The Archive of the Indigenous Languages of Latin America:
Its Mission, History, and Operations

Susan Smythe Kung, PhD
skung@austin.utexas.edu
www.ailla.utexas.org

COLING Project Workshop
University of Texas at Austin
12 December 2019

Land acknowledgement.

Thanks to the COLING Team for bringing us all together for this workshop and for inviting me to participate. Even though most of the speakers this afternoon and tomorrow are from UT-Austin, we rarely have the opportunity to discuss our work with each other, and in fact, this is the first time that some of us will be hearing about each other projects.

The Archive of the Indigenous Languages of Latin America (AILLA) is a dynamic digital online repository here at the University of Texas at Austin that was founded in 2000 to house collections of linguistic and anthropological materials about Latin American Indigenous languages. Since then, AILLA has grown into one of the premier archives for endangered languages, and it is one of the founding members of DELAMAN, the Digital Endangered Languages and Musics Archives Network. In this presentation, I briefly describe AILLA, its history and mission, and its current projects and operations.

1

Archive of the Indigenous Languages of Latin America (AILLA)

Digital language data archive founded in 2000
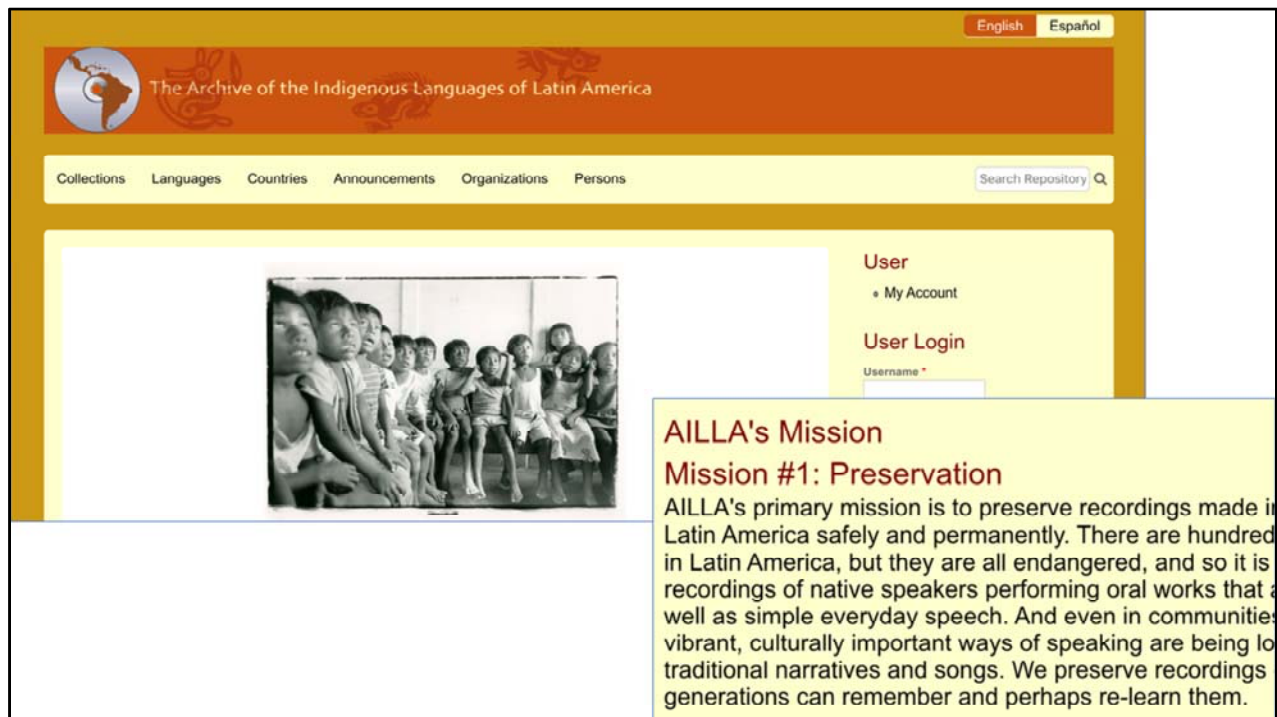
Growing multimedia collections:

- 374 Indigenous languages of the Americas represented so far

- 218 distinct collections

- **Digitized audio/video and images** from legacy collections

- **Born-digital materials** from recent/on-going research projects

- **Participatory collections** managed by depositors

- **Bilingual** (English-Spanish) website and metadata

- Metadata are **open**

- Media files are behind an access wall, but **accounts are free**

- Password-based access **restrictions for sensitive or embargoed media**

About half (174) of AILLA's languages are no longer widely learned by children

Nearly a fifth (69) of AILLA's languages holdings are spoken by few, if any people.

Anyone can deposit material into AILLA
- Depositors retain all copyrights – they give AILLA a non-exclusive license to reproduce, display, and share the media files
- AILLA has a self-deposit tool that allows Self-Depositors to update and edit metadata and ingest files via a GUI
- S-Ds must be trained and sign an agreement before being given the self-deposit role
- They have access only to their own collection, files, and metadata.

Since its beginning, AILLA's Mission is three-fold:
1. AILLA's primary mission is to safely and permanently preserve recordings made in the Indigenous languages of Latin America.
2. The 2nd mission is to make the resources in AILLA available to the people who can make good use of them:
   - the Indigenous people of Latin America,
   - the scholars who study these languages and cultures, and
   - interested members of the general public.
   - Basically, everyone.
3. The 3rd mission is to do whatever we can to support the survival of the Indigenous languages of Latin America, and to help their speakers make them flourish. (One way we can help is by fostering the

Image: Joel Sherzer helps grind sugarcane for an Inna celebration ca. 1970. Kuna Photograph Collection of Joel Sherzer. Archive of the Indigenous Languages of Latin America https://ailla.utexas.org/islandora/object/ailla:246379

- In late 80s Joel Sherzer (Emeritus, Anthropology, UT-Austin) had the idea for a clearing house for recordings of indigenous languages, a place where researchers could share their recordings with each other.
- As the Internet became more widely used, his idea evolved to depend on it (the Internet) to make those recordings of indigenous languages accessible all over the world.
- In 2000 so he called NEH about funding a project. The Program Officer told him that the University of Texas Libraries had one of the best departments for Digital Libraries Services in the country and that he should partner with the Library to make his idea a reality.
- Thus, Sherzer and his grad students (in Anthropology) collaborated with a team of software developers from the library to create a pilot site that launched in March 2001.
- Sherzer, and his collaborator--Mark McFarland (Lib) & Anthony Woodbury (UT Linguistics--used the pilot site as a proof of concept and they were awarded an NEH Digital Initiatives grant to build the permanent repository.
- The first permanent site launched Jan. 31, 2003.
- That grant was followed by others that focused on gathering together existing collections of academic linguistic and anthropological research on lgs of LA, digitizing & preserving them, and sharing them in AILLA's digital repository.

Mark Sicoli, David Yetman, Daniel Duncan and Marcelino Flores Mariscol. 2011. Map navigation Marcelino to Pedro 1. Sochiápam Chinantec Whistled Speech Collection of Mark Sicoli. Archive of the Indigenous Languages of Latin America. https://ailla.utexas.org/islandora/object/ailla:242953

Today the NSF, the NEH and other US & EU-based granting agencies refer researchers to AILLA for archiving Latin American language documentation data. Dozens of researchers archive their born-digital, multimedia research collections in AILLA every year …

Left: Filemón Díaz Sánchez and Damián López Gutiérrez. 2018. Catching Fish. Chol Collection of Juan Jesús Vázquez Álvarez and Jessica Coon. Archive of the Indigenous Languages of Latin America. https://ailla.utexas.org/islandora/object/ailla:261428. Right: Hilaria Cruz, Teodoro Leonardo Cruz, Gibrán Morales Carranza, and Natalia Salinas Cruz. 2012. Photographs. Chatino Documentation of Hilaria Cruz. Archive of the Indigenous Languages of Latin America. https://ailla.utexas.org/islandora/object/ailla:137620

… including Indigenous Peoples who take an active role documenting their own languages.

The AILLA directors and staff continue to get major grants to pursue our own projects to improve our collections:

Current NSF-funded project

Transforming Access and Archiving for Endangered Language Data Through Exploratory Methodologies of Curation BCS-1653380

ARCHIVING FOR THE FUTURE

Video still from *Sustainable File Types* https://youtu.be/2JCpg6ICrBM

We are finishing up one grant from the NSF called Transforming Access and Archiving for Endangered Language Data Through Exploratory Methodologies of Curation (BCS-1653380).
Over the past 3 years, we developed an archives-neutral workshop and animated video tutorials designed to teach language documenters how to organize their own language collections for deposit into a digital repository.

Online how-to course for archiving language documentation

Linguistic Data Curation Tutorials. Archive of the Indigenous Languages of Latin America, https://www.ailla.utexas.org/islandora/object/ailla%3A257379

Now we are turning that workshop and the video tutorials into an online course that language documenters can take as they have time. This site is still being built, and we are not yet ready to share the link. The course's landing page gives the user to select between the English or the Spanish versions of the course.

Current NEH-funded projects

Transcribing AILLA: Increasing Collection Access and Reusability through Crowdsourced Transcription (PW-259116-18)

Archiving Significant Collections of Endangered Languages: Two Multilingual Regions of Northwestern South America (PD-260978-18)

Top photo 2018 by Susan S. Kung; bottom photo 2018 by Ryan Sullivant

Currently, AILLA also has two projects funded by the NEH to improve our collections:
- One called "Transcribing AILLA: Increasing Collection Access and Reusability through Crowdsourced Transcription" is an effort to improve one of our most-used collections through crowdsourced transcription. (PW-259116-18). Ryan Sullivant will be telling us more about that project tomorrow.
- The other is "Archiving Significant Collections of Endangered Languages: Two Multilingual Regions of Northwestern South America" (PD-260978-18). This is a major digitization and preservation effort to gather together, digitize and archive eight significant collections from South America held by researchers from the US, the UK, Brazil, and France.

Left: PICT0819.jpg. Woman with baby on back close to Ingapirca. Cañar Peace Corps Collection. Archive of the Indigenous Languages of Latin America https://ailla.utexas.org/islandora/object/ailla:260223. Photo: Thomas Catchpole. Right: EGI-E1B1-033.jpg. Elsa Gómez-Imbert Collection. Archive of the Indigenous Languages of Latin America (forthcoming collection)

The two South American regions that are the focus of our digitization and collection project are Highland Ecuador and the Northwest Amazon region around the Brazil-Colombia border. Though separated by more than 700 miles and about 10,000 feet of elevation, these regions are both multilingual. Highland Ecuador is home to a number of distinct Quechuan languages (locally usually called Kichwa), and Northwest Amazon's practice of linguistic exogamy leads to very multilingual homes where children often hear languages from three or four different families in their houses growing up, to say nothing of each community's differing amounts of contact with Spanish and/or Portuguese. --Ryan Sullivant

Northwestern South America Collections

- Materials collected 1964-2010
- 4 language families represented across 8 collections
  - Tucanoan
  - Arawak
  - Quechua
  - Nadahup
- Total size of collections so far:
  - 13,810 manuscript pages
  - 2,775 images
  - 555 hours of audio
  - 53 hours of video

Right: Scan of handwritten document to appear in the AILLA collection of Stephen Hugh-Jones.
Archive of the Indigenous Languages of Latin America (forthcoming collection).
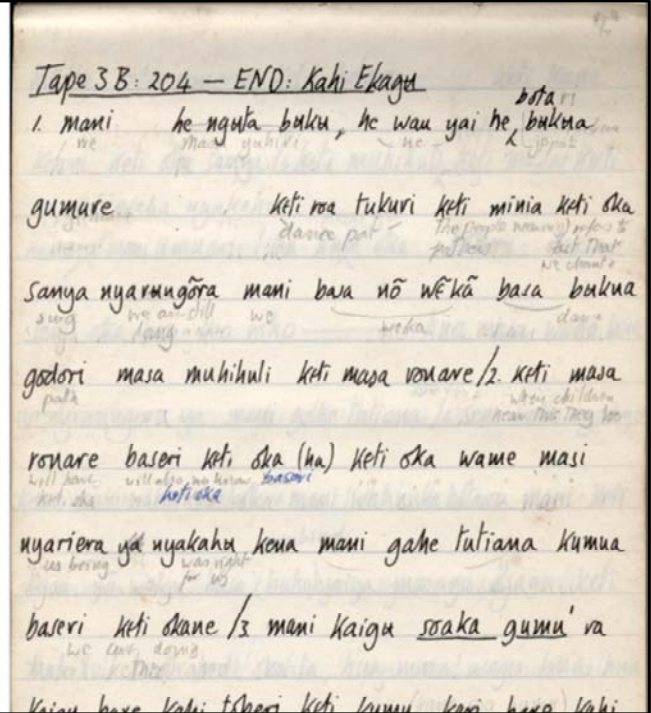
Only two collections have been fully processed and are public so far (both from Highland Ecuador), but processing is underway on the thousands of digital objects forming the remaining collections. We hope that these collections will help bring further attention to the cultures of these regions and will spur future study of these languages and the changes evident within them over the course of the last fifty years. --Ryan Sullivant

References

AILLA. 2018. Linguistic Data Curation Tutorials [Collection]. The Archive of the Indigenous Languages of Latin America, www.ailla.utexas.org. Access: Public. PID ailla:257379.

AILLA. 2018. Sustainable_File_Types-cc.mp4 [Video]. Data Curation Tutorial Videos – English. Linguistic Data Curation Tutorials [Collection]. The Archive of the Indigenous Languages of Latin America, www.ailla.utexas.org. Access: Public. PID ailla:257468.

Catchpole, Thomas. 1971. PICT0819.jpg [Photo]. Thomas Catchpole Photographs. Cañar Peace Corps Collection. The Archive of the Indigenous Languages of Latin America, www.ailla.utexas.org. Access: Public. PID ailla:260223.

Cruz, Hilaria. 2012. Photographs. Chatino Documentation of Hilaria Cruz. The Archive of the Indigenous Languages of Latin America, www.ailla.utexas.org. Access: Public. PID ailla:137620.

Sherzer, Joel. 1970. PB_KP04/tif [image]. Kuna Photograph Collection of Joel Sherzer. The Archive of the Indigenous Languages of Latin America, www.ailla.utexas.org. Access: Public. PID ailla:246379.

Sicoli, Mark. 2011. Map navigation Marcelino to Pedro 1. Sochiápam Chinantec Whistled Speech Collection of Mark Sicoli. The Archive of the Indigenous Languages of Latin America, www.ailla.utexas.org. Access: Public. PID ailla:242953.

Vázquez Álvarez, Juan Jesús & Jessica Coon. 2018. Catching fish. Chol Collectin of uan Jesús Vázquez Álvarez and Jessica Coon. The Archive of the Indigenous Languages of Latin America, www.ailla.utexas.org. Access: Public. PID ailla:261428.

Thank you