

APPLICATION FOR FEDERAL DOMESTIC ASSISTANCE - Short Organizational				
* 1. NAME OF FEDERAL AGENCY: Institute of Museum and Library Services				
* 2. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER: 45.312				
CFDA TITLE: National Leadership Grants				
* 3. DATE RECEIVED BY STATE	SYSTEM USE ONLY			
* 4. FUNDING OPPORTUNITY NUMBER: NLG-LIBRARIES-FY20				
* TITLE: National Leadership Grants for Libraries				
<b>5. APPLICANT INFORMATION</b>				
* a. Legal Name: University of North Texas				
b. Address:				
* Street 1: 1155 Union Circle #305250	Street 2:			
* City: Denton	* County: Denton			
* State: TX: Texas	* Zip/Postal Code: 76203-5017			
* Country: USA: UNITED STATES				
c. Web Address: http://				
* d. Type of Applicant: Select Applicant Type Codes(s): H: Public/State Controlled Institution of Higher Education Type of Applicant: Type of Applicant: * Other (specify):	* e. Employer/Taxpayer Identification Number (EIN/TIN): 756002149 * f. Organizational DUNS: 614168995 * g. Congressional District of Applicant TX-026			
<b>6. PROJECT INFORMATION</b>				
* a. Project Title: A Metadata Approach to Addressing Tasks of Linguistic Data Users				
* b. Project Description: A team of linguistics, librarians, and information scientists at the University of North Texas seek funding for \$149,727 for a two-year IMLS National Leadership, National Forum Grant. The team will bring new levels of access through enhanced metadata to digital language resources housed at digital language collections throughout the country. The team will determine what metadata enhancements would meet the needs of users, using two types of language deposits as case studies and discussions with language archive users, depositors, and metadata creators. Digital language archives are often the last access point for language and culture for a community. The data in these archives is the source/validation point for language-based discoveries for many scientific fields. We seek to make language archives maximally useful so they can serve the needs of users.				
c. Proposed Project:	* Start Date: 09/01/2020      * End Date: 08/31/2022			
<b>7. PROJECT DIRECTOR</b>				
Prefix: Dr.	* First Name: Shobhana	Middle Name: L	* Last Name: Chelliah	Suffix:
* Title: Professor	* Email: Shobhana.Chelliah@unt.edu			
* Telephone Number: 940-565-4458	Fax Number: 940-565-4355			
* Street 1: 1155 Union Circle, Box 305298	Street 2:			
* City: Denton	County: Denton			
* State: TX: Texas	* Zip/Postal Code: 76203-5017			
* Country: USA: UNITED STATES				
<b>8. PRIMARY CONTACT/GRANTS ADMINISTRATOR</b>				
<input type="radio"/> Same as Project Director (skip to item 9):				
Prefix:	* First Name: Kasia	Middle Name: Veronica	* Last Name: Milewski	Suffix:
* Title: Grants & Contracts Specialist	* Email: Kasia.Milewski@unt.edu			
* Telephone Number: 940-565-4596	Fax Number:			
* Street 1: 1155 Union Circle #305250	Street 2:			

* City: Denton	County: Denton
* State: TX: Texas	* Zip/Postal Code: 76203-5017
* Country: USA: UNITED STATES	
9. *By signing this application, I certify (1) to the statements contained in the list of certifications** and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances** and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 218, Section 1001)	
**I Agree ●	
** The list of certifications and assurances, or an internet site where you may obtain this list, is contained in the announcement or agency specific instructions.	
<b>AUTHORIZED REPRESENTATIVE</b>	
Prefix:	* First Name: Middle Name: * Last Name: Suffix: Carla McGuire
* Title: Director Pre-Award and Contracts	* Email: Carla.McGuire@unt.edu
* Telephone Number: 940-565-3940	Fax Number:
* Signature of Authorized Representative:	* Date Signed:

## Attachments Form

**Instructions:** On this form, you will attach the various files that make up your grant application. Please consult with the appropriate Agency Guidelines for more information about each needed file. Please remember that any files you attach must be in the document format and named as specified in the Guidelines.

**Important:** Please attach your files in the proper sequence. See the appropriate Agency Guidelines for details.

- |                                 |                                    |                            |
|---------------------------------|------------------------------------|----------------------------|
| 1) Please attach Attachment 1   | IMLS_SupplementaryChelliah10086767 | Mime Type: application/pdf |
|                                 | 93.pdf                             |                            |
| 2) Please attach Attachment 2   | IMLS_LibraryChelliah1008676761.pdf | Mime Type: application/pdf |
| 3) Please attach Attachment 3   | Preliminaryproposal1008676762.pdf  | Mime Type: application/pdf |
| 4) Please attach Attachment 4   |                                    |                            |
| 5) Please attach Attachment 5   |                                    |                            |
| 6) Please attach Attachment 6   |                                    |                            |
| 7) Please attach Attachment 7   |                                    |                            |
| 8) Please attach Attachment 8   |                                    |                            |
| 9) Please attach Attachment 9   |                                    |                            |
| 10) Please attach Attachment 10 |                                    |                            |
| 11) Please attach Attachment 11 |                                    |                            |
| 12) Please attach Attachment 12 |                                    |                            |
| 13) Please attach Attachment 13 |                                    |                            |
| 14) Please attach Attachment 14 |                                    |                            |
| 15) Please attach Attachment 15 |                                    |                            |

# IMLS Supplementary Information Form

PLEASE NOTE: Information contained within this form may be made publicly available.

## Applicant Information

The organizational unit performing the activities described in the application is the same as the legal applicant from the SF-424 Short Organizational form.  Yes  No

Select the institution type that most accurately describes the legal applicant:

Graduate School of Library and Information Science

## Funding Request

IMLS funds requested	149,727.00
Cost share/match amount*	0.00
Total costs	149,727.00

\* Enter \$0 if the budget includes no cost share or match.

## Indirect Costs

Current indirect cost rate(s) have been negotiated with a federal agency.

Select from list: Department of Health and Human Services

Rate: 48.50

Expiration Date: 08/31/2020

Indirect cost proposal has been submitted to a federal agency but not yet finalized.

Select from list:

Rate:

Proposal Date:

Applicant chooses a rate not to exceed 10% of modified total direct costs, and declares it is eligible for the 10% rate.

Applicant chooses not to include indirect costs.

## Population(s) Served

Select the target population(s) to be served by the activities described in the application. Choose all that apply.

- General population
- Early childhood/Preschool (0-5 years)
- Middle childhood/Primary school (6-12 years)
- Adolescents/High school (13-19 years)
- Adults
- Aging, elderly, senior citizens (65+ years)
- Ethnic or racial minority populations other than Native American/Native Hawaiians
- Families/Intergenerational
- Immigrants/refugees
- Military families
- Museum and/or library professionals
- Native Americans/Alaska natives/  
Native Hawaiians
- People with mental or physical challenges/  
disabilities
- People who are low income/economically  
disadvantaged
- Rural populations
- Scholars/researchers
- Unemployed
- Urban populations
- Other

## Abstract

Enter or paste your abstract below (maximum 3000 characters).

PLEASE NOTE: Information contained within this form may be made publicly available.

A team of linguistics, librarians, and information scientists at the University of North Texas seek funding for \$149,727 for a two-year IMLS National Leadership, National Forum Grant. The team will bring new levels of access through enhanced metadata to digital language resources housed at digital language collections throughout the country. The team will determine what metadata enhancements would meet the needs of users, using two types of language deposits as case studies and discussions with language archive users, depositors, and metadata creators. Digital language archives are often the last access point for language and culture for a community. The data in these archives is the source/validation point for language-based discoveries for many scientific fields. We seek to make language archives maximally useful so they can serve the needs of users.

## IMLS LIBRARY - DISCRETIONARY PROGRAM INFORMATION FORM

PLEASE NOTE: Information contained within this form may be published online or otherwise shared by IMLS; therefore, please do not include any sensitive, proprietary, or confidential information.

### 1. Agency-Level Goals and Objectives

Select the IMLS agency-level goal that best aligns with your proposed project. Once you have selected a goal, please select an associated objective.

**Promote Lifelong Learning**

**Build Capacity**

**Increase Public Access**

Support Stewardship of Museum and Library Collections

Tools, Technology, and Training Enabling Discovery and Use of Collections and Resources by All

Policies and Partnerships to Address Access Barriers to Collections, Programs, and Information

Increased Access to Knowledge through Effective Communications

## 2. Grant Program

Choose the program to which you are applying. If applicable, you will then be asked to select one choice per mandatory category for the chosen program. Refer to the Notice of Funding Opportunity for descriptions of these options and instructions for how to make selections.

**Laura Bush 21st Century Librarian Program**

**National Leadership Grants for Libraries**

1. Select one:

- Planning
- National Forum
- Project
- Research in Service to Practice

2. Select one:

- Lifelong Learning
- Community Catalysts
- National Digital Infrastructures and Initiatives

**Native American Library Services: Enhancement Grants**

**Native Hawaiian Library Services**

**Native American Library Services: Basic Grants**



**Overview:** A team of linguistics, librarians, and information scientists at the University of North Texas (UNT) seek funding in the amount of \$149,727 for a two-year IMLS National Leadership, National Forum Grant. With this grant, the team will bring new levels of access, through enhanced metadata, to digital language resources housed at digital language collections throughout the country. The project team will work with an Advisory Board and a graduate student RA to determine what metadata enhancements would meet the needs of users, using two types of language deposits as case studies. To determine those needs, the team will meet four times with language archive users, depositors, and metadata creators.

Linguists are increasingly working with libraries (e.g., academic, tribal, museum) to provide access to analog and digital versions of language datasets. The resulting language archives are a critical resource for Native American groups working on language revitalization and for linguists learning from and disseminating information to the general public on major and low-resource languages. These digital archives are often the last access point for language and culture for a community.<sup>1</sup> They are the data source/validation point for language-based discoveries for many scientific fields. When compared to other digital archives, however, their usage level is low. This is primarily, as we have learned through our ongoing IMLS-funded survey of language archives, due to scarce metadata. Through our project, we seek to make language archives maximally useful so they can serve the needs of users and their communities. Such communities include tribal members and students/researchers at US universities.

**Statement of Need:** As technologies have become increasingly available, communities working to preserve linguistic and cultural heritage are depositing audio, video, and text as digital objects to language archives. Depositors include Native American groups working on language revitalization and linguists documenting languages for their cultural and scientific value. These deposits are a part of digital collections of academic libraries or dedicated language archives in library systems, such as the California Language Archive (CLA) at the University of California Berkeley or the Alaska Native Languages Archive (ANLA) at the University of Alaska, Fairbanks. Funding agencies (e.g., NSF) require these deposits and professional organizations (e.g., Linguistic Society of America) encourage such deposits. The amount of data being ingested to language archives is thus increasing exponentially. Digital language datasets have the potential to make a strong contribution to social science research and education (e.g., Language Science, Geography, History, Sociology) and Computer Science (e.g., natural language processing). However, this potential remains largely unrealized, as language archives are rarely accessed by user communities. This is because the metadata that accompanies language data deposits does not fully serve the needs of potential users. For example, if a Karuk language teacher, who may not be an academic, wants to create a lesson with CLA data sets on how to ask questions, give directions, or perform a blessing, there is no metadata to filter the Karuk collection for that information. This same lack of adequate representation for content of information objects is also seen in language deposits of major languages, such as English. We propose to investigate, with communities of users, the levels of granularity and coverage in metadata that will help users find, identify, select, obtain, and explore language materials. For example, we will target user tasks that metadata records must serve (according to the functional requirements of the International Federation of Library Associations and Institutions (IFLA) Library Reference Model (LRM)).

**Project Design and Timeline:** *Fall 2020:* (1) The project team and graduate student will review results of an IMLS-funded project that co-PI Zavalina conducted. Zavalina's project analyzed the content of metadata used in 20 languages archives. (2) We will (a) create a description of the two common datasets, an Annotated Text corpus and a Squib corpus. Annotated texts are instances of connected, transcribed speech, such as folktales labeled for meaning at the clause, word, and morpheme level. These instances are found in all digital language archives. The Squib corpus comes from the collected papers of the influential linguist John (Haj) Robert Ross and can be found here: <http://www-personal.umich.edu/~jlawler/haj/Squibnet>. Permission to use representative datasets has already been secured. (b) We will create a project website to disseminate findings as we progress through our grant period. *Spring 2021:* (3) We will hold Meeting 1 with an Advisory Board. The board will be comprised of key stakeholder community representatives (e.g., users of language collections from linguistics, Native American studies, geography, history, sociology), metadata practice experts from library and linguistics data environments, metadata standards developers, technologists and natural language processing experts, and managers of linguistic

---

<sup>1</sup><http://miamioh.edu/myaamia-center/breath-of-life/about/history/index.html>

The University of North Texas

dataset collections and archives. At the meeting we will (a) review user tasks and a related concept base in the LRM conceptual model in relation to the two types of language datasets and existing metadata used in language archives. Specifically, we will identify key user tasks of potential language dataset users. We will also assess how much these tasks are supported by the available general metadata standards (e.g., MARC, Dublin Core, MODS, etc.), more specialized standards (e.g., OLAC, IMDI), and current controlled vocabularies (e.g., lists of terms and codes, thesauri, taxonomies for linguistic data). (b) We will create plans for user-task research to inform metadata enhancements. Specifically, we will explore how each of the general user tasks (find, identify, select, obtain, and explore) are supported in typical language-archive use scenarios by existing metadata (e.g., through mapping specific metadata elements to user tasks and determining the level of support for each user task). We will also explore what other, specific user tasks need to be supported by metadata that describe language datasets (including mapping metadata elements to these tasks). *Spring through Summer 2021:* (4) Using the guidance from Meeting 1, we will continue to implement user-task research on the two selected data sets and on language materials ingested to the UNT digital library to refine suggested metadata enhancements. *Fall 2021:* (5) We will hold Meeting 2 at UNT. Participants will be users of language archives and creators of metadata for representing language materials. With these participants, we will (a) review findings and recommendations from the user-task research for metadata enhancements; (b) use proposed metadata enhancements (including controlled vocabularies) to represent Squib and Annotated Texts objects; (c) refine proposed metadata enhancements based on the participants' experiences. (6) Following Meeting 2, we will (a) discuss the feasibility of implementing these enhancements with existing metadata standards; (b) explore the use of automatic processing methods (e.g., natural language processing) for capturing and creating the data values for the enhanced metadata elements. *Spring 2022:* (7) We will hold a Meeting 3 (virtual event): Participants will be members of the Advisory Board and project staff. They will create metadata guidelines for depositors and metadata creators. (8) For the culminating event, we will (a) present the results (user-study findings, metadata enhancements including controlled vocabularies, metadata usability for the datasets) of this project at Meeting 4, which will be held in conjunction with a relevant conference (such as the annual meeting of the Association for Tribal Archives and Museums or the International Conference on Language Documentation and Conservation held in Hawaii). Participants will be the Advisory Board, the project Team, and invited presenters (depositors and end-users, metadata developers and implementers). *Spring to Summer 2022:* (9) We will share our findings and metadata enhancements via knowledge bases, such as GitHub and Wikidata, and social media to encourage comments and suggestions. All project documents will be available through UNT's open access Scholarly Works Repository and its Data Repository.

**Outcomes and National Impact:** Falling under the project category of National Digital Infrastructure and Initiatives, the project will build on user-task understanding and pilot methods and tools for enhancing existing metadata and controlled vocabularies. Applying these enhancements will give us the potential to exponentially increase the usability of digital language collections. A common cause of concern and source of slowness in language revitalization for Native American communities has been the lack of non-technical access to language material. The impact of enhanced metadata and controlled vocabularies will greatly support language revitalization goals. We similarly strive to provide access to "digital content and collection at scale for uses of all interests and skill levels" by metadata enhancements arrived at through this project.

**Project Team:** The project team includes PI Dr. Shobhana Chelliah (Professor, Linguistics, and Associate Dean for Research in UNT's College of Information); Co-PIs Dr. Haj Ross (Professor, Linguistics); Drs. Oksana Zavalina and William Moen (Associate Professors, Information Science); and Dr. Daniel Alemneh (Digital Curation Unit, UNT Libraries). We will hire a student in the Information Science PhD program, with a concentration in Linguistics. Our team members have successfully collaborated in the past, and they will bring the extensive research and practical expertise needed for this project (e.g., metadata and controlled vocabulary development and evaluation, user behavior research, and linguistics research). Our team also has experience with successful planning and implementing projects funded by IMLS and other funding agencies.

**Budget:** We request \$149,727 from IMLS for a two-year National Forum Grant. The indirect costs will be \$43,631, as per UNT's federally negotiated indirect cost rate (Dept. Health and Human Services, 06/19/2015). Funds will support a graduate research assistant for two years (50% time for 11.5 months) for a total of \$29,520 plus \$15,940 in benefits (at 8.65%) and \$16,136 in tuition reimbursement. We request \$44,500 to support four meetings.