Indigenous Digital Archive: Building Effective Access to and Collaboration with Mass-Digitized Archival Documents
Abstract

To conduct "The Indigenous Digital Archive: Building Effective Access to and Collaboration with Mass-Digitized Archival Documents," a project to build tools to contribute to the national digital platform, the **State of New Mexico's Museum of Indian Arts and Culture (MIAC)** as lead is joined by the **New Mexico State Library Tribal Libraries Program** and the **Indian Pueblo Cultural Center** (IPCC), a nonprofit jointly run by the 19 Pueblo tribes of New Mexico.

Activities: From May 2016 to April 2019, to create an opensource online access and collaboration tool that will enable effective access to the content of mass digitized archival records, the IDA project will:

A) **Create a software tool** that will extend the capacities of Omeka-S, an opensource online content management system in use by GLAMs, through building or integrating a suite of architectural and interactive features that include:
   1. Translation of the images stored in Omeka-S to the International Image Interoperability Format (IIIF), to allow image display enhancements and use of the Open Annotation format, a digital sustainability-friendly format for annotations shared with the web annotation software Hypothe.is.
   2. Tools for user tagging and annotations that apply to a portion of an image, or across a range of images, rather than at the level of the entire digital object, and the ability for a user to make that content public, shared to a group, or private.
   3. Tools for users to improve navigation by marking the beginning and end of documents in a series, and, for material arranged chronologically, marking the beginnings and endings of years.
   4. Semantic tagging, with tools for specific local controlled vocabulary or taxonomy to use as a source for semantic tagging and to anchor topic pages to. User tagging is supported with autocomplete and synonym support, and support for multiple language.
   5. Peer-to-peer communication such as voting up or down on tags and annotations; links from documents to relevant discussion forums; and user-created collections that can be shared.
   6. The ability for a user to flag a passage for redaction, such as for reasons of cultural sensitivity.
   7. For enhanced user experience, working with use case community in development of a user-centered interface guided by the principle of creating a "generous interface," where a user is presented with useful overviews and leads of information and possibilities rather than just the traditional search box.
B) **Create a model use case** creating access to digitized copies of open public records related to Native land and the US Government Indian Boarding Schools of the late 1800s to 1930s, and support GLAM staff and community and other researchers in using the interface to create create effective access to previously unavailable public records of interest. Activities include: 1) Community building meetings to help incorporate diverse populations including New Mexico's 23 tribes, fostering a community who will help inform design of the user interface. 2) Acquiring and ingesting digital images of relevant mass digitized open public records. 3) Training tribal librarians in the IDA interface and use case content. 4) Conducting an IDA Fellows program, open to the 23 tribes of New Mexico plus Hopi, in which stipend support is provided for 3 Fellows to participate intensively over a year to help create model contributions.
C) **Disseminate information**: 1) about the IDA tools for Omeka-S to GLAM staff and developers, providing code on Github, and 2) about the IDA use case and expanded access to relevant collections to Native communities and Native-related GLAMS.

Needs: Regionally, and among Native communities more broadly, this project meets the needs of broadening access and expanding use of collections and content that has been unavailable to the communities to whose history and cultural heritage it relates, and who have been requesting access and training. Nationally, among GLAMs it meets the need of 1) providing a set of tools that extend a widely adopted Content Management System to broaden access to and expand the use of digitized content, especially mass digitized archival documents and photographs for which OCR does not provide an aid to access. 2) This project meets the need of a toolset extending the potentials of crowdsourcing beyond straight transcription or the creation of structured data from a limited range of materials, such as daily troop logs or logs of weather data. 3) The use case addresses the need for models of engagement with diverse communities to broaden access.

Outcomes and measures include 1) creation of toolset extending capabilities of Omeka-S. 2) Creation of online access to at least 140 microfilm reels (~140,000 pages) of open public records related to Native lands and to the US Government Boarding Schools of the late 1800s to 1930s, with 3) discoverability of the records via the Digital Public Library of America (DPLA). 4) At least 2% of documents will have annotations by people who have never seen them before. 5) Increasing digital literacy, access to, and use of collections in underserved communities, and encouraging intergenerational conversations and learning.

The Indigenous Digital Archive Project: Building Effective Access to and Collaboration with Mass-Digitized Archival Documents

1. Statement of Need

Since the beginning of social media and Web 2.0 those in charge of providing access to archives and special collections have talked about the need to create interfaces to bring in diverse voices and share the authority for describing material, particularly for detailed descriptions (Yakel 2007, 2011). However, even what was long a model experiment in archives for adding user-driven information to records, the Polar Bear Expedition Digital Collections (Yakel 2007), is no longer available, leaving almost no trace in the current online information, as without another major project it was unable to be integrated with tools serving needs in preservation and access (Bentley Historical Library n.d.). The need for an effective toolkit compatible with the forward migration of content management systems and digital preservation has not flagged, however. In Kate Theimer's 2011 volume collecting examples and plans of digital engagement and crowdsourcing in repositories, Yakel notes "an undiscovered power of the social web for cultural institutions" that can be engendered when "authority for the description and representation of [online materials] is shared."

In recent years, the impetus for allowing people to contribute to knowledge about collections online has given rise to projects like the Citizen Science projects of Zooniverse, Zooniverse's new AnnoTate transcription project, and the Smithsonian Transcription Center. These and similar recent projects have developed strong tools for repositories allowing people to conduct very specific tasks online, such as full-text transcription, correction of automated optical character recognition (OCR), and transcription of portions of regular records, such as weather logs, an institution's register of bird specimens, or logs of troop movements, into structured data. In each case by design there is a defined and specific task to accomplish, rather than being an interface to aid exploration of digitized material. In a recent analysis of existing tools facilitating crowdsourced tasks, Ben Brumfield and Mia Ridge (2015:17) noted the "line-at-a-time, queue-oriented, multi-track transcription workflow" characteristic of the Zooniverse interfaces don't allow for users to return to something they've worked on, or see what others have done and discuss among themselves.

These interfaces are designed to collect a limited range of structured data according to a particular research design, and are becoming increasingly sophisticated at directing the workflow. However, these are not tools for exploring collections beyond a specific kind of encounter. Sometimes the narrow task orientation and gamification of some interfaces might mean that a user is presented with a single image of text that is completely interesting to them, only to see it whisked away after they've completed a transcription task, with no clue of where it came from, or ability to see the whole item.

As a tool allowing more flexible interactions with content, FromThePage has made great advances in user interactions around transcriptions and translations, OCR correction, and exploring content based on indexed key terms. FromThePage also integrates and allows use of content stored in the Internet Archive or Omeka, an opensource content management system adopted by many GLAMs, and allows users to create articles on indexed key terms. Most recently, FromThePage also made the advance of incorporating the systems interoperability-focused International Image Interoperability Format (IIIF) standard. Being able to work with images in an IIIF compatible environment brings many advantages in user interface, such as the ability for smooth deep zooming, and being able to work with and compare material from multiple IIIF-compatible repositories in one browser interface.

However, tools are still needed to enable work with mass digitized documents. As Brumfield and Ridge note, FromThePage is designed for work with small collections, and its limited discover interface means it would be prohibitive to use on collections of hundreds or thousands of documents (28). Additionally, indexing in

FromThePage depends on a transcription in which the subjects are identified and hyper-linked from. For mass digitized archival documents, access needs are not always met by transcription. This is not only because full transcription or OCR correction is usually much more time consuming than selecting a tag (a name, event, concept, or place) that would be meaningful for someone looking for the content, but also because often times what would be used as a keyword does not actually appear in that text. (For example, a derivative, alternate, or misspelled form of a name is used, or what would receive a keyword tag of "boarding school deaths" appears in euphemistic language.)

The need to create online access to allow collections to reach a wider group of users means that repositories do continue to look to mass digitization as part of their strategies (Rocke and Klic 2015). Usability studies on the posting of a large backlog of tens of thousands of archival photographs with minimal processing found that users preferred to be able to access the material even with the most minimal metadata, and in some cases supplied information about the items that staff then used to augment the catalog record (Kelly 2014). At the Hoover Institution Archives at Stanford, which holds over 27 linear miles of mainly 20th century material, Miller (2013) has suggested mass digitizing unprocessed archival collections, with searchability to be provided through OCR, with the idea that this would give users a level of accessibility that would be familiar to them from Google searches.

The need for more user-directed tagging while exploring content, as opposed to straight transcription or the narrow task orientation of Zooniverse, have led to some projects finding a solution in outsourcing the content (and contributions) to commercial ventures such as Flickr. While experiments with putting collections on Flickr and similar services has shown that people are eager to interact with digitized historic cultural content online in a collaborative environment, and that using Web 2.0 tools can enhance a person's ability to find the digitized documents they're looking for, and provide their input into what is presented with the material (Zinkham and Springer 2010:107), using these commercial content services introduces its own problems, such as user privacy from commercial data mining now or in the future, the large expenditure of effort on creation of material that is 'locked' inside the company's system, the need to conduct additional activities for digital preservation purposes (web harvesting or web archiving of the site), and limitation of formats (e.g. single images). Additionally, there is not yet a match between such services and potentially culturally sensitive material, or where participating community members don't feel comfortable with the contributions they are making becoming part of a commercial venture, rather than being curated by a museum or archive.

The US National Archives itself has begun what it calls Citizen Archivist crowdsourcing projects. However, these are largely limited to transcription, subtitling, and the ability to add free-form tags to a small curated selection of images in a custom-built catalog. Tags can only be applied to the selected records, individual photographs which have had full item-level cataloging, greatly limiting the ability of the approach to increase access beyond a small taste of the collections.

With respect to user-directed exploration and tagging, the Mukurtu content management system, with an emphasis on responding to the needs of cultural sensitivity issues in a digital repository, has built an effective system designed for gating and providing access to material appropriate to one's tribal affiliation, clan, gender, age group, and other indigenous community-defined considerations. Users may create tags and add commentary, at the level of the entire digital object and catalog record. For now, this constraint limits the ability of Mukurtu to provide access with the same kind of interactivity to mass digitized material, or allow someone to show exactly where in a passage or multi-page document their comment applies.

Recently, in talking about the potentials for crowdsourcing in increasing access to collections at the May 2015 Crowd Consortium workshop as part of the IMLS-funded National Forum in Crowdsourcing for Libraries and Archives, HathiTrust's Jeremy York noted that while we hold collections for the benefit of our communities, "yet

many in the community do not know what is in our collections, [and] they can't find the useful materials; we are not being as effective as we could in fulfilling our vision" (Flanagan, et al., 2015:132). This challenge is still urgent to address.

This project will build tools based on international standards to help bridge that gap. This project will build software tools that interface with the Omeka-S digital content management system for cultural heritage institutions to create an online access and collaboration layer that enables creating effective access to mass-digitized archival documents, images often highly resistant to automated processes such as optical character recognition (OCR), through enabling a suite of interactive features based on the interoperabilty focused international standards of the International Image Interoperabilty Format (IIIF, http://iiif.io) and Open Annotation.

Use of the IIIF format enables a suite of abilities, such as the ability for a user to quote all or part of an image using just a URL, seamless deep zoom, the ability to add keyword tags (Appleby 2015) or annotations to a portion of or even a range of images, and even, as seen in IIIF-clients such as the Mirador viewer, being able to view and work with objects from multiple IIIF-enabled repositories in one browser interface (Snydman, et al., 2015). Use of the Open Annotation format, the standard used in the web annotation software Hypothes.is, and maintained by the W3C, the international organization responsible for managing the standards that make the internet able to communicate across different languages and applications, addresses the need to be able to create online collaborations including crowdsourcing applications where the data will stay linked to the source images, and be maintained in an internationally-agreed standard format that will allow it to be sustained and useful outside of any one particular software application, aiding long-term digital stewardship.

The IIIF standard has been adopted by US libraries and institutions such as ARTStor, the libraries of Stanford, Yale, Princeton, Harvard, and others, and DPLA encourages it as a way for contributing repositories to have better representation of their content. Internationally IIIF has been adopted by La Bibliothèque nationale de France; the National Libraries of Wales, Austria, Denmark, Israel, Poland, Serbia, Norway, Australia, and New Zealand; Oxford University's Digital Bodelian, and others, for a total of at least 50 participating organizations so far. Shims or patches have been made for existing repository systems such as ContentDM and the Internet Archive. 345 people subscribe to the email listserve of the IIIF developer and user community, IIIF-Discuss. The toolkit components of our project have been discussed as solutions to needs of common interest at in-person IIIF Community gatherings as well as during bi-monthly conference calls.

This project will also build on the Universal Viewer, an open source project to enable cultural heritage institutions to present their digital artifacts in an IIIF-compliant and highly customizable user interface, initially developed to provide an interface to the content of the many different subunits of the British Library, and the Wellcome Library's emerging Digital Library Cloud Services (DLCS), a service to provide IIIF-compliant image hosting with additional services for cultural heritage digital projects, including OCR indexing and searching, annotation storage, and easy to use APIs.

Tools this project will develop greatly complement new tools developed for Omeka-S. For example, the new LOD tools Omeka-S is implementing to aid standardization in the collection and object level catalog metadata is complemented by the IDA tool allowing semantic tagging by users within the documents. Additionally, others could take parts of the semantic tagging toolkit that best apply to their applications.

Further, as the ultimate goal of this project is to build an interface that is really focused on creating the best user experience possible, this project will build on the emerging work in developing what Whitelaw (2015) terms generous interfaces, user interfaces that rather the present the user with just an empty search box, instead organize and present preliminary faceted data to provide the most information to a person in the most

apprehendable way, with the most effective possibilities for next steps for effective interaction.

We are matching these national needs for tools among GLAMs with needs among our local communities. MIAC in 2014 conducted needs assessments among New Mexico's many tribal constituents that articulated the need and desire for online access to archival documents about their histories and for training in archives. The IDA has since conducted a pilot study with making available a small sample of records, and additional activities including a community building session that drew community members from miles away and from young adult to elder, and a survey about which of the available public records are the ones to which people would most like access.

The use case of the IDA project will address the absence of access to open public government records relating to the build up to and operations of the US government boarding and day schools for Indians in the period of the Indian Wars up to the reforms of the "Indian New Deal" in the 1930s, and records related to tribal land claims in the same period. The State Coordinator of Tribal Libraries, who often receives reference requests related to information the documents the project will make available, notes that now having even just the pilot documents of student names online (http://native-docs.org) fills a need no one has been able to respond to before in connecting people affected by these government policies generations onward with the information they're seeking.

Creating effective access locally is particularly important at this time as this is a window of opportunity where tribal people in New Mexico have the benefit of understanding the records with the input of those who are elders today who were young children at the time of the creation of the later records, and others who still have first-hand stories from their parents or grandparents in the 1920s-1930s and even earlier.

To meet the need of being able to effectively connect people with the digitized collections, the IDA project will follow a plan of meeting needs for fostering digital engagement articulated by Ridge (2013), and conduct community building sessions, involve community members in the User Interface design, and conduct a sustained program supporting community members' digital participation.

2. Impact
Given these national and digital inclusion needs, MIAC is requesting an $249,858 IMLS National Leadership Grant for Libraries under the National Digital Platform priority to support "The Indigenous Digital Archive: Building Effective Access to and Collaboration with Mass-Digitized Archival Documents." To create an opensource online access and collaboration toolkit to enable effective access to the content of mass digitized archival records, the IDA project will:

A) **Create a software tool** that will extend the capacities of Omeka-S, an opensource online content management system in use by GLAMs, through building or integrating a suite of architectural and interactive features that include:
1. Translation of the images stored in Omeka-S to the International Image Interoperability Format (IIIF), to allow image display enhancements and use of the Open Annotation format, a digital sustainability-friendly format for annotations shared with the web annotation software Hypothe.is.
2. Tools for user tagging and annotations that apply to a portion of an image, or across a range of images, rather than at the level of the entire digital object, and the ability for a user to make that content public, shared to a group, or private.
3. Tools for users to improve navigation by marking the beginning and end of documents in a series, and, for material arranged chronologically, marking the beginnings and endings of years.
4. Semantic tagging, with tools for specific local controlled vocabulary or taxonomy to use as a source for semantic tagging and to anchor topic pages to. User tagging is supported with autocomplete and synonym support, and support for multiple language.
5. Peer-to-peer communication such as voting up or down on tags and annotations; links from

documents to relevant discussion forums; and user-created collections that can be shared.
6. The ability for a user to flag a passage for redaction, such as for reasons of cultural sensitivity.
7. For enhanced user experience, working with use case community in development of a user-centered interface guided by the principle of creating a "generous interface," where a user is presented with useful overviews and leads of information and possibilities rather than just the traditional search box.

B) **Create a model use case** creating access to digitized copies of open public records related to Native land and the US Government Indian Boarding Schools of the late 1800s to 1930s, and support GLAM staff and community and other researchers in using the interface to create create effective access to previously unavailable public records of interest. Activities include: 1) Community building meetings to help incorporate diverse populations including New Mexico's 23 tribes, fostering a community who will help inform design of the user interface. 2) Acquiring and ingesting digital images of relevant mass digitized open public records. 3) With our partners the State Library Tribal Librarians Program and the Indian Pueblo Cultural Center (IPCC), a nonprofit run by all 19 Pueblo tribes, training tribal librarians in the IDA interface and use case content. 4) Conducting an IDA Fellows program, open to the 23 tribes of New Mexico plus Hopi, in which stipend support is provided for 3 IDA Fellows to participate intensively over a year to help create model contributions.

C) **Disseminate information**: 1) about the IDA tools for Omeka-S to GLAM staff and developers, providing code on Github, and 2) about the IDA use case and expanded access to relevant collections to Native communities and Native-related GLAMS.

Through the design and build of this toolkit and in our partnerships and programming this project will meet the goals articulated in the 2015 NPD meeting report, of building and enhancing interoperable tools and services, being inclusive and mobilizing a diverse range of end users, including underserved communities, through self-directed and collaborative digital engagement with collections, and initiating radical collaborations.

Benefits will be sustained through MIAC continuing the IDA model use case and related programming, with plans to later expand the digital holdings to include records of Government Boarding Schools in other states (especially relevant since students were by design distributed geographically to reduce the influence of home communities), and to expand the IDA Fellows program to more participants annually. Benefits will also be sustained by adoption of the toolkit, or in whole or in parts as needed, by members of the GLAM community nationally, especially for medium and large cultural institutions using Omeka-S.

In addition to gathering input from community building sessions in New Mexico, the IDA project will allow for input, consensus building, and buy-in through regular participation in the bimonthly teleconferences and other outreach events of the IIIF Community, which has a growing participant email list that currently numbers 345 people.

Throughout the project, we will track and measure our outcomes in a number of ways. We will track progress of the design and creation of the software toolkit, keeping aware of developments of use in ongoing complementary projects that can allow us to exceed the baseline toolkit to meet the full desired range of software requirements. (These are articulated in detail in the Summary of Work document in the Supporting Documents.) During the development, testing, and release cycles, we will ask for community feedback through in-person usability feedback and targeted surveys that address user experience and functionality. Code and documentation will be made available on github.com.

For the IDA use case, we will acquire and make available online digitized images of at least 140 reels of microfilmed records of seven series of open public records held by the US National Archives related to land rights and the federal boarding schools and day schools, 1890s-1930. We will conduct surveys at the Community Building sessions to gather demographic information and information about previous experience with primary

source records, research, and online collections. Feedback on usability of the user interface and on the project's support for their digital engagement with the IDA will allow us to make adjustments during the course of the project. In addition to supporting volunteers, we will select and support 3 people from among tribal members and support their work as IDA Fellows. We will train at least 15 participants in our tribal librarians training program, and IDA Fellows will make at least 3 presentations at tribal libraries. At least two college classes will use the IDA in the project period. Data from the IDA interface will allow us to track meeting the goal of at least 2% of records receiving tags or other augmentation by people who had never seen them before.

Finally, we will conduct program evaluation with the Fellows and with focus groups at tribal libraries and at least one college classroom. We will conduct evaluations using the Generic Learning Outcome (GLO) toolkits developed by the UK's Museum, Libraries, and Archives Council as part of their Improvement Framework for Museums, Libraries, and Archives, and the Archival Metrics Toolkits for Web interfaces and Classroom use (Duff, et al., 2010).

3. Project Design
The IDA project will create an opensource software layer that extends the capabilities of Omeka-S through enabling online digital engagement and collaboration to create effective access to mass digitized archival documents. Having identified open public records related to Native land claims and the government boarding schools of the late 19th to early 20th century as both a priority for and otherwise unavailable to Native communities in our region, we will acquire digital images of at least 140,000 pages (140 reels) of previously microfilmed records held by the US National Archives, ingest them into a hosted server instance of Omeka-S, and make them available through a new online environment characterized by rich interaction and collaboration tools and an enhanced user experience made possible by using generous interface design principles to replace the standard search box.

As part of the development of the full toolkit, enumerated in the Supporting Documents, creation of a semantic tagging tool for user tagging complements the LOD toolkits Omeka-S will soon release for catalog records. We will build for the interoperability capacities of the international standards of IIIF and Open Annotation, make the toolkit or components of it able to be shared with other applications and datasets based on the same standard, and showing themselves especially useful as standards enabling perpetuation of crowdsourced data based on or involving source images.

To inform the software design, and to populate our use case repository, MIAC in partnership with the State Library Tribal Libraries Program and the Indian Pueblo Cultural Center, a non-profit run by all 19 Pueblo tribes, will host community building sessions, include the IDA in their current collaborative training program for tribal librarians and other community members, and support and engage with volunteers and 3 Native IDA Fellows online. The project will also continue to seek input and feedback at the biweekly teleconference and other meetings of the IIIF community of practitioners from diverse GLAMs.

Preliminary work has included developing User Stories and software requirements documents (see Supporting Documents) and surveying what software components are available to be built on, and which are anticipated to be emerging over the next few months. We have also conducted a community building session and surveyed attendees about their priorities for records to be made available in the online system. We conducted a pilot project of acquiring digital images of one sample reel of records, conducting outreach, and getting community feedback. Input to and feedback about the project design has also been gathered from the IIIF Community. The IPCC, State Tribal Libraries, and MIAC are also now a year into their collaborative training program in archives and archival management for tribal librarians and other community members.

The design of this project is deeply rooted in the needs identified in Section 1. We will conduct the activities

described in Section 2. A complete description of the toolkit we will build is in the IDA Tasks document in the Supplemental Documents. Additional information is available in the included Digital Stewardship Supplementary Information Form.

The strongest test of our design for opensource software to create effective access to mass digitized archival documents through crowdsourcing will be the actual use. We will collect information about its usability at community building sessions, including with the IDA Fellows. Formal toolkit assessments such as those conducted by our performance evaluator will be complemented by input to and feedback on user interface design as it is being prototyped. We will share information about the project and results of the online collaboration through social media, at community sessions at tribal libraries, at a gathering of government boarding school descendants and interested others, and at the GLAM professional conferences of the Society of American Archivists (SAA), the Association of Tribal Archives, Libraries, and Museums (ATALM), and the IIIF Community.

4. Diversity Plan
At a 2014 World Café planning session MIAC's Native American constituents expressed wanting MIAC to provide online access to documents relating to their history, and for MIAC to provide experience with archives, to increase professional development opportunities and to inform development of their own community's archives. MIAC conducted a pilot study with images of mass digitized open public records and further community building activities and assessments to determine the requirements of the project and the archival records to select.

New Mexico is home to 23 tribes – Acoma, Cochiti, Hopi, Isleta, Jemez, Jicarilla, Laguna, Mescalero, Nambé, Diné/Navajo, Ohkay Owingeh, Picuris, Pojoaque, San Felipe, San Ildefonso, Sandia, Santa Ana, Santa Clara, Santo Domingo, Taos, Tesuque, Ute, Zia, and Zuni – all of whom were affected by the Government Indian Boarding Schools and land actions of the same time period, in ways that still have repercussions in communities today. Included in our constituency also is Hopi, geographically separate in Arizona, but culturally and genealogically related.

New Mexico is the fifth largest state in terms of area and ranks sixth in the least density of population. New Mexico ranked 49th in economic well being in the Annie E. Casey Foundation 2013 Kids Count profile, with overall rank of 50th in the nation when economic well being, education, health, and family and community statistics are combined. The current situation is that most people would have to drive long distances and be able to spend days away to be able to access primary source information pertinent to their community's recent past and trajectory.

IMLS support will allow us to create the tools we'll use in creating a supportive learning environment where Native American community members gain digital literacy, critical thinking skills, and greater knowledge of events shaping their communities while individually and collaboratively engaging with authentic public documents of community history, government actions, and civic life never before available in New Mexico. Online, and in structured, supported learning environments for collaborative reading, tagging, evaluation, and discussion of what's found, MIAC will improve opportunities for more Native communities throughout New Mexico and be able to have a wide impact throughout the state in encouraging civic participation, increased economic opportunity through digital literacy, and fostering life-long learning, inter-generational conversations and increased cultural heritage knowledge.

To address concerns that there could be culturally sensitive material in the open public documents that will be made accessible online, our software interface includes the option of flagging a passage for redaction. Confirmation of redaction will be made according to policy developed by our advisory panel, and based largely on the redaction policy piloted by the Indian Pueblo Cultural Center (IPCC), the non-profit cultural center and library run jointly by all 19 pueblos.

This program will establish culturally relevant digital literacy opportunities, increase training for tribal librarians, invite the participation of volunteers, and guide the inaugural cohort of IDA Fellows participants through a year's program that will include them sharing their knowledge and skills in their communities at their tribal library. As tribal libraries play a key role in establishing digital inclusion in many Native American communities (Jorgensen, et al., 2014), helping strengthening local connections to tribal libraries makes all online resources more accessible, in addition to strengthening opportunities to participate in digital literacy learning with the IDA online resource.

In conducting this project, MIAC will be building on an ongoing partnership with the State Tribal Libraries and the IPCC for conducting training in archives and archival management for tribal librarians and others in the community. We will also conduct outreach to Society of American Archivist Roundtables, the Association for Tribal Libraries, Archives, and Museums, boarding school descendant groups, and topical listservs.

5. Project Resources: Personnel, Time, Budget
Key Project Staff: *(Please see resumes for additional information on qualifications)*
Project Director **Dr. Anna Naruta-Moya** will draw on a decade of project planning and management experience in multi-institution collaborations to oversee this project in vision and implementation. She will serve as project archivist and engage in the online collaboration. (25% time) As Communications Coordinator, **Ms. Caren Gala** (Nambé) will be responsible for communications with the media, responding to inquiries, disseminating information about the Fellows Program, receiving applications, organizing, and providing them for the selection committee formed by the advisory board, coordinating with MIAC and performing other hosting logistics for convenings of the Advisors and Fellows, and serving as moderator in the online discussion forums at least until one or more users fill that role. Ms. Gala will also contribute to program evaluations. (6% time)

State Tribal Libraries Coordinator **Ms. Alana McGrattan** will participate in convenings of the Advisors and the Fellows, webinar training of tribal librarians, and Fellows presentations at their tribal libraries. **Dr. Rose Diaz**, Library and Archives Director of the IPCC, will take part in convenings of the Advisors and the Fellows, and will organize the in-person training of tribal librarians held at IPCC. IPCC Program Coordinator **Deborah Jojola** (Jemez) will organize the two talks and one workshop for interested persons and potential volunteers at the IPCC, and conduct publicity. **Della Warrior** (Otoe-Missouria), Director of the state's Museum of Indian Arts and Culture (MIAC) will participate in all major activities. Significant personnel resources come also in our **Advisory Board** and **Technical Advisory Board**, as detailed in the Supporting Documents.

Consultants: **Mr. Tom Crane, Mr. Adam Christie, and Mr. Ed Silverton (Digirati), IIIF Specialist Technology Consultants,** individually and collectively in Digirati are leading developers for cultural institutions including the British Library and the Wellcome Library. They bring to the project expertise in development with IIIF compatibility, and while a UK firm are an essential part of this project as currently all US IIIF implementations are done with in-house development staff, and consequently there is no US contractor or firm to hire. With Mr. Crane as lead, they will design, build, and deploy the software tools. **Ms. George Oates, Owner and Director of Good, Form, and Function,** will join the Digirati team to consult on User Experience and User Interface design. The designer of Flickr, Ms. Oates is a world leader in developing generous interface design, including for the British Museum, the Wellcome Library, and has consulted for numerous cultural institutions including the Smithsonian and Historypin. They will work in person with staff, the advisory panel, and community members at a kick off meeting during the first advisors' meeting and a community building session. (180 hours consulting time, plus Mr. Crane's pro bono service on the technical advisory panel)

Contract Program Evaluator, **Mr. George Toya** (Jemez), will conduct in-progress program evaluation with the Fellows and with focus groups at three tribal libraries and closing program evaluation with the Fellows, focus

groups, and two classrooms. Mr. Toya will conduct evaluations in conjunction with Communications Coordinator Caren Gala (Nambé) using the Generic Learning Outcome (GLO) toolkits developed by the UK's Museum, Libraries, and Archives Council as part of their Improvement Framework for Museums, Libraries, and Archives, and the Archival Metrics Toolkits for Web interfaces and Classroom use (Duff, et al., 2010). Mr. Toya is familiar with the communities and cultural sensitivities and is experienced with complex organizational dynamics from his years as a board member of SWAIA's Santa Fe Indian Market. As a trained engineer, Mr. Toya is accustomed to research methods and recording, and he has been involved in developing and implementing traditional agricultural programs for tribal youth, and conducting archival research and oral history interviews.

Service Providers: Hosting of Omeka-S is planned to be provided by Reclaim Hosting, long-time providers of Omeka installs to hosted service, and which already offers the service of loading Omeka-S installations. For a digital preservation server, we plan to use D-Space with the option to store on Chronopolis, a TRAC-certified digital repository, which Omeka-S will provide integration with. The Wellcome Library's Digital Library Cloud Services (DLCS) is planned for the IIIF layer, for reasons including their IIIF image hosting, OCR services, annotation server, and searching across annotations and OCR.

Project finances will be managed by Monica Vigil, the Museum's Financial Specialist, who has been with the department for 22 years with consistently positive performances on audits. She will be responsible for preparing purchase documents and tracking project expenses. Both Dr. Naruta-Moya and Ms. Vigil will be responsible for coordinating expenditures and reporting with the Museum of New Mexico Foundation, who will serve as the fiscal agent for this grant. Dr. Naruta-Moya has prior experience conducting grant-funded projects.

Budget: We have included $4000 annually for hosting services and $2700 for TRAC-certified digital preservation server services, to be funded by MIAC as part of operating budget. $198,500 goes to software design and implementation consultants. All staff time except the Communications Coordinator is in-kind contribution of the applicant. $500 is allocated for promotional print material. $2091 is allocated to digitize 39 reels of lands-related microfilm held by IPCC, with IPCC making in-kind contribution of digitization of 58 reels ($3109). MIAC plans to raise $17,500 to pay the US National Archives duplication fee ($150 each) to acquire digital copies of 140 reels of microfilmed records, through grants and community contributions. IDA Fellows stipends total $6000, and $29,968 in travel costs includes in person convenings of the advisory panel, IDA Fellows, and technical advisory panel, and is inclusive of $7660 to support outreach travel and presentation at conferences.

Work Plan: Year 1: Advisory Panels meeting, Community building sessions and social media work, digital records acquisition and bulk ingest, user interface studies, initial design and software build; participation in Boarding School Symposium at Carlisle. Year 2: Continuance of software build, recruitment of IDA Fellows from tribal communities; training, information and experience sharing convenings including with Advisory Panels; sharing through social media and project blog posts; outreach to additional volunteer users, IIIF conference. Year 3: conclusion of IDA fellows program and fellows research presentations; synthesis of usability data and interface redesigns or refinements; test use in selected college classes; further outreach including presentations at conferences of Society of American Archivists, Association of Tribal Libraries, Archives, and Museums (ATALM), and IIIF.

6. Communications Plan
We look to reach our regional tribes and the wider Native American community, those interested in Native American history and culture, people with government boarding school heritage, and GLAM professionals interested in crowdsourcing and online collaboration applications and in potential for increasing access to mass digitized material.

Our communications plan includes outreach and engagement through the project social media sites including Facebook and Twitter, outlets popular among potential users of the IDA use case and IDA software toolkit. We will conduct community building sessions throughout the project performance period, and sustained engagement through the IDA use case online interface. The project director, communications coordinator, and community outreach specialist will jointly contribute to this work. We will continue to use surveys at community building sessions, and the contract program evaluator will conduct the detailed evaluations with Fellows and other tribal community members as outlined in Section 5. Throughout the project, we will participate in the discussion listserv and biweekly and other meetings of the IIIF community, comprised of more than 300 practitioners from diverse GLAMs.

Stipend support is provided for 3 IDA Fellows to participate intensively over a year to help create model contributions, as recommended from museum participation research that finds that high quality model contributions to open participation opportunities can encourage people to take contributing seriously, and "can inspire and energize less-skilled visitors without making them feel inferior" (Simon 2010: 215).

In-person sharing of knowledge and training between the Fellows and their larger community will also be held at the State Tribal Libraries located at each reservation, at MIAC, and at the Indian Pueblo Cultural Center, strengthening community connection to these resources. The project director will also make presentations at the annual meeting of the SAA, ATALM, the IIIF Community, and at the 2015 Symposium at the former Carlisle Boarding School. Software code and documentation will be made available on Github.com.

7. Sustainability
The benefits of this project will continue beyond the grant period of performance. MIAC will continue to maintain the IDA model use case, continue preserving content in a digital preservation server, and plans to expand the holdings beyond the initial set of records. MIAC also plans to support continuing volunteers and expand the IDA Fellows program beyond the inaugural cohort. The code created in this project will remain available on github.com as opensource software, and as it builds on the popular opensource Omeka platform and meets many articulated needs to enable organizations to move into online collaborative projects like crowdsourcing, it is likely to be adopted, kept viable, and improved through use.

Summary
The IDA project provides an opportunity to model how information can be made effectively accessible online without the bottleneck of museum staff individually cataloging each document, and while meaningfully engaging a core but underserved museum constituency. An opensource toolkit will be created and shared to provide a means for collaborative and individual making sense of and interlinking the content of primary source documents. [conntect to] This project will extend the capacities of the Omeka-S online content management system to contribute to the national digital platform through creating and effective method for online collaboration and creating effective access to mass digitized archival records.

This project meets the needs of broadening access and expanding use of collections and content that has been unavailable to the communities to whose history and cultural heritage it relates, and who have been requesting access and training.

REFERENCE CITED:

Appleby, Mike. "IIIF Tagging and Discovery. Sharing Images of Global Culture." National Gallery of Art, 2015 May 5 http://bit.ly/1Vtl2AV

Bentley Historical Library, University of Michigan. "About the Polar Bear Expedition Digital Collections". n.d. http://quod.lib.umich.edu/p/polaread/about.html

Brumfield, Ben and Mia Ridge. "Wellcome Library Transcribing Recipes Final Report." London, UK: Wellcome Library. 2015 September. http://www.slideshare.net/Wellcome/wellcome-library-transcribing-recipes-report

Duff, Wendy M. et al. "The Development, Testing, and Evaluation of the Archival Metrics Toolkits." *American Archivist* 73.2 (2010): 569–599

Flanagan, Mary, Neil Fraistat, and Andrea Wiggins. "Enaging the Public: Best Practices for Crowdsourcing Across the Disciplines." A report of Dartmouth's 2014 IMLS-funded National Forum in Crowdsourcing for Libraries and Archives (CCLA). http://crowdconsortium.org 2015.

Jorgensen, Miriam, Traci Morris, and Susan Feller. *Digital Inclusion in Native Communities: The Role of Tribal Libraries.* Oklahoma City, OK: Association of Tribal Archives, Libraries, and Museums. 2014.

Kelly, Elizabeth. "Processing Through Digitization: University Photographs at Loyola University New Orleans." *Archival Practice* 2014 Vol. 1, No. 1. http://libjournal.uncg.edu/ap/article/view/865/563

Krause, Magia and Elizabeth Yakel. "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid." *The American Archivist* Sep 2007, Vol. 70, No. 2 (Fall/Winter) pp. 282-314

Miller, Larisa K, "All Text Considered: A Perspective on Mass Digitizing and Archival Processing," *American Archivist* 76, no. 2 (2013): 536-537.

Ridge, Mia. "Digital Participation, Engagement and Crowdsourcing in Museums." London Museums Group. August 15, 2013. http://www.londonmuseumsgroup.org/2013/08/15/digital-participation-engagement-and-crowdsourcing-in-museums/

Rocke, Michael, and Lukas Klic. "Revealing Renaissance Art: Mass Digitization of the Berenson Library Photo Archive." Berenson Library, Harvard University. 2015. http://library.harvard.edu/revealing-renaissance-art-mass-digitization-berenson-library-photo-archive

Simon, Nina. *The Participatory Museum*. Santa Cruz, Calif.: Museum 2.0, 2010.

Snydman, Stuart, Rob Sanderson and Tom Cramer. "The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images." *Archiving 2015*, Proceedings of the Society for Imaging Science and Technology Conference, Los Angeles, CA, May 19-22, 2015. http://purl.stanford.edu/df650pk4327.

Whitelaw, Michael. Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly* 2015 9.1

Yakel, Elizabth, "Balancing archival authority with encouraging authentic voices to engage with records" in Theimer, Kate, Ed. *A Different Kind of Web: New Connections between Archives and Our Users.* Chicago: Society of American Archivists, 2011.

Websites References

Annie E. Casey Foundation 2013 Kids Count profile
http://datacenter.kidscount.org/~/media/61/2013KC_state_profile_NM.pdf

Archival Metrics Toolkits (http://ArchivalMetrics.org)

Inspiring Learning: An Improvement Framework for Museums, Libraries, and Archives. Generic Learning Outcomes (GLOs). http://www.inspiringlearningforall.gov.uk

Digital Library Cloud Services: Technical Overview
http://dlcs.github.io/dlcs-doc/technical-overview/

Indigenous Digital Archive
User Stories and Interface Features

User Stories

A) Sohk'wha, an adult tribal member: Sohk'wha's grandparents attended government boarding schools in the 1900s-1920s, but it was an experience they largely didn't discuss. Sohk'wha would like to use the documents to learn more about his grandparents' lives, and the environment they were in. He would potentially contribute content from his own documents, photos, and other sources he has found, especially to enrich information about his grandparents and their peers. Sohk'wha has friends whose grandparents or parents were in the boarding school system, and collectively they would like to know more about potential shared experiences of their ancestors. For example, were they in the system at the same time? Did they share experiences in workshops, the military-style bands, athletics? When they were on months-long work assignments in the "Outing" program, did they have worksites or bosses or even just geographic areas in common? Sohk'wha is college educated but may not previously have participated in extended primary source research with large collections or with material organized archivally. He knows information from family oral histories and has gathered what he can of family documents and photos.

B) Kwanie, a graduate student and tribal member: Kwanie wants to use the Indigenous Digital Archive to conduct research for her thesis. She hopes to track the development and changes in the US Government policy towards Native people from the Indian Wars and administrations under the Department of War through the creation of boarding schools and, forty years later, the reforms of the boarding schools and the Bureau of Indian Affairs under the Indian New Deal in the 1930s. She is interested in the collaborative potentials of the research, and will be putting in a significant amount of time reading and tagging the documents for her own research. She finds the ability to point to passages with permanent URLs and return to a workscreen with her selected documents useful in communicating with her faculty advisors and in making presentations and engaging in scholarly exchange.

C) Maria, an art historian, curator, and professor: Maria is an expert on contemporary and 19th and 20th century Native American Art. She would like to trace further the development of particular techniques and schools of practice, and finds that to do this it is key to be able to closely study art educators and programs of the Boarding Schools, from the period when Native art first started to be tolerated, through the time when it became an established part of vocational curriculum. She looks forward to being able to discover and to research in the documents detail of the programs' emergence, Native instructors' developing schools of practice, choices of traditional and modern techniques, and students. She feels that having detailed historical knowledge of the contexts of the boarding schools and the governance of the Bureau of Indian Affairs before the enactment of Indian Self-Determination is key to understanding any of the 19th and 20th century museum collections, biographical information, and community histories.

D) Poe, a young person in a tribal community: Poe is in high school and found out about the Indigenous Digital Archive through a program at her local tribal library. She has been experimenting with the documents, reading about government schools in her area. When she

reads something about a happening in the school or a policy that was directed towards her tribe, she asks her grandparents if they have heard about this and if it was true. This has become occasion for her grandparents sharing with her many stories about life in their community in decades past, and what their own parents and grandparents overcame. Poe's grandparents have asked her to show them certain things she has been finding, and have asked her if she could find more information about certain topics, such as the farming program in a certain time period and how various decisions were reached.

E) John and Harriet, volunteers / hobbyists: John and Harriet are interested in working with the material in the IDA because they are interested in knowledge and cultural institutions like libraries and museums, are interested in history, and would like to make meaningful and visible contributions to well-run collaborative research projects with clear objectives. John is retired and able to devote significant time to projects that interest him, and Harriet wishes to use part of her spare time to explore interesting material and make contributions to help a larger project make information meaningfully accessible. John and Harriet are interested in the content, and are also willing to work on structural tasks that make the materials more accessible overall, such as marking the beginnings and endings of individual documents in a large series of pages, and marking where years and months begin and end in a series of documents that is arranged chronologically. John has also just introduced the project to a friend who frequently makes presentations on his work at the annual state historical society meetings and is very interested in continuing his research with documents not previously available to him.

F) Whaa pin, a community elder: Whaa pin is a tribal member who himself attended boarding school. He wants to be able to see the records of the school he attended, and the decisions that were made that affected his family and home community. He wants to see what they put on record, and what else can be learned. For him, being able to examine these documents is part of his reconciliation with the times and policies he lived through. He needs a way to look through this information in blocks of time that are right for him, from the convenience of his home or local tribal library. He would perhaps like to connect with others who attended boarding schools in that period. Whaa pin's cousin, a tribal historian, is also interested in being able to access more documents regarding tribal lands and their disposition in the mid-to-late nineteenth and early twentieth century.

G) Arthur, Sam, and Feliciana, a Tribal Librarian, a University Librarian, and a State regional history librarian: While Arthur, Sam, and Feliciana are able to direct inquiring researchers to a selection of secondary sources, prior to the IDA they had no resource to which they could direct individuals trying to find out more in detail about most boarding schools, or about mid-to-late nineteenth and early twentieth century regional history. They are relieved to have a place to direct researchers to that provides primary source material in a setting that doesn't require someone to be already expert at researching with archival material and the special ways in which it is organized. They also find it helpful that directing researchers to highly relevant material no longer requires that the researcher be able to take off work and travel hundreds or thousands of miles.

H) Brook, a College Instructor: Over the course of a semester, Brook typically involves his students in an extended research project. The class takes one main topic, and students work in

collaborative groups to conduct original research to address aspects of the topic, write reports, and make presentations. While he conducts student evaluations in part by the assessments members of a group make of each others' work, Brook would be interested to be able to see how many and what tags over what areas of the documents were contributed by particular students.

Schedule

Schedule of Completion	YEAR One – Software Build and Outreach			
(May 2016 – April 2017)	May-July	Aug-Oct	Nov-Jan	Feb-April
Confirm workplans, timelines, and lines of responsibility	■			
Logistics for Fall travel: advisory board and consultants	■			
Begin social media presence		■		
Advisory Board Convenes		■		
Community Building Sessions at MIAC and IPCC, at MIAC with Advisory and Tech Advisory Panels and Software Developers		■		
Final selection of records for project	■	■		
Digital images of microfilmed documents ordered	■	■		
CMS selected	■			
Digital images ingested into CMS, available for experimenting for Generous User Interface development		■		
Backups to Trustworthy Digital Repository service begin		■		
IDA software tool prototype for annotation interface delivered and installed in web server environment			■	■
Community Buildings Sessions at MIAC and IPCC				■
IDA software tool refinements as needed				■
Community Building Sessions – Outreach for Fellows program and Volunteers: talk on IDA at IPCC; expert blog post 1				■

Schedule

Schedule of Completion	YEAR Two – Community Building, Collaboration			
(May 2017 – April 2018)	May-July	Aug-Oct	Nov-Jan	Feb-April
Sourcecode for software tool made freely available on github.com	■			
Outreach: workshops at IPCC, MIAC for potential volunteers and possible applicants to Fellows program		■		
Outreach: Training for Tribal Librarians (webinar), publicize Fellow opportunities		■		
IPCC trainings for tribal librarians		■		■
Fellow applications accepted (Dec-Jan)			■	■
Community Building Session at IPCC			■	
Advisory Board approves selection of Fellows, first Fellows convening, keynote from advisory board member, in-person Fellows training				■
Fellows interact with and via IDA, project staff support and join				■
Outreach: Expert blog posts 2 and 3				■

Schedule

Schedule of Completion	YEAR Three – Collaboration, Final Presentations			
(May 2018 – April 2019)	May-July	Aug-Oct	Nov-Jan	Feb-April
Fellows convening at MIAC: midpoint presentations and evaluation of experience, introduction and training on online presentation options, Advisory board convenes jointly	■			
Post-convening phone conference with advisors	■			
Continued: Fellows interact with and via IDA, project staff support and join				
Outreach: workshop at MIAC for potential volunteers		■		
IPCC trainings for tribal librarians		■		
Outreach: Expert blog post 4		■		
Outreach: Fellows present on IDA at their Tribal Library, feedback gathered		■	■	
Some IDA Fellows present on IDA at UNM and at NNMC, at least two interested high school or college classes identified through outreach use IDA, feedback gathered		■	■	
Fellows convening: closing program at MIAC, Fellows presentations		■		
Outreach and recognition of Fellows: Fellows online presentations shared during Indian Market		■		
Program Evaluation interviews and surveys of Fellows, selected community members, volunteers, and tribal librarians			■	■
Final evaluation and debrief with advisors				■
Presentation at conferences (ATALM, SAA) and at IPCC (August, September)		■		
Any final updates to sourcecode. Software tool made freely available on github.com				■

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded research, data, software, and other digital products. The assets you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products is not always straightforward. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and best practices that could become quickly outdated. Instead, we ask that you answer a series of questions that address specific aspects of creating and managing digital assets.

Your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

If you propose to create any type of digital product as part of your project, complete this form. We define digital products very broadly. If you are developing anything through the use of information technology (e.g., digital collections, web resources, metadata, software, or data), you should complete this form.

Please indicate which of the following digital products you will create or collect during your project

(Check all that apply):

	Every proposal creating a digital product should complete …	Part I
	If your project will create or collect …	**Then you should complete …**
☑	Digital content	Part II
☑	Software (systems, tools, apps, etc.)	Part III
☑	Dataset	Part IV

The IDA project has completed all sections.

PART I.

A. Intellectual Property Rights and Permissions

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1   What will be the intellectual property status of the content, software, or datasets you intend to create? Who will hold the copyright? Will you assign a Creative Commons license (http://us.creativecommons.org) to the content? If so, which license will it be? If it is software, what open source license will you use (e.g., BSD, GNU, MIT)? Explain and justify your licensing selections.

Content of the IDA will be assigned a Creative Commons license of Attribution-NonCommercial CC BY-NC. This is to encourage people to contribute tags and other information to a library/museum/archive project without worrying that their contributions will be used for commercial purposes. The opensource software tool created for the project will be assigned the GNU AGPL v3.0, a Free/Open Source license. This is to encourage further development of the tool to be returned to opensoure use.

A.2 What ownership rights will your organization assert over the new digital content, software, or datasets and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users about relevant terms or conditions.

The digital content will be assigned a Creative Commons license of Attribution-NonCommercial CC BY-NC. Access will be open and freely available on the Web, with notice that use is restricted to noncommercial uses. We will restrict use to noncommercial uses so that people can participate and not fear that their stories will be sold. We will notify potential users of relevant terms and conditions via the Creative Commons notices on display pages and as part of the interface for signing up for an account.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

The content we will post is all from open federal records, copies of which can currently be purchased by anyone possessing the required funds. However, we want to respond to concerns that there could be culturally sensitive material in the documents, and so we are building into the tool the means for a person to nominate a passage for redaction, flagging it for attention by an administrator, who would confirm the redaction. Our Advisory Panel will determine policy for redaction, with the aid of some local working policies by area repositories with sensitive materials.

Part II: Projects Creating or Collecting Digital Content

A. Creating New Digital Content

A.1 Describe the digital content you will create and/or collect, the quantities of each type, and format you will use.

Using matching funds, we will order digital images of 140 reels of microfilmed records from the US National Archives and digitize another 97 already purchased reels. The records relate to land rights and the government Indian Boarding and day schools in New Mexico, 1890s-1930. There are about 1000 images on each microfilm reel, scanned into the corresponding number of jpgs. A sample from one digitized reel are viewable at http://native-docs.org. We will convert the files to JPEG 2000. From the 140 newly acquired reels, the project will produce approximately 420 GB of files, the equivalent of approximately 78 linear feet of paper records.

A.2   List the equipment, software, and supplies that you will use to create the content or the name of the service provider who will perform the work.

The US National Archives will scan the microfilm and deliver the scans to us on CD or DVD, per their standard procedure for duplication orders.

A.3   List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to create, along with the relevant information on the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

Digital copies of microfilm reels from the US National Archives will arrive as JPG scans of 300 ppi. We will batch convert these to store as JPEG 2000 (jp2) images.

B. **Digital Workflow and Asset Maintenance/Preservation**

B.1   Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

Professional scans of microfilmed records will be purchased from the National Archives. After conversion to JPEG 2000, checksums will be generated as a baseline for monitoring file fixity. The scans will be visually spot-checked upon ingest into the repository system.

B.2   Describe your plan for preserving and maintaining digital assets during and after the award period of performance (e.g., storage systems, shared repositories, technical documentation, migration planning, commitment of organizational funding for these purposes). Please note: You may charge the Federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the Federal award. (See 2 CFR 200.461).

We will use the DSpace interface with storage option of an OAIS Compliant TRAC Certified Trusted Digital Repository, Chronopolis (http://chronopolis.sdsc.edu), a project originally funded by the Library of Congress. Chronopolis maintains content on professionally-staffed servers located in geographically separated areas: the San Diego Supercomputer Center and the UCSD Libraries, the National Center for Atmospheric Research, Boulder, Colorado; and the University of Maryland Institute for Advanced Computer Studies). Chronopolis provides continuous checksum monitoring of files to monitor file fixity and restore in case of bitrot or other corruption. We will use this digital preservation service during and after the project, with MIAC committing to the costs.

Similarly, MIAC commits to hosting the online resource beyond the completion of the project.

C. **Metadata**

C.1   Describe how you will produce metadata (e.g., technical, descriptive, administrative, or preservation). Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, or PREMIS) and metadata content (e.g., thesauri).

Descriptive metadata will be transferred from the original archival descriptions of the microfilmed records creating DACS-compliant descriptive metadata. As Omeka-S will be the content management system, the metadata will be entered in Dublin Core format. As needed, we will use the convention modeled by the Bentley Library to represent hierarchical arrangement in Dublin Core format, developed for use with DSpace (http://archival-integration.blogspot.com/2015/09/what-we-talk-about-when-we-talk-about.html).

User-added metadata will be allowed to be entered free-form, but autocomplete will suggest the remaining text of a tag based on previous usage or availability of a semantic tag, and administrators will be able to edit or consolidate tags to form an evolving data dictionary.

C.2   Explain your strategy for preserving and maintaining metadata created and/or collected during and after the award period of performance.

Digital objects and metadata will be backed up on the TRAC Certified Trusted Digital Repository when created and ingested. The live server containing user-supplied metadata will be daily backup up and periodically backed up on the TRAC Certified Trusted Digital Repository.

C.3   Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of digital content created during your project (e.g., an API (Application Programming Interface), contributions to the Digital Public Library of America (DPLA) or other digital platform, or other support to allow batch queries and retrieval of metadata).

The Digital Public Library of America (DPLA) is also working towards developing International Image Interoperability Format (IIIF) compatibility with its interface and encouraging use among contributors, and anticipates harvesting discovery metadata from the IDA via IIIF endpoints.

D. **Access and Use**

D.1   Describe how you will make the digital content available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

The digital content will be openly available online, with the exception of any user-contributed annotations that the user has designed as being private.

Underlying software platforms include Omeka-S on leased server service, leased services on the Digital Library Cloud Services (DLCS), and leased preservation server services on Chronopolis via DSpace.

Accessibility will be via standard web browsers with no special software tools required in order to use the content.

D.2   Provide the name and URL(s) (Uniform Resource Locator) for any examples of previous digital collections or content your organization has created.

MIAC does not yet have any online digital collections.

The sample records used in formulating the IDA project are online at http://native-docs.org .

Examples of online digital access projects of other organizations executed by our technical contractors:

Digirati:
http://wellcomelibrary.org/ for the Wellcome Trust (UK)
Universal Viewer for access of all of the British Library's online collections (under development)

Good, Form & Spectacle:
http://whatsinthelibrary.com/ for the Wellcome Trust
http://wb.britishmuseum.org/ for the British Museum

Part III. Projects Creating Software (systems, tools, apps, etc.)

A. **General Information**

A.1   Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) this software will serve.

We will create an opensource software tool that will extend the abilities of Omeka-S and use the potentials online collaborative and self-directed work to create effective access to mass digitized archival documents, particularly late 19th and early to mid 20th century documents that are resistant to Optical Character Recognition (OCR). Drawing on the material and catalog metadata we ingest into our Omeka-S content management system, our interface layer will display the contents in any web browser and allow users to sign in

to add tags at the page level. Permalinks will allow sharing or returning to a particular page. Users will also be able to flag for redactions portions of pages, with redactions confirmed or removed by an Administrator according to policy set by our advisory panel. Users will be able to discuss findings and suggest additional information and resources in a discussion forum. The Administrator will be able to moderate, edit and manage the users and the content of the system. User-added metadata will be allowed to be entered free-form, but autocomplete will suggest the remaining text of a tag based on previous usage or availability of semantic tags, and administrators will be able to edit or consolidate tags to form an evolving data dictionary. Documents can be searched and retrieved based on metadata criteria. Documents can be added to public or private collections. Users will be able to see a representation of what collections or documents need more attention in terms of tagging.   Users will be able to view Documents via their location data on a map. Any document can be downloaded.

Primary audiences: For the particular content, the primary audiences are those wanting to access records relevant to the build up to and period of early US government boarding schools for Native Americans. The content we put into this tool will be of particular interest to Native community members, genealogists, historians, educators, curators. The tool itself will be useful for all standards-based digital repositories interested in enabling annotations, online collaboration, crowdsourcing, and/or interopability of images. It will be of particular interest to repositories wishing to increase effective access to mass digitized holdings.

A.2   List other existing software that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

A number of digital content management systems such as the opensource Mukurtu and Omeka enable users to register and add tags or comments. The main difference is that our Indigenous Digital Archive software tool will allow users to add tags at the level of a passage or portion of an image rather than at the level of a digital object. This allows making large digitized series of documents available online without individual item cataloging. Additionally, we will provide semantic tagging tools, and the overall toolkit is designed on the international standards of the International Image Interoperability Format (IIIF) and Open Annotation, aiding digital preservation and forward migration of the user-contributed content.

Also different from these existing tools will be our focus on the design of the user interface. We will draw on the emerging work designing generous interfaces, where a search box is replaced by some effective leads into characterizations of the content available and possibilities for exploration.

The software toolkit we will create is uniquely useful in allowing libraries and archives to post records scanned and cataloged in bulk and provide people increased access and learning opportunities with their collections.

B. **Technical Information**

B.1   List the programming languages, platforms, software, or other applications you will use to create your software (systems, tools, apps, etc.) and explain why you chose them.

Working with Omeka makes the choice of application logic and database platform for us: PHP and MySQL. We

would add any additional database tables to MySQL rather than introduce another database.

So, the bulk of the development (and all that requires direct support) is:

Omeka-S with extensions and plugins in PHP for the application logic, MySQL for the database, Git for version control, PHPUnit and Behat for testing, HTML, CSS, JavaScript with jQuery and other common libraries. Hosting is anticipated to be on Amazon Web Services.

Additional applications - if https://www.discourse.org/ is chosen for the "talk" functionality, it is a Ruby on Rails application. This will also be installed, supported, and hosted alongside Omeka. The Universal Viewer (https://github.com/UniversalViewer/universalviewer) would be extended as part of this project if it lacks features required for the annotation scenarios.

External services (for example DLCS) are written in a variety of technologies but they are third party services as far as the IDA project is concerned, so their internal implementation details are opaque.

B.2 Describe how the intended software will extend or interoperate with other existing software.

The IDA tool will be accessible from any web browser. Designing annotation and semantic tagging around the International Image Interoperability Format (IIIF) means that this tool will be able to extend software such as Omeka-S, and can build on developments by IIIF-compatible FromThePage.

B.3 Describe any underlying additional software or system dependencies necessary to run the new software you will create.

The technology stack proposed in B.1 above allows the software system to be accessible via any modern web-browser across all devices. There is no need for special software to be downloaded, or for anything to be installed for users to access the system since it all happens through a web-browser.

B.4 Describe the processes you will use for development documentation and for maintaining and updating technical documentation for users of the software.

Documentation for the technical system will appear in two main places:

1. A main application README which covers the general layout of the application and aspects of it
2. Comments within the code pointing out specific logic and code rationale.

Documentation for users of the system will occur through "How To" sections of the site, along with intuitive guides and tooltips at the appropriate locations.

B.5 Provide the name and URL(s) for examples of any previous software tools or systems your organization has created.

Examples of online digital access projects of other organizations executed by the experts we will contract with for the IDA:

Digirati:
http://wellcomelibrary.org/ for the Wellcome Trust (UK)
Universal Viewer for access of all of the British Library's online collections (under development)

Good, Form & Spectacle:
http://whatsinthelibrary.com/ for the Wellcome Trust
http://wb.britishmuseum.org/ for the British Museum

C. **Access and Use**

C.1 We expect applicants seeking federal funds for software to develop and release these products under an open-source license to maximize access and promote reuse. What ownership rights will your organization assert over the software created, and what conditions will you impose on the access and use of this product? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain any prohibitive terms or conditions of use or access, explain why these terms or conditions are justifiable, and explain how you will notify potential users of the software or system.

The software will be released as open source software on Github under the GNU AGPL v3.0, a Free/Open Source license prior to the end of the performance period.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

The source code will be available via Github.com.

C.3 Identify where you will be publicly depositing source code for the software developed:

Name of publicly accessible source code repository: URL:
Github.com

Part IV. Projects Creating a Dataset

1. Summarize the intended purpose of this data, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

User-generated tags and annotations are anticipated to be generated many times a day. These will be maintained in the annotation server, with regular automated backups to the digital preservation server. Users will generate this data themselves in order to improve navigability of document, take notes or create explanations for others, and to provide supplemental information on topics.

2. Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

Not applicable

3. Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

Not applicable

4. If you will collect additional documentation such as consent agreements along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

Not applicable

5. What will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

The technical stack is described in B.1. Our opensource software tool based on the IIIF and Open Annotation standards will receive annotations and semantic tags. The results will be available on a standard web browser without any special requirements.

6. What documentation (e.g., data documentation, codebooks, etc.) will you capture or create along with the dataset(s)? Where will the documentation be stored, and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

Not applicable

7. What is the plan for archiving, managing, and disseminating data after the completion of the award-funded project?

User generated data (tagging, annotations) will be maintained throughout the project with backups to the digital preservation server. User generated data will continue to be generated after the award performance period as MIAC continues the Indigenous Digital Archive use case repository.

8.   Identify where you will be publicly depositing dataset(s):

Name of repository: Indigenous Digital Archive
URL: http://native-docs.org

9.   When and how frequently will you review this data management plan? How will the implementation be monitored?

We will review the data management plan annually, and monitor implementation through checking that the server fees are paid, logs are consistent with backup needs, and that digital preservation server content is increasing.

Original Preliminary Proposal

The Indigenous Digital Archive: Building Effective Access to and Collaboration with Mass-Digitized Archival Documents related to Native Peoples

The Indigenous Digital Archive project (IDA) is a collaborative project led by the State of New Mexico's Museum of Indian Arts and Culture (MIAC), a library, archives, and museum located in Santa Fe, New Mexico. The Museum of New Mexico Foundation, the applicant, serves as fiscal agent for all grant applications of MIAC.

The Indigenous Digital Archive project will leverage developments brought about by international cooperation in developing particular application programming interface standards (APIs) for image interoperability and an open annotation format to develop tools to allow efficient online access to mass digitized open archival documents. This project takes for its target public government documents that due to typescript quality or inclusion of handwriting or photos are highly resistant to optical character recognition (OCR). This work will extend technological advances that have been developed so far mainly in the arena of medieval manuscripts and individual images, and with newspapers and other printed material in national libraries in Europe, Australia, and New Zealand.[1] This project will extend the potentials of these tools to a next logical conclusion, and draw on the capabilities for tagging and otherwise annotating precise locations within a document, to create a platform where community and crowd sourced keyword tagging and annotation[2] create effective waypoints for navigating through a large mass digitized archival document set. Further, as the ultimate goal of this project is to build an interface that is really focused on creating the best user experience possible, this project will build on the emerging work in developing generous interfaces, user interfaces that rather the present the user with just an empty search box, instead organize and present preliminary faceted data to provide the most information possible to a person in the most apprehendable way possible, with the most effective possibilities for next steps for effective interaction.[3] Following initial build, this project will also include usability testing with a cohort who will over a year's time individually and together undertake exploration with the documents, and the user interface design will be recursively adjusted.

In particular, the project will draw on and extend work enabled by the creation of the International Image Interoperabilty Format (IIIF, http://iiif.io) and the related ecosystem of open source tools, some of which are already incorporating semantic tagging. The IIIF model, drawing on a shared canvas concept familiar from such applications as Photoshop and Powerpoint, provides a model for computationally referring to any point or area on a canvas through a set of coordinates, and provides a powerful standardized tool for being able to refer to or even annotate that area. Annotations created in IIIF compliant systems comply with the W3C standard for Open Annotation, and so are also themselves interoperable with other IIIF compliant systems. (The W3C is the international organization responsible for managing the standards that make the internet able to communicate across different languages and applications.) As annotations made in a IIIF compliant tool are in an international standardized format, they avoid the pitfall of getting stuck within any one particular system, are in a format beneficial for digital preservation and forward migration, and are in a standard API that supports varied user friendly interfaces.

While there are systems that already make it possible for users to add a keyword tag or other annotation, so far this is possible only at the level of the digital object. This means that the annotation applies to the entire image, or entire pdf of a multi-image document, for example. Exciting developments in citizen scientist/archivist crowdsourcing platforms have so far been designed to create structured data from relatively structured and finite portions of records and books, such as creating a tool to allow crowdsourced identification of record type and transcription of a log of troop movements and activities, or transcriptions of text in scientific illustrations. These tools work well to allow people to make incremental contributions to a particular defined and delimited research project.

However, the need still exists to develop the use of annotations as a way of increasing access to and allowing for collaboration among people accessing documents, and allowing a range of annotations that are located at the precise point of reference in the document. Crowdsourced annotations provide a bridge between the situation of records effectively not being available until they are cataloged or described in detail by library or archives staff. The tools produced by this project will meet the IMLS strategic goal of promoting the use of technology to facilitate discovery of knowledge and cultural heritage, and has the potential of providing a means of creating effective access to masses

---

[1] Snydman, Stuart, Rob Sanderson and Tom Cramer. The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images. *Archiving 2015*, Proceedings of the Society for Imaging Science and Technology Conference, Los Angeles, CA, 2015 May 19-22. http://purl.stanford.edu/df650pk4327.

[2] Appleby, Mike. IIIF Tagging and Discovery. Sharing Images of Global Culture, NGA, 2015 May 5 http://bit.ly/1Vtl2AV

[3] Whitelaw, Michael. Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly* 2015 9.1

of public records that are otherwise often considered too big to work with, or other papers or manuscript collections that would benefit from people being able to access and work on them in a collaborative setting.

This work will also benefit any existing CMS that adopts IIIF capability, so, for example, if (and for some, it's already when) they incorporate IIIF compatibility, this tool will be available as an additional enhancement of their platforms. Examples include the national digital library infrastructure DPLA; Mukurtu, a CMS that enables object-level commenting and gating access to content based on tribal affiliation, clan, gender, or age group, for example; and Omeka, the effective opensource package chosen by many smaller institutions and research and digital humanities projects, and now thanks to the IMLS being enhanced with linked open data, and for which a software bridge "shim" for IIIF is currently being built.

In our case, the records we will start with building access to are open public government records relating to the build up to and operations of the US government boarding and day schools for Indians in the period of the Indian Wars up to the reforms of the "Indian New Deal" in the 1930s. In our three year project we will purchase digital copies of extant microfilms of these records, ingest them into the system we build, and work with a cohort of 23 IDA fellows, one from each tribe of New Mexico plus Hopi, to work with the system and provide usability feedback that will inform design improvements, and commission the improvements to the software. We will also conduct outreach and training with and through the state's tribal libraries system, and with the librarian training program of the Indian Pueblo Cultural Center, an institution collaboratively run by all 19 Pueblo tribes. (Year 1: Advisory Board meeting, digital records acquisition, initial design and software build, bulk ingest of records, beginning of outreach. Year 2: recruitment of IDA fellows from tribal communities; training, information and experience sharing convenings including with Advisory Board; sharing through social media and project blog posts; outreach to additional volunteer users; in-production refinements to system. Year 3: conclusion of IDA fellows program and fellows research presentations; synthesis of usability data and interface redesigns or refinements; test use in selected college classes; further outreach including presentations to Society of American Archivists [SAA] and American Library Association.)

The IDA project director, Dr. Anna Naruta-Moya, is a Certified Archivist, holds an SAA Digital Archives Specialist Certification, and has served as an archivist for the Hoover Institution Archives, Stanford University, and the US National Archives. She earned a PhD at the University of California, Berkeley and is an Associate Research Professor at the University of New Mexico and an archivist for the State of New Mexico and as a member of the SAA Standards Committee. A member of the IIIF research and development community, she has extensive experience in project management. MIAC in 2014 conducted tribal constituent needs assessments that articulated the need and desire for online access to archival documents about their histories and for training in archives, and has since conducted pilot studies and other assessments. These two institutions serve as major partners to the project, the State Tribal Libraries system and the collaborative cultural center of all 19 Pueblo tribes, the Indian Pueblo Cultural Center. In addition to an advisory panel including academics and former boarding school students, we will also be guided by technical advisors from Stanford University (Dr. Rob Sanderson, Information Architecture) and the National Library of Wales (Glen Robson, developer experienced in focused crowd-sourcing IIIF applications). UI designed will be guided by George Oates (Good Form and Spectacle, for Wellcome Library, others; formerly of UI for Flickr).

Through the design and build of this tool and the partnerships and programming this project will meet the goals of building and enhancing interoperable tools and services, being inclusive and mobilizing a diverse range of end users, including underserved communities, and initiating radical collaborations (2015 NPD meeting report). Performance goals include creation of the annotation tool and generous interface UI design, ingest of 140,000 pages, sustained usability testing by 23 tribal IDA fellows, at least 4 public blog posts, and release of refined design tool on Github.

Total cost of project is estimated at approximately $447,500 (Estimated $100,000 for software development, UI design, testing, and refinements, $162,000 [as match] for salaries and in-kind time, $30000 [as match] fringe benefits, $500 equipment, $500 printing, $23,000 fellows stipends, $10,000 webhosting and digital preservation repository [as match], $60,000 travel expenses [outreach, conferences, contractor and fellows travel and meeting refreshments], $5000 for meeting refreshments [as match], and $17,500 for digitized records [purchased with matching funds].) $223,000 is requested, and $224,500 is cost share matching.