VRA EAC Community Hour

Topic: Found in Translation: Multilingualism in GLAMs

From metadata and descriptions to whole interface translations, multilingualism in GLAMs is an important step towards opening our collections to a broader range of users. Do you have a multilingual project you would like to share with us? Do you want to start a project but don't know where/how to begin? How can we build capacity for these projects in small and/or under-resourced repositories? Join us and repeat special guest Devon Murphy for an informal discussion/brainstorming session on multilingual support in our collections.

Guest: Devon Murphy

Moderator: Summer Shetenhelm

- Please keep your microphones muted--please use a reaction, raise your hand, or use the chat to indicate that you would like to speak.

- **Video not required**, virtual backgrounds welcome (be mindful of your virtual background selections). Feel free to add your pronouns to your Zoom name.
  - How to change your name during a Zoom meeting
  - How to provide captions via PowerPoint or Google Slides live captions
    - https://webcaptioner.com/ - Free Captioning Right In Your Browser

- If you'd like to ask a question anonymously, please send your question through a private chat to any moderators listed above.

- **Community Hours are not recorded** for the sake of attendees' privacy and so that all attendees feel safe to express any concerns or questions they may have on various topics.

Attendees are encouraged to take notes under the 'Community Notes' section of this document.

EAC webpage: http://vraweb.org/about/committees/equitableaction/
EAC email: ea@vraweb.org
Submit a topic for a future Community Hour: https://forms.gle/fENKfkEJjAC6Fqew5

Thank you! Visual Resources Association's Equitable Action Committee

---

Discussion Prompts:

- Do you have a multilingual project you would like to share with us?
- Do you want to start a project but don't know where/how to begin?
- How can we build capacity for these projects in small and/or under-resourced repositories?

Community Notes:
- Where UT Austin began, and where they are now for multilingual metadata.
- Website shows levels of metadata. Basic - only 1 to 3 fields translated, Intermediate - more than 3 fields in another language, Extensive - most or all of the record is translated. Some people on staff have the language proficiency, but sometimes have to hire someone who can translate. Decide which records need to be translated for public view, versus fields that are used in house.
  https://wikis.utexas.edu/display/utlmetadata/Multilingual+Metadata
- We are a Hispanic serving institution, so Spanish and Portuguese are languages used most often. Visitors come from Mexico, Argentina, Guatemala, other countries, and Spanish and Portuguese speakers who live in the US.
- Latin American Digital Initiatives - using Drupal 9. Drupal function called Views will translate English to other languages, including the field names.
- Focus on our subjects and genres. Use University of Florida for authoritative sources. Use terms that will be used in other Spanish thesauri.
- PANA Pan American Authorities Workflow. https://guides.uflib.ufl.edu/pana/workflow
- We meet as a group every month to discuss terms, building a spreadsheet with LCSH terms, local terms, the Spanish terms. UT_SPanish_Terms etc. Decide on final terms that are then shared with University of Florida , everyone is then using the same terms. From the University of Florida side, a group of the bilingual staff are now working with UT-Austin.
- Question: Would an individual record have to be indexed under both acuarelas and watercolors? Yes, it would be. In a fancier site, not necessary, but in our Arabic collections, and Southeast Asia collections yes.
- There are some older collections that did have some collections in multilingual data.
- 2021 was the time we started diving deeper into the multilingual data.

- Question: Are you using or storing ISO language codes? YES! ISO
- Also [AILLA Indigenous Languages of Latin America](#). We sometimes add new codes.
- Question: So you use 2 for common languages and 3 for indigenous? 3 more frequently.
- Question: Will screen readers recognize 3? I do not know!
- Question: Question: I know that if you have 3 for a rare language it does not have a voice engine.
- Friction between what we need collections to be able to do and what our technology is able to do or what our in-house technology is currently capable of
- Super early pilot stages of a digital humanities project. Tamil language issues due to Tamil sites. Currently copying the place names from Google Maps in Tamil script. Sometimes have to work with an outsourced vendor, if there is no one available on staff who is familiar with the language. Money becomes an issue
- Some places have hired students specifically for larger translation projects here. Of course, more and more student employees speak multiple languages. In fact, all 8 current student workers at one institution do speak multiple languages
- Translating a project using Google Translate?- Google translations for single terms can be okay (for example, Chinese is okay, but flow/grammatical structure is not great). It could be the first step, but definitely not the only step.
- The Heidelberg Library has done a language code project for Asian and South Asian languages.
- Has anyone else hired student workers or worked with students in classes to help with translation of materials? Partnering with a faculty member?
- One summer we had a student who spoke Arabic. The student was interested in translating data for 200 images.
- Right before I started at the library, there was a crowdsourced event for Arabic materials https://library.ucr.edu/about/news/arabic-manuscript-leaves-described-through-community-collaboration Any other events related to this? One follow up event, probably for the faculty member, and for community building.
- Digitizing a bilingual student newspaper. Optical character recognition in English & Spanish, though the repository is not yet fully text searchable. Getting the students to describe the collection was pretty cool. Not enough time and resources for complete bilingual metadata. https://calisphere.org/collections/27876/
- It is clear that for bespoke collections there is translation, compared to other collections that have been around longer. We need to go back and translate the sites that have migrated. Finding aids also need to be updated.
- Community building events to build multilingual collaboration.
- Not so much in terms of language, definitely for terminology, inconsistently, worked with donors, on terminology that they want to use. We are seeking out groups that can help with translation. Also university to university collaboration. https://calisphere.org/collections/27876/
- Other challenges? Challenging materials?
- Video has to have a transcript and captions. How would you translate a video, and provide captions and transcript? Large ask. The University requires captions and transcripts.

- How do you store the caption and the transcript? Stored in the same DAMS with the video. Indigenous languages are harder to find translators for.
- Separate objects of metadata.
- Bringing up IIIF - Any power to gather information from IIIF to use in your own database? Somebody who is working on storing maps of 1960's 1970's activism in Guatemala.
- Subject librarians will make exhibits for certain classes and use IIIF data.
- Do you use IIIF Change Discovery for updating records?
- [IIIF Change Discovery API 0.3](#) Project at Yale, use change discovery to update data in future.
- Yale has a lot of Indigenous and Cherokee manuscripts in its collection. Can take the IIIF manifests and work with individuals from the communities to translate the manuscripts.
- Point: There is no voice engine for Indigenous languages. Can you begin to give raw material to a process to improve discoverability?
- There are some language revitalization groups and projects, working with collection materials. Although be aware of intellectual property, the language belongs to the Indigenous community, if you are recording a speaker.
- The Welsh language is the only language with a relatively small number of speakers that has a voice engine.
- Ethical issues. What are the results of these community partnerships and translation projects? Does the community control the outcome?
- Multilingualism at your institution?
- ISO codes.
- Language Concepts - won't work with a screen reader, but helps with links.
- [Lux is the project at Yale](#)
- How are you providing screen reader support for non-English speakers if you're not using ISO language codes?
- RS (Rob Sanderson): The data has multiple languages, but the UI text is only in English. The language concepts do have ISO codes, so we can include them in the data as and when possible.
- Eg: https://lux-front-prd.collections.yale.edu/view/concept/7bf852bf-a0b0-4cab-ab31-15b5869a9bc7

Resources:

- [PANA (Pan-American Authorities) workflow: Background/Overview](#)
- [PANA (Pan-American Authorities Project)](#)
- [Presentation: Enabling and Reusing Multilingual Citizen Contributions in the Archival Record](#) [Starts at ~29:00 in downloadable .mp4 file]
- [White Paper on Best Practices for Multilingual Access to Digital Libraries](#)
- [Latin American Digital Initiatives](#)
- [Arabic manuscript leaves described through community collaboration](#)