

# Creating Workflow for Mediated Archiving in CoRSAL

Merrion Dale

Department of Information Science

University of North Texas

Denton TX, USA

JessicaDale@my.unt.edu

## ABSTRACT

Language archiving involves the collection and curation of a variety of language materials. As an emerging language archive, CoRSAL caters to a range of different language depositors with different research needs. As such, we have developed a workflow process that can accommodate this diversity.

## CCS Concepts

• Information systems~Information systems applications~Digital libraries and archives • **Human-centered computing~Interaction design~Interaction design process and methods~User centered design • Human-centered computing~Collaborative and social computing~Collaborative and social computing theory, concepts and paradigms~social media • Human-centered computing~Collaborative and social computing~Collaborative and social computing theory, concepts and paradigms~Social content sharing**

## KEYWORDS

Language archiving, Participatory archiving, Community archiving, Archival workflow, Archival engagement, Social media

## 1 Introduction

The Computational Resource for South Asian Languages (CoRSAL) is an archive housed within UNT's Digital Library that specializes in the curation and preservation of source audio, video, and textual language data. CoRSAL is a dedicated repository for linguists to archive grammars and related grammatical information of South Asian languages, many of which are endangered and/or under-resourced. Since its inception in 2016, a primary objective for CoRSAL has been to engage closely with community depositors in each step of the archiving process, from pre-archival decision making to post-archival collection promotion. The ingest process is purposely mediated and participatory because of our diverse depositor base. This paper offers an exploration of CoRSAL's curation and ingest workflow and our findings at this fairly early stage in the growth of CoRSAL. I also include a brief discussion of our desired direction for the future. The results and feedback from our mediated approach have been encouraging so far. For this reason, we will explore additional opportunities to involve speaker-community

members in the archiving process, and in increasing their engagement with archived collections.

## 2 Background

The emergence of a participatory archiving framework in the twenty-first century is contributing to a shift in archival power dynamics [10]. The participatory framework privileges the power of those connected through ownership or cultural heritage to archival materials so that they may play integral roles in the archiving process as archive designers, researchers, depositors, and users [7, 15]. As outlined in Cook as well as in Roeschley and Kim [7, 14] participatory archives may, depending on the wishes and needs of the community, be facilitated, and established by institutional archives.

CoRSAL's approach to archiving falls within the participatory framework by implementing a mediated workflow that is very much a conversational exchange between depositor and curator and resulting in a collection that is representative of each depositor/community's research needs. Additionally, community depositors can engage in post-archival participatory measures in the sharing and promotion of their collections. It is an ongoing concern that general engagement with archived language collections is low among members of the language communities. One contributing factor, based on Burke and Zavalina's [2] interviews with linguistics researchers/depositors, is that some depositors, especially those from Indigenous communities, access information through social media. The researchers also found that many of the Indigenous have limited access to reliable, high-speed internet. As a result, the community's individuals use mobile devices in place of computers and find streaming media platforms like Facebook, Youtube, and Instagram more realistic to access than potentially downloading an archive's large files. This contributes to an overall preference for a "social media-type interface over the archival access point" [2, 3].

Thus, as a final piece of the mediation between the depositor and the archive, and to facilitate the use of the archive as framed by the depositor, CoRSAL has worked with recent depositors to use that existing interest [4, 5, 6, 8, 9, 11, 12, 13] as a means to increase engagement with language archives. Social media has become a tool for Indigenous communities to utilize for purposes of connection, outreach, and activism. As of October 2020 the CoRSAL team has been exploring how to utilize Facebook's private group feature as a tool for depositors to share items with

their community and for the community to discuss the items with the depositors and one another.

### 3 Methods

Since 2016 the CoRSAL team has done a great deal of outreach work in the form of lectures and workshops. The purpose of this was to explain the process and the purpose for archiving with CoRSAL and offer examples of the ways archived language data can be useful for language revitalization and pedagogy development. As a result of this type of work the team has often received interest from young community researchers who are interested in discussing their own potential collections in greater detail. These conversations have recently resulted in several new language collections in CoRSAL, such as Azamgarhi and Bhojpuri language collections deposited by Maaz Shaikh, and the Boro Language Collection deposited by Prafulla Basumatary.

At present, CoRSAL has eleven published language collections which reflect its diverse depositor pool. For instance, some of the collections, like the Burushaski Language Resource and Lamkang Language Resource, are the result of funded documentary linguistic projects. Others are legacy collections, that is, materials collected in pre-digital times. CoRSAL uses a mediated approach here as well as researchers in possession of a lifetime's worth of fieldwork notes and recordings are often overwhelmed by the tasks of data management and file transfer, and metadata creation. To create metadata for legacy collections, CoRSAL curators collaborate with depositors to aid with data management and metadata recall.

When a depositor is interested in archiving a collection with CoRSAL we work with them hand-in-hand to assist with every step of the process. The workflow consists of communicating regularly over Zoom and email to discuss selection and collection of content, file transfer, the process of filling in the CoRSAL metadata spreadsheet, and general data management. This is the longest part of the process, taking less or more time depending on the state of data management when we begin the process. When there are gaps in the metadata, the CoRSAL team attempts to help fill those. For example, in the case of the Azamgarhi collection, we lacked an identifying language code. A member of the CoRSAL team applied for and obtained a Glottocode for the language on the depositor's behalf. Depositors also have several opportunities to provide context for their collection. For example, we work with depositors to write a brief description of the language and the collection for the collection's landing page in the CoRSAL archive. We assist and edit only as needed. Depositors are also given the option to contribute a thumbnail image for their collection.

Once the digital files and their metadata are transferred to a dedicated CoRSAL drive, the materials are made available to staff at the UNT Digital Library to upload the items to the archive. Despite being published, collections can still be edited after this stage. When a depositor wants to modify their metadata, they can inform us of changes they have made to the spreadsheet and a member of the team implements the changes on their behalf.

Again, keeping our diverse depositor pool in mind, CoRSAL does expect that deposits will be incrementally added to the CoRSAL drive. This possibility of incremental growth of what is to be deposited supports weaker infrastructure for collectors - lack of backup media, unstable computers, a possibility for loss of data. If depositors need to move files into the CoRSAL drive in smaller batches, we allow and encourage this. We don't ask for all of the metadata at once, but rather the key pieces, which depositors can then build upon. We also believe that it is not essential to immediately expect depositors to be ready or able to provide intellectual access to the individual items, i.e., through transcription and translations. Our curators acknowledge that the source data is the most precious and under the most pressing danger of loss. We encourage depositors to approach archiving by starting small and holding the view that the other important items can be obtained as long as we have adequate metadata for the source files. As part of the metadata, we do encourage at least a rough translation. It is our goal to provide training so that depositors are ultimately able to provide different levels of access through transcription and translation using more interlinear glossed texts, and through encouraging the publication of text collections.

Another way in which CoRSAL encourages and then provides help for depositors to "own" their collections is through depositors and associated community members improving metadata, especially in enriching keywords with ethnographic information. In the case of two language collections, Lamkang and Burushaski, two community researchers contacted their peers requesting additional context on a number of photos and videos. The community researcher for the Burushaski collection was able to crowdsource helpful feedback by posting in an existing Burushaski Facebook group [1].

We encourage community depositors to promote their collections on social media as they are interested/able. For example, we have assisted several depositors in creating private Facebook groups. Within these groups the depositor/group admin shares items from their collections with members of their communities. Some depositors share their collection in existing Facebook groups, namely the Burushaski Research Academy group. Maaz Shaikh, the depositor of the Azamgarhi and Bhojpuri collections, recently created his own Azamgarhi Facebook group after observing the models of existing groups.

For each group created as part of this promotional initiative the depositor acts as the primary group admin and is responsible for inviting new members to join. They are also responsible for accepting/rejecting any membership requests the group receives. The admin shares items directly from their collection which they believe will be of specific interest to group members. Members are then able to click on the shared link and view the item in the archive.

### 4 Findings

Depositors who have completed the archiving and publication process with CoRSAL have given very positive feedback on the workflow and their resulting collections and publications. CoRSAL is also seeing encouraging growth and engagement from the newly created Facebook groups. By monitoring the usage

statistics feature embedded in each collection we are able to see that the shared items are being used via their Facebook referral links. We can thus track the number of uses an item has in general, and more granularly, over the last 30 days. Additionally, Boro group admin/depositor Prafulla Basumatary has reported that he has received useful fieldwork recommendations from community members who have commented on his shared posts.

## 5 Statement of Significance

A participatory approach to archiving corrects for imbalances in power that unjustly take intellectual, cultural, and material information from one set of people for inquiry by another. The goals of participatory archiving can be greatly supported through a mediated archiving workflow. This helps non-professional collectors and depositors develop an understanding of the archiving process and of how to enhance the usability and accessibility of each collection. While this mediated methodology is time consuming and labor intensive, it gives depositors the power to frame and contextualize a collection. It also requires a sympathetic connection with speakers.

## 6 Conclusion and Future Work

The CoRSAL team feels encouraged by depositors' responses to the workflow. Boro depositor Prafulla Basumatary is currently preparing a collection guide which will assist users in how best to utilize the collection for revitalization efforts, particularly relating to development of pedagogy. We expect to see a similar guide for the Azamgarhi and Bhojpuri collections come to fruition this year. CoRSAL curators are actively in the process of curating more legacy collections and would like to have them reviewed and vetted by community members before their publishing. We are hoping that inviting community members to promote these collections on Facebook will lead to even greater community viewership and feedback.

Additionally, as the community Facebook groups continue to grow, we are actively researching more ways we can potentially use social media to enhance the community experience engaging with the archive. Azamgarhi depositor Maaz Shaikh has recently taken the initiative of creating an Azamgarhi Instagram account to promote his Facebook posts to a wider audience and we are excited to see the response this will receive. We are also interested in exploring other social media platforms like WhatsApp and TikTok that may be more appealing to younger community members.

## ACKNOWLEDGMENTS

I'm very grateful to Dr. Shobhana Chelliah for her helpful reviews and feedback.

## REFERENCES

- [1] Mary Burke, Sumshot Khular, Javid Iqbal, and Mark Phillips. 2021. Enriching ethnographic metadata with the help of native speakers Poster presentation. *7th International Conference on Language Documentation and Conservation (ICLDC)*. March 4-7, 2021, Honolulu, HI. URI: <http://hdl.handle.net/10125/74477>.
- [2] Mary Burke and Oksana Zavalina. 2020. Identifying Challenges for Information Organization in Language Archives: Preliminary Findings. In Anneli Sundqvist, Gerd Berget, Jan Nolin, and Kjell Skjerdingsstad (eds) *Sustainable Digital Communities. iConference 2020*. March 23-26, 2020,

- Boras, Sweden. Springer, Cham, Switzerland, 622-629. DOI: [https://doi.org/10.1007/978-3-030-43687-2\\_52](https://doi.org/10.1007/978-3-030-43687-2_52).
- [3] Mary Burke. In press. User Needs in Language Archives: Findings from Interviews with Language Archive Managers, Depositors, and End-Users.
- [4] Laura Buszard-Welcher. 2001. Can the Web Help Save My Language? In Leanne Hinton & Ken Hale (eds.) *The green book of language revitalization in practice*, 331-345. Brill, Leiden, Netherlands. DOI: <https://doi.org/10.1163/9789004261723>
- [5] Morgan Cassels. 2019. Indigenous languages in new media: Opportunities and challenges for language revitalization. *Working Papers of the Linguistics Circle of the University of Victoria* 29, 1 (Sept. 2019), 25-43.
- [6] Coppellie Cocq. 2015. Indigenous Voices on the Web: Folksonomies and Endangered Languages. *Journal of American Folklore* 128, 509 (Aug. 2015), 273-285.
- [7] Terry Cook. 2013. Evidence, memory, identity, and community: four shifting archival paradigms. *Archival Science* 13, 2 (Jun. 2013), 95-120. DOI: <https://doi.org/10.1007/s10502-012-9180-7>.
- [8] Dario De Falco and Alfonso Cesarano. 2016. Endangered Languages in the Era of Social Media: The Case of the Kenyah Lebu' Kulit Language. In Joseph Cru (Ed.) *Linguapax Review 2016: Digital Media and Language Revitalization*. 55-65. Linguapax International, Barcelona, Spain.
- [9] Gabriel Djomeni and Etienne Sadembouo. 2016. African Languages and Digital Media: Practice, Challenges, and Perspectives in Cameroon. In Joseph Cru (Ed.) *Linguapax Review 2016: Digital Media and Language Revitalization*. 33-54. Linguapax International, Barcelona, Spain.
- [10] Ryan Henke and Andrea L. Berez-Kroeker. 2016. A Brief History of Archiving in Language Documentation, with an Annotated Bibliography. *Language Documentation & Conservation* 10, (Dec. 2016), 411-457. URI: <http://hdl.handle.net/10125/24714>.
- [11] David Lee. 2011. Micro-blogging in a mother tongue on Twitter. (April 2011). Retrieved June 16, 2021 from [http://news.bbc.co.uk/2/hi/programmes/click\\_online/9450488.stm](http://news.bbc.co.uk/2/hi/programmes/click_online/9450488.stm).
- [12] Brook Danielle Lillehaugen. 2016. Why write in a language that (almost) no one can read? Twitter and the development of written literature. *Language Documentation & Conservation* 10, (Sept. 2016), 356-393. DOI: <http://hdl.handle.net/10125/24702>.
- [13] Brook Danielle Lillehaugen. 2019. Tweeting in Zapotec: Social Media as a Tool for Language Activists. In Jennifer Carolina Gómez Menjívar and Gloria E. Chacón, (eds.) *Indigenous Interfaces: Spaces, Technology, and Social Networks in Mexico and Central America*. University of Arizona Press, Tucson, Arizona.
- [14] Ana Roeschley and Jeonghyun Kim. 2019. "Something that feels like a community": the role of personal stories in building community-based participatory archives. *Archival Science* 19, (March 2019), 27-49. DOI: [10.1007/s10502-019-09302-2](https://doi.org/10.1007/s10502-019-09302-2).
- [15] Christina Wasson. 2021. Participatory Design of Language and Culture Archives. *Oxford Research Encyclopedia of Anthropology*, (Mar. 2021). DOI: <https://doi.org/10.1093/acrefore/9780190854584.013.234>