri svojem strokovnem delu. Predlogi, ki jih prispevek podaja glede sprememb v delovanju TGDA, so podkrepljeni z relevantno literaturo o dobrih praksah na področju urejanja in uporabe arhivskega gradiva.