



ETHICS & EVOLUTION OF
TRUTH
& INFORMATION
October 14-17 | Portland, Oregon

Developing Graduate Curriculum for Digital Language Archive Stewardship

ALISE 2024 juried paper presentation, 10/16/2024

Oksana L. Zavalina oksana.zavalina@unt.edu

& Hugh J. Paterson III hughpaterson@my.unt.edu

General Needs of Digital Language Archive Users

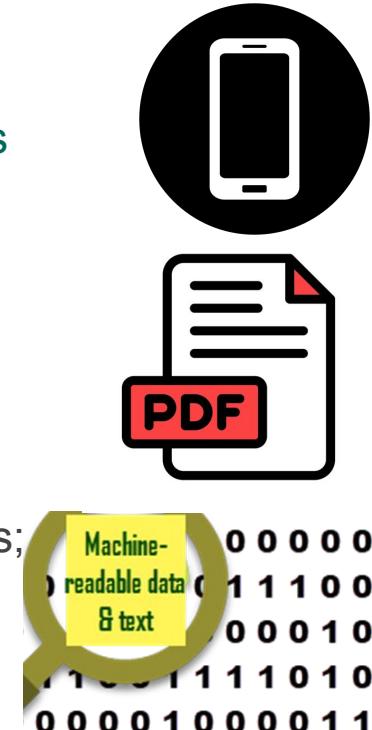
General information needs do not depend on the specific use purpose:

- The main general need is **discoverability** of information resources (achieved through providing quality **metadata**)
- **Metadata** should support **find**, **identify**, **select**, **obtain**, and **explore** user tasks.

Some examples of **digital language archives users' general needs** are for **services** available:

- Streamable audio and video with transcriptions and translations;
- User interface accessible on mobile devices;
- Downloadable machine-readable text files and bulk download option.

[Burke, Zavalina, Chelliah, & Phillips \(2022\)](#)



Specific Needs of Digital Language Archive Users

Community users have specific needs for supporting language revitalization ([Burke, 2023](#) findings with Boro language community):

- **Resources available:**
 - Dictionaries
 - Textbooks and teaching aids for different subjects
 - Storybooks for children; Folktales, stories
 - Resources on religion and culture
 - Attractive items that capture attention
- **Services available:**
 - Grouping resources by reading level, grade level, etc.



Current State of Meeting These User Needs: Linguists

Challenges for linguists interacting with digital language archives:

- Observations ([Burke et al., 2022](#)) identified deficiencies in completeness and accuracy of metadata as an important challenge:
 - *"This is a text...it's just a scan, not really accessible in any way other than hand transcription. So, in the cases where text does exist, it would be useful to denote the type of text. Some PDFs are searchable, and some are not."*
 - *"There are so many things tagged as text that are just an audio file....if I needed data, how would I systematically gather data? it would be a little difficult for me, given just the disparity of formats and availability."*



Language archive users

Current State of Meeting These User Needs: Communities

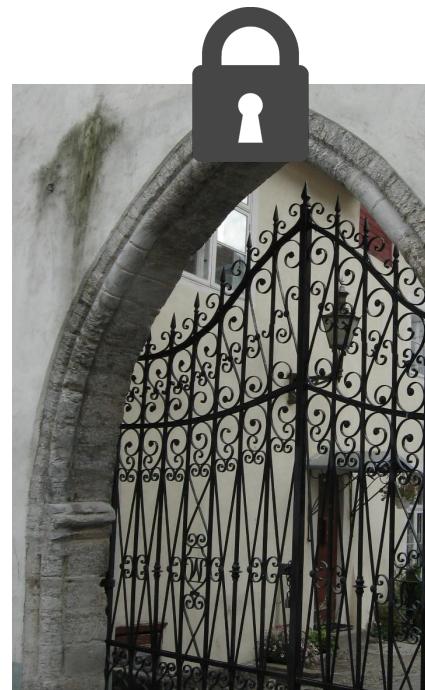


Language community members have reported access barriers (focus group):

Academics are building these archives...and so you build it for people like yourself. So, the door is an academic door, right? So other academics walk along and say, "Oh! I know how to open this door. And it's for me! And everything in there is for me!"

And for other people who are not academics, they look at these archives and they're like looking at tools from some foreign thing...the door isn't made for them

([Wasson et al., 2016](#), p. 675)



Need & Context

- Speakers of endangered and minority languages *along with those who support these people*, **need the services of memory institutions.**
- **How do we prepare information professionals** to align the service offerings of memory institutions with the currently-felt and long-term needs many endangered language using groups have?

Preparing Information Professionals to Help Support Language Archive User Needs

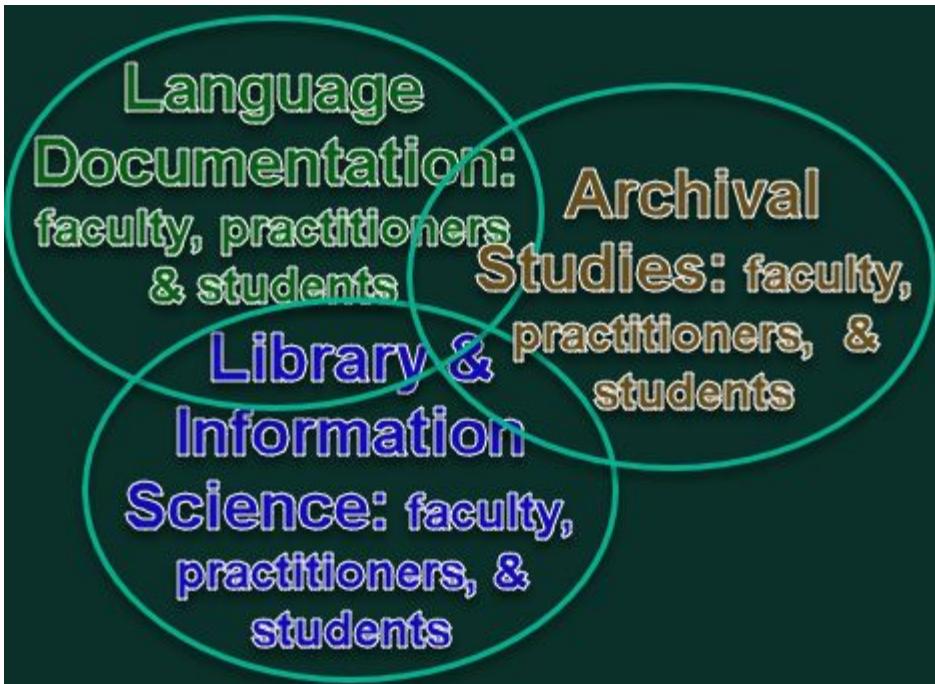
**Laura Bush 21st Century Librarian Program grant-funded project
RE-254860-OLS-23**



- collaboration between CoRSAL digital language archive, UNT College of Information, UNT Libraries, IU Linguistics Department
- Developing (2023-2025) open source modular curriculum with a strong practical component to educate information professionals in the archiving and curation of resources that provide the means to **revitalize community memory and language**.
- Learning materials suitable for use in:
 - courses offered in Library & Information Science, Archival Studies university programs
 - on-the-job trainings of library/museum/archive staff working with language collections



Interdisciplinary Project Team



- Co-PI **Dr. Shobhana Chelliah** (Linguistics, Indiana University)
- Co-PI **Dr. Mark E. Phillips** (Library & Information Science: digital repositories, University of North Texas - UNT)
- PI **Dr. Oksana Zavalina** (Library & Information Science: information organization, UNT)
- **Dr. Ana Roeschley** (Archival Studies, UNT)
- **Dr. Brian C. O'Connor** (Library & Information Science: digital Imaging, UNT)
- **Graduate students** (Information Science + Linguistics):
 - Sergio I. Coronado, Merrion Frederick, Hugh Paterson III at UNT
 - Alexandra O'Neil at IU

Community Language Archiving and Curation for Information Professionals Course: Overall Structure (1)

Learning modules	Learning objectives (top-level)
M1: Planning, developing, and managing a community language archive	<ol style="list-style-type: none">1. Define community language archives and their functions.2. Describe procedures and considerations for community language archive planning and development.
M2: Ethical archival practices and digital curation in community language archives.	<ol style="list-style-type: none">3. Examine archival theory and practice and digital curation trends and perspectives relevant for community language archives.4. Explain important access, preservation, and description issues and practical problems associated with community language archives.

Community Language Archiving and Curation for Information Professionals Course: Overall Structure (2)

Learning modules	Learning objectives (top-level)
M3: Metadata, digital content management and web archiving for community language archives	<ol style="list-style-type: none"><li data-bbox="547 372 1843 556">5. Identify metadata standards that can be utilized in representing community language archive materials to support information needs and user tasks of community language archive users.<li data-bbox="547 556 1843 729">6. Describe the problems related to selection of digital content management tools, depositing and web archiving for community language archives.
M4: Dissemination and use of community language archive content, evaluation of archival services	<ol style="list-style-type: none"><li data-bbox="547 729 1843 847">7. Identify efficient and ethical ways of disseminating the content of community language archives.<li data-bbox="547 847 1843 1022">8. Discuss approaches for the evaluation of the services provided by community language archives.

Training with Strong Experiential Component

(all 4 assignments are practical or real-life analytical): One Example

In Module 1, students plan, create -- by following best practices for language archives -- and document the process of creation of a **mini-collection of 4 born-digital & digitized archive items based on student's own 5-words wordlist:**

- 1 image, 1 text, 1 video, & 1 audio

Summer 2024 students' collections represented **12 languages/dialects and various topics:**



Project-Developed Course Testing at UNT

Target audience at UNT:

current and future
graduate students in
Master of Library
Science, Master of
Information Science, &
Ph.D. in Information
Science programs
(including but not limited
to Archival Studies
concentration, Linguistics
concentration)

*Developing learning
materials, getting ADA &
copyright compliance
approvals*

1st offering:

Summer 2024
5-week semester
as a section of
INFO 5680
seminar, with 11
federally-funded
stipends (23
students)

2nd offering:

Summer 2025
10-week semester
as INFO 5385
specialized course,
with 14
federally-funded
stipends

Broader audience for this training: Information Professionals in the following Roles

- ※ Digital archivists and other informational professionals
 - responsible for **inventory** and **implementation of data management plans**
 - working with university faculty on **adding legacy data collections to repositories**
 - whose responsibilities include **establishing a new archive** and/or development of existing digital collections
 - working on system/repository **migration tasks**.
- ※ Programming Librarians / Exhibit Production Specialists
- ※ Special Collections Archivists, Curators, and Collection Maintenance Specialists

Topics to emphasize in DLA stewardship training for LAM professionals

1. Fiduciary responsibility
2. Language identification and representation
3. Analysis and representation of **relations between works, their versions, formats, etc.**
 - Including but not limited to **part-whole relationships** representation
4. Resource **of-ness, interactive modality and materiality** representation
5. Evaluating community language archives:
 - Including their services, technical platforms, content description and arrangement, discovery, and use

Fiduciary responsibility topic: Example

1. *DEFINITION IS NEEDED here*
2. *Add example here from the course content or from a language archive*

Language identification and representation topic: Example

Students learn to navigate, compare and contrast, and apply in practical tasks the metadata standards, including:

- Controlled vocabularies for language, language group, and dialect representation: [ISO 639-3](#), [marclanguage](#), [Glottolog](#), [AUSTLANG](#)
- Guidelines for language representation included in the Open Language Archive Community [\(OLAC\) metadata scheme](#), [CoDA curriculum metadata guidelines](#) for CoRSAL digital language archive
- Other relevant OLAC controlled vocabularies: [Discourse type](#), [Linguistic data type](#), [Linguistic subject](#), [Role](#)

Inter- and Intra-work relations analysis & representation topic: Example

Language archive resources are highly interrelated, e.g.:

- a video recording of the conversation about,
- its audio recording,
- regular text transcript of it,
- translation and linguistic annotation of that transcript
- photographs taken during this event, etc.

Students learn about 2 alternative approaches to representing language archive materials: individual (Dublin Core) and bundle (EMELD)

In one of the 4 major assignments, students explore typical language archive items -- including an item they select from their own digital language collection (created in Module 1) -- for:

- WEMI entities, attributes and relations (based on IFLA's [FRBR](#) and [LRM](#) models)
- Degree to which user tasks are supported by representing these attributes, relations, and entities

Resource of-ness, interactive modality and materiality representation topic: Example

1. *Add example here from the course content or from a language archive*

Evaluation topic: Example of Practical Skills Developed

In Module 4 assignment, each student:

- examines a finding aid of a digital language archive collection (self-selected from the provided list) and compares it with general guidelines in [DACS archival standard](#) and specific guidelines for language collection guides in [Sullivant \(2020\)](#).
- selects 2 collections within the [Computational Resource for South Asian Languages \(CoRSAL\) digital language archive](#) hosted by the University of North Texas Libraries -- and 2 items from within each of these collections -- to:
 - comparatively examine patterns of use for the selected collections and items (based on several usage metrics)
 - discuss factors that could have contributed to relatively higher popularity of the most popular item in this small sample.

Some Findings of Course Surveys for Summer 2024



The level of learning-objective-related confidence developed by students:

- Meets & exceeds our expectations for course-level objectives (**91%-100%**)
- Overall, is as expected for module-level objectives (**70%+**)

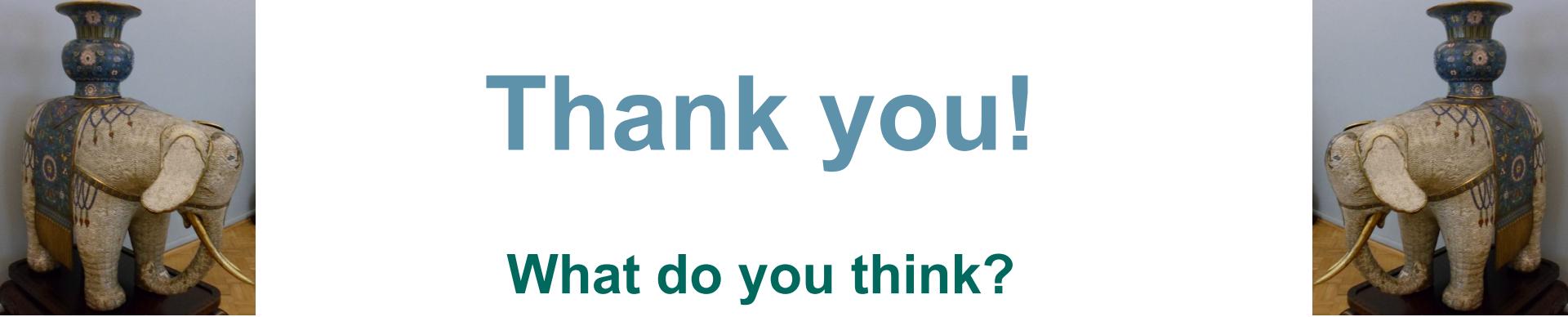
Learning objectives that were addressed not only by instructor presentations but also by module assignments, tended to be met by higher percentage of students

- **Conclusion:** Need to revise/expand budgeting & funding content in M1 instructor presentation, and content on Digital Content Management tools in M3 instructor presentation, possibly add components focusing on these issues in the module assignments
 - **Please share with us any ideas you have!!!**

Next Steps



- Obtain LIS profession & language community feedback for refining project-developed learning content
- Develop 10-week & 16-week versions of the course with more detailed content and more tasks in practical assignments
- Integrate project-developed content in other LIS courses, including:
 - [UNT INFO 5224](#) Advanced Metadata (Modules 2-3 continued integration since Spring 2024)
 - [UNT INFO 5960 Cultural Heritage Stewardship](#) (Module 2 continued integration since Spring 2024)
 - [UNT INFO 5742](#) Web Archiving (Module 3 integration planned for Spring 2025)
- Finalize project-developed learning materials and make available for anyone interested as open source [CC BY NC 4.0](#) materials (through [UNT Scholarly Works](#) repository)



Thank you!

What do you think?

Please share your feedback with our project team:

In-person during ALISE 2024

and/or any time via contact information on the project website



Works Cited in this presentation

- Bird, S., & Simons, G. F. (2022). The Open Language Archives Community: A 20-year update. *The Electronic Library*, 40(5), 507–524. <https://doi.org/10.1108/EL-08-2022-0192>
- Burke, M., & Zavalina, O. L. (2019). Exploration of information organization in language archives. *Proceedings of the Association for Information Science and Technology*, 56(1), 364–367. <https://doi.org/10.1002/pra2.30>
- IFLA Study Group on the Functional Requirements for Bibliographic Records & Plassard, Marie-France. (1998). Functional Requirements for Bibliographic Records: Final Report (19; 2nd ed., [UBCIM Publications, New Series] IFLA Series on Bibliographic Control). K.G. Saur. <https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr.pdf>
- Riva, P., Le Bœuf, P., & Žumer, M. (Eds.). (2017). IFLA Library Reference Model: A Conceptual Model for Bibliographic Information (December 2017). International Federation of Library Associations and Institutions (IFLA). <https://www.ifla.org/publications/node/11412>
-
- Burke, M. (2023). *Designing Archival Collections to Support Language Revitalization: Case Study of the Boro Language Resource*. [Doctoral dissertation, University of North Texas-Denton]. UNT Libraries.
- Burke, M., Zavalina, O. L., Chelliah, S., & Phillips, M. (2022). *User needs in language archives: Findings from interviews with language archive managers, depositors, and end-users*. *Language Documentation & Conservation*, 16, 1-24.
- Cook, T. (2013). *Evidence, memory, identity, and community: Four shifting archival paradigms*. *Archival Science*, 13(2–3), 95–120.
- Dale, M., Basumtary, P., Iqbal, J., & Shaikh, M. (2022). *Case study of using Facebook groups to connect community users with archived CoRSAL content*. *Language Documentation & Conservation*, 16, 399-416.
- Henke, R., & Berez-Kroeker, A. L. (2016). *A brief history of archiving in language documentation, with an annotated bibliography*. *Language Documentation & Conservation*, 10, 411–457.