Museums for America Grants

Sample Application MA-245543-OMS-20
Project Category: Collections Stewardship and Public Access

University of Oklahoma (Sam Noble Oklahoma Museum of Natural History)

Amount awarded by IMLS: $73,882
Amount of cost share: $74,283

The project description can be viewed in the IMLS Awarded Grants Search:
https://www.imls.gov/grants/awarded/ma-245543-oms-20

Attached are the following components excerpted from the original application.

- Narrative
- Schedule of Completion

Please note that the instructions for preparing applications for the FY2021 Museums for America grant program differ from those that guided the preparation of FY2020 applications. Be sure to use the instructions in the Notice of Funding Opportunity for the grant program and project category to which you are applying.

Project Justification

The Native American Languages collection (NAL: http://samnoblemuseum.ou.edu/collections-and-research/native-american-languages/) at the Sam Noble Oklahoma Museum of Natural History (SNM) at the University of Oklahoma is requesting urgently needed funds for new digital infrastructure to safeguard and provide access to the collection into the future. Our hard drives and physical systems which store the collection have not been updated since the collection was founded in 2003. We are rapidly exceeding both our digital storage capacity and the backwards compatibility of our hardware, and are therefore requesting $80,142 in funds from IMLS to update our systems, which the exceeds the capabilities of both the collection and the museum to provide.

NAL is an archival repository for materials in and pertaining to the indigenous languages of the Americas, with a specialization in the indigenous languages of the central United States. The materials in the NAL collection are an irreplaceable resource on the history and traditions of Native America, which are of enduring value to the communities they come from and the world at large. The collection centers around protecting the cultural heritage of Native peoples across the continent which continues to be threatened by social, political and economic forces which have stemmed from settler colonialism, relocation and genocide. *All* the indigenous languages of the Americas are either no longer spoken or endangered, many critically so, which is especially pronounced in communities like those in Oklahoma that have weathered dislocation from their ancestral lands. The advanced state of endangerment for many of the languages best represented in NAL makes the information we curate particularly rare and valuable.

Figure 1. A letter from Durbin Feeling to his family from Vietnam, in Cherokee (November 3rd, 1968)

The collection currently contains approximately 8,500 items representing 316 languages. This includes primary materials such as field notes, audio/video recordings, and correspondence, as well as secondary materials like datasets, pedagogical materials, descriptive materials, translations, transcriptions, and other media productions. A sample collection item is given in Figure 1. While we are still in the process of producing comprehensive finding aids for our collections, more information on ~30 of our 56 current processed collections can be found by visiting http://samnoblemuseum.ou.edu/collections-and-research/native-american-languages/native-american-languages-collections/.

Within the SNM, the collection's physical spaces include a climate-controlled archival space, a digitization lab, a recording studio, and a reading room, which houses a reference library with supplementary materials on Native American languages, linguistics, archaeology, and ethnology. The collection consists of both born-digital items and ~140 linear feet of analog materials. All physical materials in the collection get digitized, so the collection is effectively entirely digital, requiring 10TB of storage space at present, with another ~1TB in known incoming accessions.

Each year the collection receives about ~100 in-person visitors, as well as over 2,000 requests for copies of materials (often many items per request). We also have thousands of people from all over the country participate in our annual public events, the largest being the Oklahoma Native American Youth Language Fair, which attracts about 3,000 people over two days (http://samnoblemuseum.ou.edu/collections-and-research/native-american-languages/oklahoma-native-american-youth-language-fair/). The largest user demographic for the collection is Native people—researchers, students, and members of the surrounding community. Collection materials are also regularly utilized by scholars (Native and non-Native), and we get requests topics ranging from language to biodiversity to traditional ecological knowledge to history to food sovereignty.

NAL follows best practices in all curation activities. Members of our staff are involved with several archival organizations which help set best practices for the field (e.g., the Society of American Archivists, the Society of Southwest Archivists). NAL is also a member of DELAMAN (Digital Endangered Languages and

Music Archives Network), a language archive consortium that has provided guidelines for NSF and the Linguistic Society of America on the preservation of language documentation corpora (http://www.delaman.org/). Additionally, the Sam Noble Museum adheres to all protocols associated with maintaining accreditation from the American Alliance of Museums.

Statement of need

The digital infrastructure which houses the entire NAL collection is both too old and too small to meet with current needs and best practices. The digital collection is currently stored across a single set of hard drives (5 G-Tech and 14 SATA hard drives, physically labeled for what is on each of them, see Figure 2), that range in age from 5-14 years old. Some of these drives are so old that they are no longer compatible with most of the computers we have available for our archivists and for our visitors, meaning we regularly have to transport ourselves, our master copy hard drives, and our guests between different collection spaces (on different floors) in order to access digital collection materials.

The hard drives that currently house the collection are mirrored (one set of copies on the same drives), which is our only backup system. Without any other backup, the collection is exposed to catastrophic data loss. This is made worse by the need to constantly transport these spinning hard drives and read from the master copies, practices which increase the possibility of data

Figure 2. NAL G-RAID 1

corruption. Additionally, we have only about 2.4TB of remaining space on those hard drives. Given that the collection has been growing by ~2TB every year in recent years, and that file sizes have been increasing as audio-visual technologies improve, the lack of secure digital storage is an immediate and growing problem which must be addressed soon.

In order to operate effectively, the collection needs a better digital storage system, with enough room to accommodate growth of the collection for the lifetime of that system. Such a system would need to provide a way for files stored on different hard drives to be searched, added to, and read from centrally, and a way to access the materials from anywhere in the museum (i.e., network attached storage (NAS)). Additionally, there needs to be a long-term, archival-quality off-site backup of the collection to prevent against catastrophic loss. The need for these infrastructure components is not only based on the observations of NAL staff, but also on best practices in digital curation set by the field ((see e.g., the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation (https://ndsa.org/activities/levels-of-digital-preservation/); Open Archival Information System (OAIS) Recommended Practice (https://public.ccsds.org/pubs/650x0m2.pdf), notably section 4; also the NARA Trustworthy Repositories Audit and Certification: Criteria and Checklist (http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf), notably section C).

Digital infrastructure

To address our current storage and preservation needs, NAL is requesting funds for new NAS systems to house the collection, an LTO tape backup system, a computer system with checksum software for managing these new infrastructure components, and a student worker to assist in data migration.

In terms of storage, we are looking at two Dell PowerEdge R540 servers striped at RAID 10 (which have 48TB of usable space on each server). With RAID 10, two hard drives can be lost or damaged and no information will be lost, ensuring the safety of the collection in the case of mechanical failures. One server will house the full-resolution master copies, and will be behind the museum firewall, accessible only by NAL collection personnel. As a NAS server, the information on it can be retrieved and edited by authorized users anywhere in the museum without having to physically plug into the system. A second server is needed to interface with the public: it will both house copies of materials that we share with patrons (access copies for non-restricted collection materials), and act as a storage center for sending out digital copies of collection items. Because it is a separate server, any potential data breach on the public side would not affect the safety of our master copies.
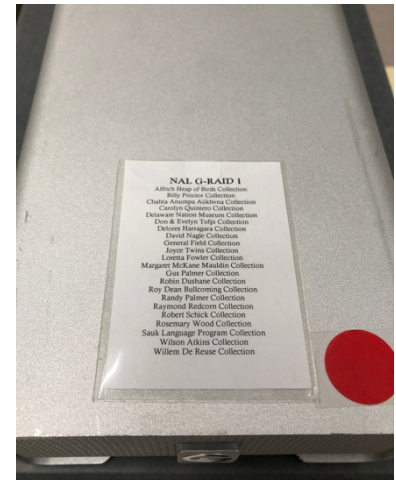
In terms of archival preservation, LTO is a widely used archival backup system, not only for language archives but also in other industries (e.g. film production and archiving). The latest in LTO tape technology is LTO-8, which has a native storage capacity of 12TB per tape (30TB compressed). The stated lifetime of each tape from the manufacturer is 30 years, which is unfathomably long in the world of modern technology. Since we will store the tapes in climate-controlled spaces and read/write to them infrequently (once a year), we expect tapes to last close to 30 years. NAL will write the collection to two sets of LTO tapes: the first copy will be kept in the Sam Noble Museum in NAL's safe for ease of access, while the second copy will be given to the University of Oklahoma police department (OU PD) for safe keeping off-site at their facility. We calculate that we will require 8 LTO tapes (4 tapes x 12TB uncompressed per tape = our maximum 48TB server capacity, x 2 copies) to store the full digital collection, factoring in future expansion. This LTO system will bring NAL into compliance with best practices for archival emergency backups and storage.

With this much-needed hardware, we are also in need of an updated computer system to run it and its associated software. Our current collection computer is an early 2008 Mac Pro, which is barely capable of running our data management program. The new collection computer will be the primary point of access for adding and organizing files on the server with the master copies, using LTO software to read and write LTO tapes, and running checksum software to prevent files from being corrupted in transit between the master copy server, the public-facing server, and our LTO tapes. The technical specifications and initial price quotes for these servers, a computer, checksum software, and the LTO system are provided as Supporting documents (1-4).

Beneficiaries and benefits

With the completion of this project, digital versions of NAL collection items will be housed in modern NAS servers and backed up to LTO tape, which helps bring NAL in line with current best practices in digital storage. The most important benefit of the project is that the collection will be secure, which it currently is not. Not only will the RAID systems prevent local data loss, but also having an off-site backup is crucial for ensuring that a copy of the collection exists that is not susceptible to local environmental disasters or hacking. Securing the collection is also an integral part of fulfilling our obligations to our patrons and donors, who have entrusted us with their materials and depend on us to maintain the collection for future generations.

This update to our infrastructure will also greatly improve the user experience for the ~100 people per year who visit the collection to access our materials. Patrons will be able to access our public materials directly, from any single location, without waiting on archivists to gather hard drives and travel to computers that can read them. The new system also has the benefit of separating the access copies from the master copies, which eliminates the danger to the collection that comes from regularly accessing and reading from the master copies.

The new system will also greatly improve our internal workflows. For example, we provide more than 2,000 digital copies of collection materials a year for our visitors. Historically, if people requested large amounts of data we would have to ship them a small hard drive. With the new system, we can create a password-protected folder on the public server and share it with them remotely, regardless of how many files they request. Additionally, having checksum software is important not only for accurately reading and writing LTO tapes, but also in protecting against data loss and corruption in all of the transfers between our hard drives, the master copy server, and the public-facing server. Checksum software also helps with monitoring and sustainability (see the NDSA's third and fourth levels of digital preservation, cited above), since it logs all transfer activities and helps us track the obsolescence of our storage and media.

Since we are a Native language archive built largely by the contributions of our local tribal partners, we are continually consulting the Native community in major projects and decisions about the collection. With respect to this project, discussions with our Native partners strongly influenced decisions about what infrastructure to acquire. First, Native depositors expressed discomfort with the idea of their data being stored in the cloud, were it would not be locally and exclusively controlled by the entity (NAL) to whom they are entrusting their materials. Regardless of US property law, many tribes consider their languages to be their intellectual property (see e.g., Hutton 2010), and the unethical sharing of Native language materials, sacred ceremonies and songs, etc. is a serious ongoing concern. This request to avoid cloud storage precipitated the decision to acquire a stand-alone server for NAL collection materials. We also learned through these discussions that people are uncomfortable with any restricted access items being on the same server as the materials being made available to

the public. To address this concern, we have decided to acquire two separate servers, so that all restricted items can be housed only on the server for our master copies, which is physically separate from the public-facing server.

Furthering SNM, NAL, and IMLS goals

The SNM currently curates over 12 million objects across 12 biological and social science collections. The museum is fully committed philosophically and financially to the preservation and care of these collections and has worked aggressively to upgrade the professional care and preservation of its collections. The museum's efforts were recognized in 2004, when we received the Award for Outstanding Commitment to the Preservation and Care of Collections from Heritage Preservation and the American Institute for Conservation of Historic and Artistic Work. The current strategic plan (see the Strategic Plan Summary) outlines objectives for 2019-2024 which include building and improving stewardship, ensuring collection integrity and accessibility of collections, and strengthening our connection with the community. This project directly addresses all of these objectives. The funds for this project will provide the infrastructure that will allow NAL to continue and improve our collection stewardship operation as it pertains to our valuable digital collection items, and ensures the integrity of the digital collection going forward. With respect to community, a central part of NAL's mission is to provide ethical, collaborative curation of Native cultural heritage materials for our Native communities. This requires maintaining secure, reliable preservation infrastructure. Additionally, the strategic plan specifically states that the museum seeks to "expand the role of the Museum as a leader in the preservation of Native American Languages". This emphasizes the strong commitment of the museum to the development of the NAL department, and the prioritization of our preservation and stewardship efforts.

As a unit of the University of Oklahoma, the collections improvement objectives of the museum support the larger university mission "to provide the best possible educational experience for our students through excellence in teaching, research and creative activity, and service to the state and society." NAL regularly provides opportunities for students to work in the collection via internships, independent studies, and SNM's volunteer program. We also plan to employ OU students as part of this project. Additionally, the project actualizes IMLS's strategic investment to support museums in collection stewardship in ways that increases access to information and facilitates the use of collections into the future.

In addition to addressing NAL's most pressing digital storage and access needs, the completion of this project also constitutes an important step in NAL's long-term goal to serve the collection to the public. Currently, people interested in accessing the archive must contact our archivists or appear in person to locate resources and get copies of the materials they need. This means that many people, particularly community elders and speakers of the languages we are committed to preserving, cannot easily interact with our materials because they are unable to make the trip to the museum. This limits the audience and utility of the collection. We hope in the next few years to write a grant to bring the collection online, which will allow researchers, community members, and the public to make use of the collection's Native American language materials in ways not previously possible. The current project lays the foundation for this important development in the future of the collection, in that the first step in this process is to have a central storage location for our files and a way to serve them to the internet. One reason we are requesting a public-facing server (in addition to the need mentioned above to physically separate restricted items from public items) is so that it can eventually host our collection website and serve our open access materials, inside and outside the museum.

Project Work Plan

NAL will undertake the following procedures as part of this project, which has an anticipated duration of one year (September 2020-August 2021):

Pre-award tasks:
1) Re-evaluate our data structures. We will take the opportunity to re-think our filing system to center around single-point access for all of our files. We will maintain our core organizational structure (Collection/Deposit → Item folder → Item files), but we will need to have a plan in place for post-migration that is ready to implement upon the arrival of the new system.

2) Formalize NAL's agreement with OU PD to keep backup LTO tape copies of the collection in a secure location, to be updated annually.

Grant-funded tasks:
1) Order the digital infrastructure (two servers, computer, LTO system, LTO tapes, checksum software) that will replace our current aging hard drives. Quotes for the specific equipment that we intend to purchase are provided in Supplementary documents 1-4, which were compiled by the museum's IT department. These orders will be placed as soon as possible at the start of the grant period.
2) Once it arrives, install the hardware and software in our secure, climate-controlled collection space. NAL staff will work with museum IT to make sure that everything is set up properly and complies with museum IT policies. IT staff will also connect the new equipment to museum systems and secure everything behind the museum's firewall. We have allotted a month to get these systems set up (ca. October 2020).
3) NAL staff migrate the collection off of the old hard drives and onto the new NAS system. Museum IT helps ensure that the system is properly RAIDed, and protocols are set up which identify when individual hard drives need to be replaced. We expect this to take about two months (November-December 2020).
4) NAL staff implement re-organize the file structure to reflect the new organization developed before the award (December 2020).
5) Hire and train two undergraduate students (January 2021). These students will assist with tasks 6-9 from the point of hire to the completion of major grant activities at the end of August. The student will be recruited from the Native American Studies department (the PI's academic unit) and will have interests in language, museums, and/or historic preservation. Previous experience interning or volunteering with NAL and therefore having experience with our systems and operations is preferable. Training will include an in-depth tutorial on NAL operations and procedures, with special attention to our digital ingestion protocols, file formats, and archival standards (https://samnoblemuseum.ou.edu/collections-and-research/native-american-languages/accepted-electronic-formats/), and also our access policies for restricted and unrestricted items (https://samnoblemuseum.ou.edu/collections-and-research/native-american-languages/native-american-languages-access-policy/).
6) Project staff work with museum IT to create the first LTO backup of the collection. Once the tapes are written, one copy will be stored in the department safe and another will be stored off-site with OU PD (January 2021).
7) Project staff will migrate access copies to the public-facing server. While many of the items in the collection already have both master (high-resolution) and access (lower-resolution) digital copies, there has never been a systematic effort to create both sets of formats for all collection items. As such, migrating access copies to the public server will require the recreation of ca. 2,000 new access copies for files which currently only have high-resolution versions. Once this step is complete, the collection as it exists on the new system will be ready for use with the public (February-April 2021).
8) Project staff will update location metadata for all items in the collection. We currently have two metadata fields ("digital file location" and "NAL collection call number") which reflect the physical locations of our files. Both of these fields will need to be updated for all items to reflect their new locations on the servers and LTO tapes (May-August 2021).
9) Project staff will update museum-level and collection-level procedural documents. The NAL operations manual will need to be updated to reflect the new infrastructure and the new organizational system. We will also revise museum-level curation documents, including the Emergency Operations Plan, the Salvage/Recovery Plan, and the Curatorial Procedures Manual (May-August 2021).
10) The PI will use the final month of the project to document and evaluate project outcomes and complete required reports.

This project is best characterized as "scaling" in terms of its maturity level, were we are implementing infrastructure improvements based on best practices in the field (see previous references) and the input of our tribal partners. We have also solicited advice from the curators of other language archives (e.g., the Archive for the Indigenous Languages of Latin America (AILLA), Kaipuleohone, the Pacific and Regional Archive for Digital

Sources in Endangered Cultures (PARADISEC)) about what types of digital systems they have for preservation and access for those collections. The current proposal reflects these different sources of input as well as our unique situation as a regional archive and a department within a university museum.

Budget and resources

The total cost of this project is $162,044. The majority of the funding from IMLS is for new digital infrastructure, which includes two servers each with 48TB of hard drive space ($28,250), an LTO system and 8 LTO tapes ($6,398), checksum software (ten-year subscription) ($490), and a computer to manage these systems ($639). All equipment quotes can be found in Supporting documents 1-4. The remainder of the IMLS funds are for two undergraduate collections assistants ($14,042), summer salary for the PI (a 9-month employee; $8,972), and associated indirect costs totaling $18,413. The salaries of NAL staff for the duration of this project as well as compensation for the time contributed by museum IT and associated indirect costs are provided as cost share (totaling $81,902). The museum is also contributing space for the new equipment, as well as assuming responsibility for providing technical support for NAL systems going forward.

Project personnel

This project has been developed by NAL staff in consultation with our Native stakeholders and curatorial staff from other language archives (described above). However, the core participants in charge of executing the project include NAL staff (Dr. Raina Heaton and Nicholas Wojcik), with additional support from our two hired student assistants (TBD), and museum IT. Dr. Raina Heaton is an Assistant Professor of Native American Studies and the curator for the NAL collection. Dr. Heaton is responsible for overseeing the execution of the project, hiring and oversight of the student worker, evaluation of the project's progress, dissemination of project results, and producing all documentation associated with this grant. Nicholas Wojcik is the Collections Manager for NAL, responsible for managing all daily operations that pertain to metadata, digitization, and the archival records system. With respect to this project, he is responsible for overseeing the migration of collections materials onto the new servers, training the student worker, and overseeing updates to NAL procedures, organizational structures, metadata, and file formats. The resumés of NAL core staff are provided in the resumé section of the application. Additionally, the student workers will be responsible for assisting with the creation of access copies, assisting with the migration of access copies to the new public server, updating the file location metadata for the collection, and helping make changes to NAL's operating documents. Finally, museum IT is responsible for overseeing all of the technology and computer systems for the various departments within the museum. For this project, museum IT will order the required equipment, install the equipment in the collection, and connect it to other museum systems. IT is also responsible for assisting with maintaining the new system beyond the duration of the grant.

Tracking progress, measuring performance, and risks

The progress of the project will be overseen by the PI (Dr. Heaton). Particular attention will be given to maintaining the integrity of the new systems, ensuring that format guidelines are being adhered to, and ensuring that no restricted items ever appear on the public server. Project staff will meet weekly to discuss progress, address any questions/concerns, and evaluate any tweaks that might be necessary to the workflow.

There are no risks associated with the execution of this project. The only potential risks to the collection are related to the eventual need to replace these systems as digital infrastructure inevitably evolves and becomes obsolete. First, SNM assumes responsibility for all systems in operation under its roof. The new NAL infrastructure will be integrated into museum systems, and NAL and museum IT will be responsible for their upkeep. Second, NAL has an annual maintenance and operations budget which is capable of funding all anticipated systems maintenance for the lifetime of the system (replacing faulty hard drives and needs of that nature). Additionally, while the collection will be responsible for replacing the servers eventually, the lifetimes of the infrastructure components we plan to acquire provides sufficient time for these maintenance expenses to be budgeted for before the point at which those system will need to be replaced.

Sharing results

The museum will disseminate and share information about the project with the public through press releases, social media (Facebook, twitter, blog, etc.), annual reports, the museum and department websites, on-site visitor kiosks, museum newsletters, tours, public events, and in our informal and formal presentations. The project will provide an opportunity to promote the support of IMLS and collection and research activities at the museum, including how the project increases the research value of museum collections and collections data. NAL will also notify our depositors and the Native community when the new system is functional for public use (expected April 2021) via our own networks as well as the Native community networks at OU (Native American Studies, the OU Tribal Liaison's Office).

Project Results

This project is critical to both the short-term and long-term preservation of an irreplaceable collection of language materials from 316 indigenous languages. Through this project we will drastically improve digital stewardship of the NAL collection with new hardware and software, which will in turn improve access to the collection for our visitors. The project also provides training for two students, who will be able to leverage the experience for their future career goals.

Tangible results
1) Rehousing of all digital materials (master and access), which increases our compliance with best practices in digital curation;
2) Creation of an archival-quality backup of the collection on long-term storage media, where none previously existed;
3) Improvement in the user experience for the archive, both for in-person visitors and those making copy requests;
4) Increased security for the collection at multiple levels: RAIDing of the new hard drives to prevent local data loss, off-site storage to prevent catastrophic data loss, removal of any public interaction with master copies, and separation of storage for restricted vs. public items in the collection;
5) Improvement to internal workflows, including mechanisms for identifying errors in transmission and storage (checksum);
6) Creation of access copies for all open-access collection items;
7) Establishment of the foundational infrastructure necessary for NAL to actualize our long-term goals.

Measurements of success
The success of the project is measured by the completion of the following key tasks over the course of the one-year grant period:
1) Successful equipment acquisition and installation;
2) Successful migration of our master copies;
3) Creation of long-term archival backups to LTO tape;
4) Creation of ~2,000 access copies, and successful migration of all access copies to the public server.

Sustaining project benefits
Funding this project will immediately address critical issues of security and preservation facing the collection by rectifying current substandard conditions for digital stewardship and access. The SNM is dedicated to the stewardship of its collections, and this project advances the mission and strategic goals of the museum as a regional resource center and steward of invaluable cultural and natural materials pertaining to Oklahoma. This project is also the first step in larger project to make the collection accessible online, which we hope to be able to make a reality within 5 years of the completion of this grant.

HUTTON, CHRISTOPHER. 2010. Who owns language? Mother tongues as intellectual property and the conceptualization of human linguistic diversity. *Language Sciences* 32(6):638-647.

Schedule of Completion

	Sept 2020	Oct 2020	Nov 2020	Dec 2020	Jan 2021	Feb 2021	Mar 2021	Apr 2021	May 2021	Jun 2021	July 2021	Aug 2021
Order hardware	■											
Install hardware and software in the collection		■										
Migrate master copies of all collection materials to the master copy server			■	■								
Hire and train students					■							
Make and distribute LTO backups					■							
Migrate/create access copies for public use						■	■	■				
Update file location metadata									■	■	■	■
Update procedural documents									■	■	■	■
Grant wrap-up												■