ABSTRACT

In the face of rapid and far-reaching global changes, the concept of the "Anthropocene" has recently captured the imaginations of scientists, cultural heritage professionals, policy makers, and many public communities. The Anthropocene describes the period when human agency began to measurably alter the environment at regional, continental, and global scales, beginning perhaps as early as 15,000 years ago.

We seek IMLS support to extend the Digital Index of North American Archaeology (DINAA). DINAA helps to map the Anthropocene in the continental United States by documenting the human presence on the landscape since the Pleistocene. It aggregates archaeological and historical data from governmental authorities that manage United States heritage resources. DINAA contributes to the national digital platform by curating and opening access to decades of data collection— all resulting from annual public investments of hundreds of millions of dollars in historical preservation. In making these data available, DINAA will provided critical infrastructure for future indexing of tens of thousands of reports now languishing as nearly inaccessible "grey literature." Open Context, an open access data publishing service for archaeology, hosts DINAA, where researchers and the public can download over 380,000 site file records (with sensitive data redacted), free of charge and intellectual property restrictions. DINAA has already successfully integrated archaeological site data from 15 states (11 currently online), encompassing the rich chronological, cultural, and anthropological metadata used by authorities and researchers alike.

In this project we propose to expand DINAA to encompass the remainder of the United States, adding an estimated two to three million archaeological sites. In doing so, DINAA will provide researchers, museums, libraries, government offices, and the public with a powerful gazetteer of all known archaeological sites in the United States. As a Linked Open Data (LOD) resource, DINAA will help integrate government, museum, library, archive, and scientific datasets, as well as repositories of scientific literature. Over the course of this project, several collaborating scientific data sharing systems will use DINAA for spatial metadata. Similarly, the Phoebe A. Hearst Museum, which maintains the largest archaeological collection in the western U.S., will link DINAA to records in CollectionSpace, a powerful, open source collection management system. This integration will open doors for many other museums to use DINAA, and with LOD methods, further expand the Web of cultural heritage data. To facilitate adoption and use of DINAA, our project will support iterative improvements in design, navigation, and visualization. Assessment of DINAA's use in classroom and heritage management settings will guide improvements.

Finally, DINAA works to make the understanding and stewardship of North American cultural heritage more inclusive. Currently, Native American tribes must interface with siloed and opaque government systems to review relevant cultural heritage data. DINAA makes key descriptive information about North America's rich cultural heritage available for inspection, evaluation, and use by descendant communities, historically marginalized from administrative and political processes. In this sense, DINAA is an "open government" project that will help make cultural heritage management more accountable to wider constituencies, especially descendant communities. Making these data linked and accessible will allow sovereign tribal nations to effectively manage and protect their ancestral cultural heritage, while improving government-to-government relationships, between tribal nations, states, and the U.S. federal government.

Narrative

1. Statement of Need

In the face of rapid and far reaching global changes, the concept of the "Anthropocene" has recently captured the imaginations of scientists, cultural heritage professionals, policy makers, and many public communities. While variably defined, the Anthropocene refers to the period when human agency began to measurably alter climate and biota at regional, continental, and global scales (Lane 2015; Lightfoot and Cuthrell 2015). Our project builds critically needed information infrastructure to document and understand how the Anthropocene unfolded across North America. To this end, we seek IMLS support to extend the Digital Index of North American Archaeology (DINAA). DINAA aggregates archaeological and historical data from state and tribal governmental authorities that manage United States cultural resources.[1] DINAA contributes to the national digital platform by providing the most comprehensive and detailed database documenting human settlement in North America currently available. To date, the public can download over 380,000 site file records (with precise location and other sensitive data redacted) free of charge, and free of intellectual property restrictions, via Open Context (opencontext.org), an open access data publishing service (Fig. 1). These records will cross-reference reports, museum collections, bibliographic references, and other online datasets. The nation's investment in archaeology and historic preservation has produced a vast but poorly curated literature, and this project will help make that effort, often overseen by State Historic Preservation Offices (SHPOs) more accessible to scholars and the public alike.

Recent collaborations among archaeologists, archivists, museum curators and Native American Indian Nations, Tribal Groups, traditional landowners, and other sovereign indigenous groups attempt to better understand and address damage caused by colonialism. Executive Order 13175 - Consultation and Coordination With Indian Tribal Governments (November 6, 2000), initiated federal government policies that further promote such partnerships, leading to collaborative land management, cultural heritage preservation, research, education, and community development programs (eg. Atalay 2008, Colwell-Chanthaphonh et al. 2010, Conkey 2005, Henderson 2000, Lightfoot et al. 2013, Martinez 2012; Watkins 2000; Colwell-Chanthaphonh and Ferguson 2006, Houde 2007, Irlbacher-Fox 2014, Menzies 2006, Simpson 2014, Wiethaus 2007, Wildcat et al. 2014). DINAA builds upon and further enables these new partnerships by making key data more accessible for descendent communities and Native American officials that manage the historical preservation efforts of sovereign tribal nations.

2. Impact

Archaeological data constitute the direct evidence of past human behaviors and are essential for identifying and describing patterns of change in past human societies. Recent estimates demonstrate the magnitude of public investments in archaeology (Altschul and Patterson 2010). Conservatively, the public invests over $500 million per year to comply with historical and archaeological protection measures required by federal law. This level of public investment nearly matches the total *combined* budgets of the IMLS (roughly $240 million in 2015), the NEH (roughly $140 million in 2015), and the NEA (roughly $150 million in 2015). These surprising numbers demonstrate archaeology's relative importance in public cultural heritage investments. Unfortunately, much of this work and investment goes largely unnoticed. Up to now, decades of investment in managing and protecting America's archaeological heritage, have led to few publicly accessible impacts. Cultural resource management (CRM) largely takes place within relatively opaque bureaucratic processes that regulate construction and development. CRM work has resulted in an estimated 350,000 reports nationwide

[1] For more about DINAA, visit: http://ux.opencontext.org/archaeology-site-data/

as of 2004 (NADB 2004), but because of poor access and cataloging, irreplaceable cultural heritage documentation in these "gray literature" reports goes unappreciated or is ignored. Furthermore, since most CRM projects receive minimal attention, reports see little external reuse in research orr other publications that greater peer recognition and review would bring.

DINAA represents a required first step to leverage this tremendous public investment in cultural heritage for wider public impacts and outcomes. DINAA compiles, cleans, and publishes site file data aggregated from state and other agencies that enforce U.S. historical protection laws. Site files include information about periods of use, associated artifact collections and documents, preservation condition, National Register of Historic Places eligibility status, and a host of other variables useful for research and management purposes, as well as geospatial coordinates. In 1993, the last time primary site file data was compiled nationally through the National Park Service's National Archaeological Database (NADB) effort, just under one million archaeological sites had been recorded (NADB Maps 1993). This total has grown over the past two decades (see also Anderson and Sassaman 2012:32) and we estimate the number now to reach nearly 2.5 million sites.

DINAA works to build a comprehensive free and open access inventory of U.S. archaeological and historical resources. DINAA has already successfully integrated archaeological site data from 15 states (11 now public)[2], encompassing the rich chronological, cultural, and anthropological metadata used by both government compliance officials and the research community. In this project, we propose to continue this work to encompass the remainder of the United States, to add an estimated two to three million archaeological sites (Fig. 2). In doing so, DINAA will provide researchers, museums, libraries, government offices, and members of the public with a powerful gazetteer of all documented historical and archaeological sites in the United States.

Field-wide need addressed by the project

DINAA as a stand-alone dataset has important research and public outreach applications. For instance, in documenting sea-level rise threats to tens of thousands of sites known through DINAA, a study by Anderson et al. (in prep) demonstrates how unprecedented data access can reveal the vast scope of cultural preservation challenges resulting from accelerating Anthropocene climate change. However, DINAA's greatest value for the museum and library communities centers on Linked Open Data (LOD) applications. Open Context, like other LOD systems, emphasizes the use of stable Web Uniform Resource Identifiers (URIs, i.e. stable URLs that serve as universally unique "primary key" identifiers) to identify concepts and other entities so they can be easily and precisely referenced and related across different data collections on the Web. DINAA uses Open Context and the EZID service to mint persistent URIs for each site files record. In archaeology and historical geography, the "site" is a key organizational entity. Minting stable Web URIs and offering rich temporal, geographic, and cultural metadata (also available in machine-readable the JSON-LD format) about sites will therefore create significant LOD resources essential for broadly integrating museum, library, and scientific datasets.

It would be naïve and unrealistic to impose a single data standard, expected to be broadly applicable for a continent full of archaeological sites collected by many different organizations for decades, and representing >13,500 years of differing cultures in widely varied environmental settings. Most of the state systems currently in place, in fact, encompass tens of thousands of sites and have been in place for more than half a century leading to many separate database systems with unique constraints on data types and coding solutions. Often overworked and understaffed, state and tribal site file managers lack the resources needed to restructure their datasets to meet external standards (the first

[2] See map of current DINAA data in Open Context: http://bit.ly/1OvF2zR

series of DINAA workshops gave many of our governmental data suppliers a rare chance for professional development with their counterparts from other jurisdictions).

Requiring contributors to implement predetermined data standards will likely to be ignored. Instead, our project imports diversely structured data into Open Context's generalized and abstracted global schema. This global schema implements elements of the CIDOC-CRM, Dublin Core Terms, SKOS, RDFS and other widely used vocabularies to model and represent a dataset's original descriptive system together with some general metadata. We use other Linked Data vocabularies such as Geonames, the British Museum Thesaurus, and the Library of Congress Subject Headings to further annotate project specific descriptions. Our use of annotation preserves original recording schemes while making relationships to more widely used metadata schemes explicit for computation and indexing. In emphasizing annotation over standardization, DINAA avoids the massive costs and disruptions inherent in changing how different states and researchers organize their databases. Although annotation cannot overcome all problems (e.g. missing data, incompatible classification granularity, sampling differences), our approach of schema mapping and annotation at least offers critical near-term functionality, and can open pathways to future improvements. This project builds capacity in the following ways:

1. *Open Data, Reproducible Research:* Open Context, referenced by both NSF and NEH for archaeology grant data management, provides open access data publication services for archaeology and related fields and hosts the massive DINAA dataset. Open Context publishes with open licenses to advance the "open science" goals articulated in the 2013 White House Office of Science and Technology Policy memorandum requiring open access and open data for federally funded research. Our research on integrating editorial practices and public version control (with GitHub; see Kansa et al. 2014) won the "Best Paper" award at the 2014 Digital Curation Conference.
2. *Open Services*: Open Context's services offer powerful and publicly-available RESTful (a simple, best practice for Web architecture) APIs (application program interfaces) to enable interoperability, extensibility, and alternate forms of visualization and user interfaces. This project will use Open Context's APIs for a variety of purposes aimed at maximizing interoperability and extensibility by using widely supported open standards and conventions.
3. *Integrating Existing Infrastructure*: This project capitalizes on prior NSF, NEH, and IMLS investments by promoting data integration and cross referencing across scientific data repositories, museum and library systems. Our project will support iterative interface improvements (search and data visualization), as well as project workshops and training materials to promote wider use of these powerful repositories. Building a wider community will multiply the impact of this project (Fig. 3).
4. *Tribal Nation Collaboration:* Our project will work in close collaboration with Tribal Historic Preservation Offices (THPOs) to exchange knowledge, develop technical capacity by THPOs to evaluate, use, and participate in Linked Data, and ensure that Native American historical perspectives have greater representation in the Web of (cultural heritage) Data.
5. *Assessing Impact:* We will measure our progress via the number of new SHPO datasets successfully published during the project. In addition, we will measure the impact of our efforts to promote DINAA adoption by tracking the number of digital collections that reference DINAA.

3. Project Design

The DINAA project is well poised to achieve significant positive impacts for libraries, museums and researchers with interests in North American prehistory and history. With IMLS support, we can further expand the coverage of DINAA and pilot the following key applications:

1. *Mapping publications*: Since the 1960s, many researchers have published scholarly papers and books identifying historical and archaeological sites with "Smithsonian Trinomials." DINAA curates

Smithsonian Trinomial identifiers, and with IMLS support we will pilot text-mining of literature in JSTOR to find trinomials and associate these with DINAA records. DINAA sites discovered in JSTOR will be documented by the Dublin Core Terms `IsReferencedBy` relation linking to the article's URI, and indexed by Open Context. This will power map-based search and browse interfaces to discover scholarly literature. We can also display "heat maps" showing where academic scholarship has focused, helping to illustrate the history and focus of research.

2. *Cross-referencing with other data sources*: By matching Smithsonian Trinomials, we have established links between DINAA site file records and metadata records in other datasets and repositories. These include the Paleoindian Database of the Americas, the Eastern Woodlands Household Archaeological Data Project, and tDAR (a digital repository for North American archaeology, managed by Digital Antiquity). We have developed powerful entity reconciliation services to enable others to find DINAA URIs (and other metadata) for Smithsonian Trinomials. With IMLS support we will host workshops and develop online training materials to help others, especially libraries and museums, use DINAA as linked data to enhance their metadata and broaden the impact and reach of their collections. The Phoebe Hearst Museum (UC Berkeley), the Canadian Radiocarbon Database (CARD; with extensive U.S. coverage), and others (see *Partner Letters*) will use DINAA for spatial metadata documentation (see Fig. 3–5).
3. *Developing APIs for entity reconciliation and data reuse*: Open Context offers powerful APIs for data discovery, reuse, visualization, and entity reconciliation. For example, rOpenSci (rOpenSci.org) created an open-source R statistical programming client for use with Open Context. IMLS funds will support improvements in the usability of Open Context's APIs that serve DINAA data. Open Refine, a popular open source application for cleaning and annotating structured data, can send requests to external APIs hosted on the Web. Our APIs and OpenRefine will power entity-reconciliation to match site identifiers or site names against DINAA URI identifiers. This will enable researchers to link their data to DINAA, promoting data integration. We used this approach to cross-reference DINAA with metadata records in tDAR, the Eastern Woodlands Household Archaeological Data Base, and the Paleoindian Database of the Americas. Feedback from additional museum and research data implementations of DINAA will help improve these API services and develop better documentation and "recipes" for entity reconciliation.

Contributions to the National Digital Platform: Digital Public Library of America

DINAA will provide key contributions to the national digital platform by offering a critically needed gazetteer of U.S. archaeological sites to enable indexing and discovery of museum collections, grey literature, and scholarship documenting America's rich cultural heritage. The Digital Public Library of America (DPLA) has emerged as a key aggregator and discovery portal to many cultural heritage collections from American libraries, archives, and museums. However, since DINAA aggregates administrative database records, DINAA by itself is not directly suitable for inclusion into DPLA as a content provider. Instead, DINAA will work like a specialized GeoNames.org by providing key geospatial metadata suitable for aggregators like DPLA. The DPLA Metadata Application Profile describes a metadata scheme for modeling relationships between content and geographic place entities (DPLA 2015:15-16). DINAA provides the label, geospatial coordinates (in DPLA-supported WGS-84 and GeoJSON specifications), and notes needed to describe place entities for DPLA. Moreover, the DPLA Metadata Application Profile also supports reference to DINAA URIs through the `SKOS:exactMatch` relation (DPLA 2015: 16). Finally, DINAA has collaborated with the PeriodO project since its inception, and will incorporate PeriodO URIs to document sites with chronological metadata. Like DINAA, PeriodO has described a model that is well suited for use by the DPLA Metadata Application Profile (see DPLA 2015: 16 if extended with `SKOS:exactMatch` URI references). Thus, DINAA place entities, chronological periodization, and other descriptions can be used immediately as geospatial metadata for DPLA. Documentation produced by this project will help library, museum and archive managers understand how to use DINAA as DPLA place metadata. Consultation with digital library experts Christina Harlow (Linked Data, metadata) and Susan Powell (geospatial) will help maximize DINAA's value for library information systems (see *Partner Letters*).

Site Security Measures

The security of archaeological sites must be protected for ethical and legal reasons. In the United States, the locations of archaeological sites are highly sensitive data and their release could have grave repercussions. It is difficult to develop adequate information security measures for public-facing websites and prevent accidental data releases or data theft through hacking and other leaks. Even if we deployed appropriate security measures, our systems would need extensive auditing for compliance to Archaeological Resource Protection Act (ARPA) regulations and our project team would be legally liable for any release of sensitive data. For these reasons, managing sensitive site location data lies beyond the scope of this project, and no such information will be released.

To eliminate the risk of accidental or malicious disclosure of sensitive data, DINAA will only store and release spatial coordinates at a reduced level of geographic precision (in a roughly 20 km grid cell). We will negotiate the exact spatial resolution we will use for public data with SHPO and agency personnel; we expect it to be at the 20 km resolution which is used in the current iteration of DINAA, or no larger than county level, which was used in earlier efforts (e.g., Anderson and Horak 1995; NADB Maps 1993). Site file managers for 15 states have already endorsed these measures for site protection and have shared data now public with DINAA. DINAA's 20 km resolution still permits important research programs and LOD applications. DINAA will associate appropriate SHPO contact information with each data record to enable qualified investigators to directly obtain higher resolution spatial data from state officials.

Entity Identification in JSTOR

LOD can play an important role in revitalizing archaeology's vast legacy of literature. Project Director E. Kansa's experience with the Google-funded "Google Ancient Places" (GAP) project helps illustrate how a site-file gazetteer can be used in LOD programs (Barker et al. 2011). The GAP project used the openly licensed Pleiades Gazetteer (hosted by New York University) and the open source Edinburgh GeoParser (entity-identification software) to automate the identification of ancient places discussed in literature (books digitized by Google). That project resulted in novel information retrieval, book visualization, and mapping tools, as well as quantitative data on co-occurrence of place references in texts. DINAA can facilitate similar text-mining powered services for North American archeological literature and grey literature. The current project will conduct exploratory "proof-of-concept" text-mining on relevant journals in the JSTOR repository, which offers researchers text-mining services (JSTOR 2015). Initial testing showed that specialized text analysis software treats alphanumeric Smithsonian trinomial strings unpredictably, even with simple tokenization. Therefore, we favor simpler and more readily debugged approaches that use string matching regular-expressions to find Smithsonian trinomials in the abstracts, titles, keywords, and contents of articles provided by JSTOR. With student help, Wells, Yerka, and Anderson will manually identify Trinomials in a sample of at least 100 articles from different journals and years (to account for variation in Trinomial formatting styles). Their manual identification will help validate automated Trinomial identifications, helping us trouble-shoot string matching algorithms to improve precision and recall. JSTOR bibliographic metadata will be added to appropriate DINAA records, and will be indexed and visualized through map-based interfaces on Open Context. Given the variability in how authors express trinomials in articles and diversity in colloquial site names, this initial text-mining effort will be preliminary and exploratory rather than comprehensive, although we do expect to generate large numbers of linkages. Nevertheless, it will demonstrate ways to enhance discovery of archaeological literature (see Kintigh 2015). This invaluable experience will guide future entity identification efforts beyond JSTOR, to also encompass the Hathi Trust (digitized books), reports in tDAR, and other document archives that contain poorly catalogued gray-literature reports.

Learning from Implementation: Use of DINAA by the Phoebe A. Hearst Museum and the Chippewa Cree Tribal Historic Preservation Office

DINAA will cross-reference diverse museum, library, and archival resources. However, using DINAA may involve a number of technical challenges for organizations that may have limited technical support and staffing. In our experience, efforts to reuse data offer some of the best ways to discover problems in data and data services (see Kansa et al. 2014). Therefore, this project will not only work to expand DINAA's coverage, it will dedicate significant effort to understanding and facilitating (re)use of DINAA. To build experience needed to guide institutions in using DINAA, this project will support implementation in two institutional settings, the Phoebe A. Hearst Museum and the Chippewa Cree THPO office. These institutional settings provide good representation of DINAA's core stakeholders. Carefully documenting and responding to implementation challenges will help improve services and user interfaces, as well as develop guidance and "how-to" manuals that will reduce the costs and difficulties of future implementation.

Founded in 1901, the Hearst Museum has a global collection of approximately 3.8 million objects and media, making it the largest anthropology museum west of the Mississippi and one of the largest in North America. The Museum cares for over 1 million objects from California and Nevada. Collections from the Pacific Northwest, Southwest, and other parts of the U.S. complement the California collection. The Museum's collections are documented in CollectionSpace, an open source museum information system sponsored by the Andrew W. Mellon Foundation. CollectionSpace is used by several institutions that also curate significant North American collections (most notably the San Diego Museum of Man). Extending CollectionSpace to use DINAA, as well as documenting cost-effective implementation methods, will help future institutional partnerships.

Similarly, collaboration with the Chippewa Cree THPO (with additional consultation with the Eastern Band of Cherokee Indians THPO) will be essential in crafting and promoting ethical and responsive "best practices" in developing and using DINAA. The Chippewa Cree THPO works within the larger Chippewa Cree Cultural Resources Preservation Office on Rocky Boy's Indian Reservation, and has developed a tribal cultural monitoring program and consultation database, Tribal106, that utilizes the traditional knowledge of tribal members in meeting the requirements of NHPA Section 106 CFR Part 800 to allow commentary by stakeholders on the effects of undertaking on identified historic and culturally significant properties. During the initiation of tribal consultation, THPO representatives receive detailed archaeological survey reports for each project generated by cultural resource management firms, but these often lack general background information on the archaeological resources previously documented in each area by SHPOs. Team member Noack Myers (with the Chippewa Cree THPO) identified the following needs:

- State agencies do not allow access to their databases without archaeological credentials, in effect, keeping information from tribal communities. DINAA must reduce these barriers.
- Tribal community members need multiple routes to find information. Straightforward user interfaces, direct links from THPO webpages, and other measures may be required.
- Tribal interests extend over multiple state boundaries. The Chippewa Cree monitors a ten-state area and the Eastern Shoshone monitors a 16-state area. By aggregating across state lines, DINAA can facilitate discovery of needed information with good search and mapping features.

- Technical jargon and complexity will limit use. DINAA needs to develop clear and accessible tutorials, especially videos and explanatory graphics.

Interface Design and User Experience Methodology and Assessment
DINAA aims to serve diverse communities in many sectors. For this reason, interface design and user experience factors directly impact the success of the project. IMLS support will improve the search, navigation, and data visualization functions on Open Context so that published DINAA datasets will be accessible to end users. During this phase of development, we will emphasize "findability and discoverability" (see Morville 20015) to enable easier browsing of DINAA datasets, and to improve user encounters with content and functionality unique to DINAA in the Open Context platform. We will address issues of representing DINAA's large, cross-referenced databases in user-friendly ways by developing interface designs that orient to user needs and values, and read clearly across multiple interaction channels (Web, tablets, smartphones, etc.). We will direct specific attention towards data visualization efforts that help end users access and understand the extensive scope and depth of DINAA. Key interface components that are critical for the success of expanding DINAA datasets across the United States include Open Context's faceted search interface, data output functionality (tables), and site navigation (efficient workflows and successful task completion) (Rosenfeld et al. 2015). The aim of investing in these usability improvements as a part of scaling-up the number of DINAA datasets is to reduce friction in user experiences as the database grows in size and complexity. This should help reduce future development and support costs due to usability issues over both the short and long term (Bias and Mayhew 2005).

The Masters of Arts in Cultural Resource Management (CRM) program at Adams State University (ASU) has agreed to test and assess DINAA in a classroom setting (see *Partner Letters*). DINAA will provide pedagogical support (instruction packets, sample queries, videoconference lectures) for the use of the DINAA Web interface as part of their curricular training in archaeological data management. In return, ASU faculty will provide feedback on their classroom implementation of DINAA, and will provide students with contact information to voluntarily submit their own assessments. This agreement provides DINAA with an advantageous combination of both dedicated online delivery and a user group familiar with archaeology and the heritage management processes behind DINAA. To facilitate usability goals, Phoebe France will conduct targeted web application design evaluations; gather user feedback through interactions with the other project researchers and end users (for example, team collaborations and workshops as outlined in this proposal); and design interface improvements. Qualitative assessment and use of Piwik (web tracking with privacy protections) will help P. France evaluate the success of iterative design improvements.

4. DIVERSITY PLAN
Our project recognizes significant challenges in ethical data management, especially given the often tragic histories of colonialism and appropriation of indigenous land, arts, and culture (Chandler and Sunder 2004; Kansa et al. 2005; Kansa 2012; Christen 2012, 2015). Recently, the NEH and IMLS invested in projects like Mukurtu to address indigenous information privacy needs. DINAA complements these prior investments. Empowering communities with respect to digital cultural heritage involves a host of issues beyond access controls and intellectual property claims (the focus of Mukurtu). Native American communities must also interface with sometimes opaque and unresponsive government agencies that hold relevant cultural heritage data. DINAA makes key information used in the management and preservation of North America's rich cultural heritage available for inspection, evaluation, and use by descendant communities that have often been marginalized from administrative and political processes. In this sense, DINAA is an "open government" project that will make cultural heritage management more accountable to wider

constituencies, especially descendant communities, and will improve the government-to-government relationships that are essential to cultural heritage management by sovereign tribal nations.

While DINAA itself is and will be freely accessible open data, it will empower tribal and non-tribal institutions managing sensitive, access-restricted data. Interoperability measures between tDAR and Open Context illustrate synergies between open data and access-restricted systems. Open Context's login-free and highly granular data will facilitate access and use of site data with location information redacted (see *Site Data Sensitivity and Security Measures,* above), for museums, educators, and research applications built on the Open Web. At the same time, DINAA URIs minted for these sites can be used to discover documents and data held in the more restricted-access tDAR repository. DINAA records can similarly point to information systems governed by tribal agencies or museum partners that assert their own access and reuse rules. DINAA's federated approach to data management will enable interoperability while promoting local oversight and governance of information access.

While facilitating federated approaches to information access, DINAA will also help highlight representation issues in classification. To do so, DINAA will adopt PeriodO for modeling historical periods and chronological schemes. PeriodO provides a general model for classifying and modeling chronological periods, but it does not demand agreement where agreement does not exist. This opens the door for including indigenous community perspectives in organizing histories represented in DINAA site files. Besides periodization, alternative classifications can be created (and modeled formerly with SKOS or OWL) for other descriptive attributes. Thus, the project will invite THPO representatives to supply alternative classification schemes to model site file data aggregated in DINAA. These alternative classifications will be displayed and indexed by Open Context along with the government agency provided classification schemes. In that sense, this project will complement other projects that seek to better align digitized cultural heritage with community needs.

This project integrates collaboration with THPO representatives and other cultural heritage experts for all activities. The project organizes collaboration most closely with the Chippewa Cree and Eastern Band of Cherokee Indians THPOs, but will also gain guidance from the Hearst Museum's Native American Advisory Council and the Native American Graves Protection and Repatriation Act (NAGPRA) (25 U.S.C. 3001 et seq) office at Indiana University. Due to the composition of its North American ethnographic and archaeological collections, the Hearst Museum is among the nation's most active institutions in compliance with NAGPRA. The Hearst Museum's dedicated Cultural Policy and Repatriation (CPR) Division processes multiple NAGPRA claims at a time, conducting intensive scholarly research and consulting collaboratively with claimant tribes to gather evidence for review by Campus and UC System administrators. The CPR coordinates due diligence research for potential acquisitions, helping to resolve Museum registration issues, reaching out to stakeholders for new exhibitions and educational programs, conducting training for staff and volunteers, and vetting social media outreach for cultural sensitivity. The Hearst Museum has found that the substantial, collaborative, and lasting relationships that develop through NAGPRA consultations inform and improve many other areas of the Museum's work. The Hearst Museum Native American Advisory Council (NAAC) advises the Museum on its relationships with California and Nevada Native communities. Created in 2013, the eleven-member committee is made up of individuals with a wide range of backgrounds— including tribal officials, scholars, museum professionals, and artists—from tribes from both states, both federally recognized and not. The Council advises the Museum's work on matters ranging from repatriation policy to loans, exhibitions, educational programs and traditional collections care. Our team will work with the Hearst CPR Division and the NAAC to supplement guidance from THPO partners so that DINAA can better meet a broad set of tribal oversight, education and outreach goals.

5. Project Resources: Personnel, Time, Budget

This project emphasizes expanding an existing dataset and service, incrementally and iteratively improving user interfaces, and widening community engagement. As such, there are few dependencies that require completion to allow other activities to begin. These factors reduce risk of cost and schedule problems. Basecamp and GitHub will support project management and software enhancements. The budget for our IMLS request reflects the project priorities for Native American engagement, improved usability, cross-referencing with literature (JSTOR), and expanded coverage (see *Budget Justification* for details). Our project team has successfully collaborated on conceptualizing and developing the DINAA project for the past three years. Consultants have been chosen based on previous collaboration and relevant expertise. All team members have access to the requisite computing facilities and services, as well as professional research, outreach, instruction and data dissemination goals that motivate time and effort commitments to DINAA. The *Budget Justification* and *Schedule of Completion* provide further details about the timing of tasks, commitment level, and budget for each person listed below.

Project Director Eric Kansa, PhD, RPA, directs software development of Open Context and will implement design improvements provided by P. France (below) and develop, execute and share results of entity identification on the JSTOR corpus. He has technical proficiencies in Python, PHP, Javascript, relational databases, information retrieval, text analysis for entity identification, linked data, and RESTful API deployment. His research explores archaeological informatics, intellectual property issues in cultural heritage, scholarly communications, and research data management.
Sarah Whitcher Kansa, PhD, RPA, Executive Editor for Open Context & Director of the Alexandria Archive Institute, will provide overall project management for this project. She oversees the full cycle of data publication, from solicitation and/or management of submissions to archiving with the CDL. She is responsible for planning all aspects of the major workshop in Year 2. She has 15 years' experience in nonprofit management and will oversee the financial aspects of the grant.
Joshua Wells, PhD, RPA, is Associate Professor of Social Informatics at Indiana University South Bend. His role in this project centers on ontology annotation and research and instructional applications of DINAA. Wells will ensure the security of sensitive site information through redaction, changes in granularity, and encryption, as necessary. He provides access to Indiana University's IUScholarWorks digital repository for additional archiving of project content.
David G. Anderson, Professor of Anthropology at the University of Tennessee, Knoxville, will assist with the collection, integration, analysis and reporting on project site file datasets and the site number data mining exercise. As an expert in North American Archaeology, he will advise, evaluate, and oversee metadata documentation, especially regarding chronological periodization.
Stephen J. Yerka, MA, RPA, will assist on all technical aspects of the project, and will work with Kelsey Noack Myers to provide outreach and assistance to tribal archaeologists and federally-funded THPOs. Yerka is co-director of PIDBA, co-PI on the first phase of DINAA, former GIS/IT manager for the Archaeological Research Laboratory, University of Tennessee, and is published in American archaeology and informatics practices.
Kelsey Noack Myers, MA, RPA, Tribal Archaeologist for the Chippewa Cree of Rocky Boy's Indian Reservation in Montana, oversees Section 106 compliance and cultural resource consultation both on-reservation and across the ten-state area identified and federally recognized as ancestral homeland for the Chippewa Cree people. She will provide outreach and assistance to other tribal archaeologists and federally-funded THPOs through her national professional networks. She will coordinate with tribal representatives to provide feedback on the content, ontology, user experience, and use of DINAA for tribal projects. She will also help develop DINAA instruction materials, also applicable to tribal schools and colleges.

Phoebe France, MA, will lead the interface design and usability aspects of the proposed project. She has conducted user testing and interaction design in the commercial and non-profit sectors, and cross-cultural interaction design and development in Cambodia, and for Meedan.org, a multilingual nonprofit communications system serving communities in the Middle East.
John Lowe, PhD, is a Service Manager & Senior Software Engineer for UC Berkeley's Research IT division and leads software development for CollectionSpace, the system managing Hearst collection data. For this project, he will enable Collection Space to reference external Linked Data gazetteers, including DINAA, Geonames, and others.

6. COMMUNICATIONS PLAN

The proposed work has a global reach, serving communities interested in the archaeological and historical past. DINAA focuses on addressing the needs of Native American cultural heritage experts and professionals, as well as museums, libraries, instructors, and researchers. This project takes an iterative approach to development by responding to identified needs in a cyclical manner. Our *Schedule of Completion* illustrates this iterative process, with key team members participating in annual meetings of Native American cultural heritage experts, including the Hearst Museum's Native American Advisory Council (NAAC), and the annual Native American Tribal Historic Preservation Officers (NATHPO) conference. Regular engagement with SHPO and THPO representatives, among others, at these various venues will help DINAA better align to diverse needs. Based on prior successes, the thematic workshop proposed here will build lasting community and collaborative ties. DINAA datasets will be published open access over the Web using Open Context. We will also present DINAA at annual conferences of NATHPO, as well as at archaeology and digital library venues, such as the Society for American Archaeology, and the Digital Curation Conference. Appropriate publication venues include: *Advances in Archaeological Practice*, *American Antiquity*, the *International Journal of Data Curation*, and the *Data Science Journal.* All publications will be open access (at minimum via self-archiving in institutional repositories). Social media, including blogging, Twitter, and Facebook will also share project news and updates, and enable conversation.

7. SUSTAINABILITY

Charging for access to data created by public agencies, funded by taxpayer dollars, poses ethical problems and would greatly undermine the public benefit of this project. Therefore, DINAA makes all data open access under a Creative Commons Zero (CC0) public domain dedication. A variety of complementary approaches are used to promote the sustainability of DINAA beyond the lifespan of this proposed project, these include income from Open Context's research data management services, consulting and training services, and philanthropic donations. Beyond the digital preservation methods discussed in the *Digital Stewardship Supplementary Information Form*, we recognize that sustainability needs extend beyond preservation of data created by this project. As a comprehensive map of U.S. archaeological sites, and in its key role in linking diverse museum, archive, and library collections, DINAA will play a central role in the stewardship of cultural heritage across America. Once completed, DINAA will require ongoing maintenance, curation, and updates as SHPOs register new sites. DINAA will need organizational support and a governing body after this initial phase of development. Therefore, we budgeted for a major workshop hosted at the Hearst Museum that will both promote the use of DINAA and start planning among key stakeholders about DINAA's long term continuity, governance, and growth.

Building a Gazetteer of Anthropocene North America (The Alexandria Archive Institute)

Schedule of Completion*

*Black = intense activity; Gray = less intense activity

Year 1 (July 1, 2016 – June 30, 2017)	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Data Collection & Interface Design												
Collect SHPO / THPO / NPS databases	■	■	■	■	■	■	▒	▒	▒	▒	▒	▒
Clean databases, redact location/sensitive info, link to OA database/GIS platforms	▒	▒	▒	▒	▒	▒	■	■	■	■	■	■
Collaborate w/ PeriodO, metadata annotation							▒	▒	▒	▒	▒	▒
Publish integrated data												■
Interaction evaluation, interface redesign, implementation (iterative)	■	■	■	■	■	■	■	■	■	■	■	■
JSTOR Entity Reconciliation												
Obtain full-text articles from JSTOR	■											
Manual trinomial ID of sample articles		■	■	■	■	■						
Iterative development & testing of automated trinomial matching of JSTOR articles					■	■	▒	▒	▒	▒	▒	
Associate DINAA records with JSTOR articles							▒	▒	▒	▒	▒	▒
Meetings & Implementation												
Hearst Museum CollectionSpace extension							■					
Native American Council at Hearst Museum				■								
Annual NATHPO conference			■									

Year 2 (July 1, 2017 – June 30, 2018)	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Data Collection & Interface Design												
Collect SHPO / THPO / NPS databases	▒	▒	▒	▒	▒							
Clean databases, redact location/sensitive info, link to OA database/GIS platforms	■	■	■	■	■							
Collaborate w/ PeriodO, metadata annotation	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒		
Publish integrated data	■	■	■	■	■	■	▒	▒	▒	▒		
Interaction evaluation, interface redesign, implementation (iterative)	■	■	■	■	■	■	■	■	■	■		
JSTOR Entity Reconciliation												
Associate DINAA records with JSTOR articles	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒		
Public Presentation / Demo Best Practices												
Summarize / synthesize lessons learned					▒	▒	■	▒	▒	▒	■	■
Post complete description of project effort											■	■
Meetings, Implementation & Workshop												
Hearst Museum DINAA cross-referencing	▒	▒	▒									
Native American Council at Hearst Museum				■								
Annual NATHPO conference		■										
Year 2 DINAA Workshop (and planning)			▒	▒	▒	▒	■	■				

Data Collection & Integration: DA**, JW, SY will collect data from SHPOs and redact sensitive fields. SY, SK will train and supervise RAs/students to use Open Refine to fix common typographic errors and normalize vocabularies. SY, JW, SK, EK will add PeriodO annotations and metadata documentation. SK, EK will incrementally published cleaned and documented data, on a state by state basis, with Open Context (see *Digital Stewardship Form* for details).
Iterative Interface Design & User Experience Improvements: Iterative interface improvements (esp. search and data visualization), as well as project workshops and training materials, will promote wider use of these powerful repositories. Interface design and user feedback will occur monthly. We will collect feedback, formulate design responses, and implement solutions on a cyclical basis. A recurrent schedule of usability best practices will help increase team member and stakeholder communication around usability issues, and allow each new design cycle to build upon the lessons of previous iterations. The initial interface assessment will consist of a heuristic review of the Open Context site design and architecture to report usability issues related to search, navigation, and data visualization. Key improvements will be designed to contextualize solutions for the specific needs of Open Context and DINAA. Some implementations will be immediate, while others may be tested with target end user groups. The timing of feedback will coincide with scheduling for other project collaborations, such as workshops and institutional partnerships with the Hearst Museum and the Chippewa Cree THPO, to maximize the benefits of cross-organization team efforts and iterative design processes. GitHub issue tracking will support project management and communication.
JSTOR Entity Reconciliation: After gaining JSTOR full-text article data, DA will select 100 papers for manual (human) identification of Smithsonian Trinomials. He will supervise RAs/students in manual trinomial identification. Two people will review each paper independently to validate results. EK, JW will develop Python scripts to use regular expressions to identify trinomials and compare with manual identifications, iteratively improving until results converge with manual identification. EK will annotate DINAA records in Open Context with relevant bibliographic information and URI references to JSTOR article URIs containing matched trinomial IDs.
Hearst/CollectionSpace Implementation: JL, EK will perform entity-reconciliation activities to link Hearst Museum provenance information with DINAA. JL will extend CollectionSpace to include references to gazetteer URIs for use with DINAA (and other linked data gazetteers).
Training and Instructional Media: All project participants, especially KM and PF will contribute to development of instructional media to promote the effective use of DINAA. SK will finalize materials and post them on Open Context or the Heritage Bytes weblog.
Meetings & Workshop: We will select key team members to attend Hearst Museum Native American Council (NAC) meetings and NATHPO meetings THPO professionals. Attending these meetings will help obtain guidance, additional collaborations (for design and user testing), and collaboration for creating culturally-relevant metadata documentation. Travel to the NAC meeting will also be used for face-to-face meetings for our project team. A Year 2 workshop will involve DINAA project team members and data contributors, prototype testers, and knowledgeable representatives of cooperating entities that manage related databases. Participants will be encouraged to discuss their medium and long-term expectations for the DINAA, how their own efforts can be improved, and how legal and ethical codes are implemented by agencies. Participants will better understand how to use DINAA for research, management, and collection documentation. A key outcome will include plans for continuity, sustainability and governance.

** See *Key Project Staff and Consultants* for names of team members

REFERENCES CITED

Altschul, J.H. and T. Patterson. 2010. Trends in Employment and Training in American Archaeology. In *Voices in American Archaeology*, edited by Wendy Ashmore, Barbara Mills, and Dorothy Lippert, pp. 291-316. Society for American Archaeology Press, Washington, D.C.

Anderson, D.G., T. Bissett, S. Yerka, J. Wells, E. Kansa, S. Kansa, R. Demuth, and K. Myers. No Date. Climate Change and Archaeological Site Destruction: An Example from the Southeastern United States Using DINAA (Digital Index of North American Archaeology). Manuscript in possession of authors for submission to PLOS ONE in 2016.

Anderson, D.G., and V. Horak, editors. 1995. *Archaeological Site File Management: A Southeastern Perspective*. Interagency Archeological Services Division, National Park Service, Southeast Regional Office, Atlanta, Georgia.

Anderson, D.G., A.M. Smallwood, and D.S. Miller. 2015. Pleistocene Human Settlement in the Southeastern United States: Current Evidence and Future Directions. *Paleoamerica* 1(1):7–51.

Atalay, S. 2008. Multivocality and Indigenous Archaeologies. In *Evaluating Multiple Narratives*, edited by Junko Habu, Clare Fawcett, and John M. Matsunaga, pp. 29–44. Springer, New York.

Barker, E., K. Byrne, L. Isaksen, and E. Kansa. 2011. Googling Ancient Places. Paper presented at Digital Humanities 2011, Stanford University Palo Alto, CA, 17-22 June 2011. Also available online at http://googleancientplaces.files.wordpress.com/2011/07/gap_dh11.pdf.

Bias, R.G. and D.J. Mayhew. 2005. *Cost-Justifying Usability: An Update for the Internet Age, Second Edition*. Morgan Kaufmann Publishers, San Francisco.

Canadian Archaeological Radiocarbon Database. "Future Directions." Accessed December 16, 2015. http://www.canadianarchaeology.ca/future_directions.

Chander, A. and M. Sunder. 2004. The Romance of the Public Domain. *California Law Review* 92. Retrieved April 6, 2012 online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=562301.

Christen, K. 2015. "Tribal Archives, Traditional Knowledge, and Local Contexts: Why the 's' Matters." *Journal of Western Archives* 6(1). Retrieved December 1, 2015 online at http://digitalcommons.usu.edu/westernarchives/vol6/iss1/3

Christen, K. 2012. "Does Information Really Want to Be Free? Indigenous Knowledge Systems and the Question of Openness." *International Journal of Communication* 6(0):24. Retrieved June 1, 2015 online at http://ijoc.org/index.php/ijoc/article/view/1618.

Clarke, M., 2015. The Digital Dilemma Preservation and the Digital Archaeological Record. *Advances in Archaeological Practice 3*(4), pp.313-330.

Colwell-Chanthaphonh, C. and T.J. Ferguson. 2006. Memory Pieces and Footprints: Multivocality and the Meanings of Ancient Times and Ancestral Places among the Zuni and Hopi. *American Anthropologist* 108(1): 148–162.

Colwell-Chanthaphonh, C., T.J. Ferguson, D. Lippert, R. McGuire, G. Nicholas, J. Watkins, and L. Zimmerman. 2010. The Premise and Promise of Indigenous Archaeology. *American Antiquity* 75(2): 228–238.

Conkey, M. 2005. Dwelling at the Margins, Action at the Intersection? Feminist and Indigenous Archaeologies, 2005. *Archaeologies* 1(1): 9–59.

DPLA (Digital Public Library of America). 2015. Metadata Application Profile, version 4.0. Retrieved June 1, 2015 online at http://dp.la/info/wp-content/uploads/2015/03/MAPv4.pdf.

Faniel, I., E. Kansa, S.W. Kansa, J. Barrera-Gomez, and E. Yakel. 2013. The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. *JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 295-304. New York, NY: ACM [http://dx.doi.org/10.1353/ema.2013.0007].

FISH (Forum on Information Standards in Heritage). 2005. FISH Interoperability Toolkit. English Heritage and the Council for British Archaeology. Retrieved June 1, 2015 online at http://www.heritage-standards.org.uk/.

Henderson, J. 2000. The Context of the State of Nature. In *Reclaiming Indigenous Voice and Vision*, edited by Marie Ann Battiste, pp. 11-38. UBC Press, Vancouver.

Houde, N. 2007. The Six Faces of Traditional Ecological Knowledge: Challenges and Opportunities for Canadian Co-Management Arrangements. *Ecology and Society* 12(2). Retrieved June 1, 2015 online at http://www.ecologyandsociety.org/vol12/iss2/art34/.

ITRDB (The International Tree Ring Database). 2015. Available online at http://www.ncdc.noaa.gov/data- access/paleoclimatology-data/datasets/tree-ring.

Iowa Office of the State Archaeologist (IOSA). "Florida and Georgia Site Files Launch DINAA Project – Heritage Bytes." *Iowa Archaeology*. Retrieved December 16, 2015 online at http://iowaarchaeology.tumblr.com/post/80072317378/florida-and-georgia-site-files-launch-dinaa.

Irlbacher-Fox, S. 2014. Traditional Knowledge, Co-Existence and Co-Resistance. *Decolonization: Indigeneity, Education & Society* 3(3):145-158.

JSTOR. Data for Research. *About JSTOR*. Retrieved December 16, 2015 online at http://about.jstor.org/service/data-for-research.

Kansa E.C., S.W. Kansa, and B. Arbuckle. 2014a. Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology. *International Journal of Digital Curation* 9(1): 57–70. [http://dx.doi.org/10.2218/ijdc.v9i1.301].

Kansa, E.C., S.W. Kansa, M.M. Burton, and C. Stankowski. 2010. Googling the Grey: Open Data, Web Services, and Semantics. *Archaeologies* 6(2):301-326. Open access at https://escholarship.org/uc/item/8jc6s6zn.

Kansa, E.C. 2009. Indigenous Heritage and the Digital Commons. In *Traditional Knowledge, Traditional Cultural Expressions and Intellectual Property Law in the Asia-Pacific Region*, edited by C. Antons, pp. 219–44. Kluwer Law International, New York.

Kansa, E.C., J. Schultz, and A.N. Bissell. 2005. Protecting Traditional Knowledge and Expanding Access to Scientific Data: Juxtaposing Intellectual Property Agendas via a 'Some Rights Reserved' Model. *International Journal of Cultural Property* 12(03):285–314. Open access at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.3323&rep=rep1&type=pdf).

King, T.F. 2007. *Saving Places that Matter A Citizen's Guide to the National Historic Preservation Act*. Left Coast Press, Walnut Creek California.

Kintigh, K.W., 2015. Extracting Information from Archaeological Texts. *Open Archaeology* 1:96-101. DOI:10.1515/opar-2015-0004.

Kintigh, K.W., J.H. Altschul,, M.C. Beaudry, R.D. Drennan, A.P. Kinzig, T.A. Kohler, W.F. Limp, H.D.G. Maschner, W.K. Michener, T.R. Pauketat, P. Peregrine, J.A. Sabloff, T.J. Wilkinson, H.T. Wright, and M.A. Zeder. 2014a. Grand Challenges for Archaeology. *American Antiquity* 79(1):5-24.

Kintigh, K.W., J.H. Altschul,, M.C. Beaudry, R.D. Drennan, A.P. Kinzig, T.A. Kohler, W.F. Limp, H.D.G. Maschner, W.K. Michener, T.R. Pauketat, P. Peregrine, J.A. Sabloff, T.J. Wilkinson, H.T. Wright, and M.A. Zeder. 2014b. Grand Challenges for Archaeology. *Proceedings of the National Academy of Science* 111(3): 879-880.

Lane, P.J. 2015. Archaeology in the age of the Anthropocene: A Critical Assessment of its Scope and Societal Contributions. *Journal of Field Archaeology* 40(5):485-498. DOI: 10.1179/2042458215Y.0000000022.

Lightfoot, K.G., and R.Q. Cuthrell. 2015. Anthropogenic burning and the Anthropocene in late-Holocene California. *The Holocene* 25:1581-1587.

Lightfoot, K., R. Cuthrell, C. Striplen, and M. Hylkema. 2013. Rethinking the Study of Landscape Management Practices among Hunter-Gatherers in North America. *American Antiquity* 78(2):285–301.

Martinez, Doreen E. 2012. Wrong Directions and New Maps of Voice, Representation, and Engagement: Theorizing Cultural Tourism, Indigenous Commodities, and the Intelligence of Participation. *American Indian Quarterly*, 36(4):545–573.

Menzies, C.R. 2006. *Traditional Ecological Knowledge and Natural Resource Management*. University of Nebraska Press, Lincoln.

McManamon, F.P. (editor). 2016 (in review). *40 Years of CRM (1974-2014): Accomplishments, Challenges, and Opportunities*. Routledge, New York.

Morville, Peter 2005 Ambient Findability: What We Find Changes Who We Become, First Edition. O'Reilly and Associates, Inc., Sebastapol, California.

NADB (National Archeological Database) Maps 1993. *NADB-Maps Archeological Site Counts (State Historic Preservation Officers).* Archeology Program, National Park Service, Washington, D.C. Original URL http://cast.uark.edu/other/nps/maplib/USsittot.1993.html archived by University of Arkansas Libraries with Archive-It.org on August 25, 2015 at https://wayback.archive-it.org/6471/20150825214608/http:/cast.uark.edu/other/nps/maplib/USsittot.1993.html.

National Monuments Record Thesauri. 1999. English Heritage. National Monuments Record Centre, Swindon, UK. Also available online at http://thesaurus.english-heritage.org.uk/

Neotoma. 2015. *Neotoma Paleoecology Database* [http://www.neotomadb.org].

OSTP (White House Office of Science and Technology Policy). 2013. Increasing Access to the Results of Federally Funded Scientific Research. Memorandum. Washington, DC. Electronic document at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Pampel, H. and S. Dallmeier-Tiessen. 2014. Open Research Data: From Vision to Practice. In *Opening Science*, edited by S. Bartling and S. Friesike, pp. 213-224. Springer Open, New York.

Pautasso, C. and E. Wilde. 2009. Why is the Web Loosely Coupled? A Multi-Faceted Metric for Service Design, In *Proceedings of the 18th International World Wide Web Conference (WWW2009)*, pp. 911-920. Madrid, Spain.

Rabinowitz, A. 2014. It's About Time: Historical Periodization and Linked Ancient World Data. *ISAW Papers* 7(22). Retrieved 1 June 2015 online at http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/rabinowitz/.

Raviele, M.E. Archaeology and an Interdisciplinary Digital Age. *Institute of Museum and Library Services*. Retrieved December 16, 2015 online at https://www.imls.gov/news-events/upnext-blog/2014/05/archaeology-and-interdisciplinary-digital-age.

Richards, J., S. Jeffrey, S. Waller, F. Ciravegna, S. Chapman, and Z. Zhang. 2011. The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing. In *Archaeology 2.0: New Approaches to Communication and Collaboration*, edited by E. Kansa et al., pp.31-56. Digital Archaeology Series 1, Cotsen Institute of Archaeology Press, University of California Los Angeles.

Rosenfeld, L., P. Morville, and J. Arango. 2015. *Information Architecture: For the Web and Beyond, 4th Edition.* O'Reilly and Associates, Inc., Sebastapol, California.

Shaw, R. A. Rabinowitz, P. Golden, and E. Kansa. 2015. A Sharing-Oriented Design Strategy for Networked Knowledge Organization Systems. *International Journal on Digital Libraries*. doi:10.1007/s00799-015-0164-0 Preprint available online at https://www.researchgate.net/publication/280529967.

Sheehan, B. 2015. RDAP14: Comparing Digital Archaeological Repositories: tDAR vs. Open Context. *Behavioral and Social Sciences Librarian* 34(4):173-213.

Simpson, L.B. 2014. Land as Pedagogy: Nishnaabeg Intelligence and Rebellious Transformation. *Decolonization: Indigeneity, Education & Society* 3(3):1-25.

Smith, Jolene. 2015. Archaeology for Everyone: A Virginia Digital Repository. *Institute on Digital Archaeology Method and Practice*. Retrieved December 16, 2015 online at http://digitalarchaeology.msu.edu/archaeology-for-everyone-a-virginia-digital-repository.

Watkins, J. 2000. *Indigenous Archaeology: American Indian Values and Scientific Practice.* Alta Mira Press, Walnut Creek, CA.

Wells, J.J. 2011. Four States of Mississippian Data: Best Practices at Work Integrating Information from Four SHPO Databases in a GIS-Structured Archaeological Atlas. Paper presented at the 76th Annual Meeting of the Society for American Archaeology, Sacramento. Electronic document, http://visiblepast.net/see/americas/four-states-of-mississippian-data-best-practices-at-work-integrating-information-from-four-shpo-databases-in-a-gis-structured-archaeological-atlas, accessed November 2015.

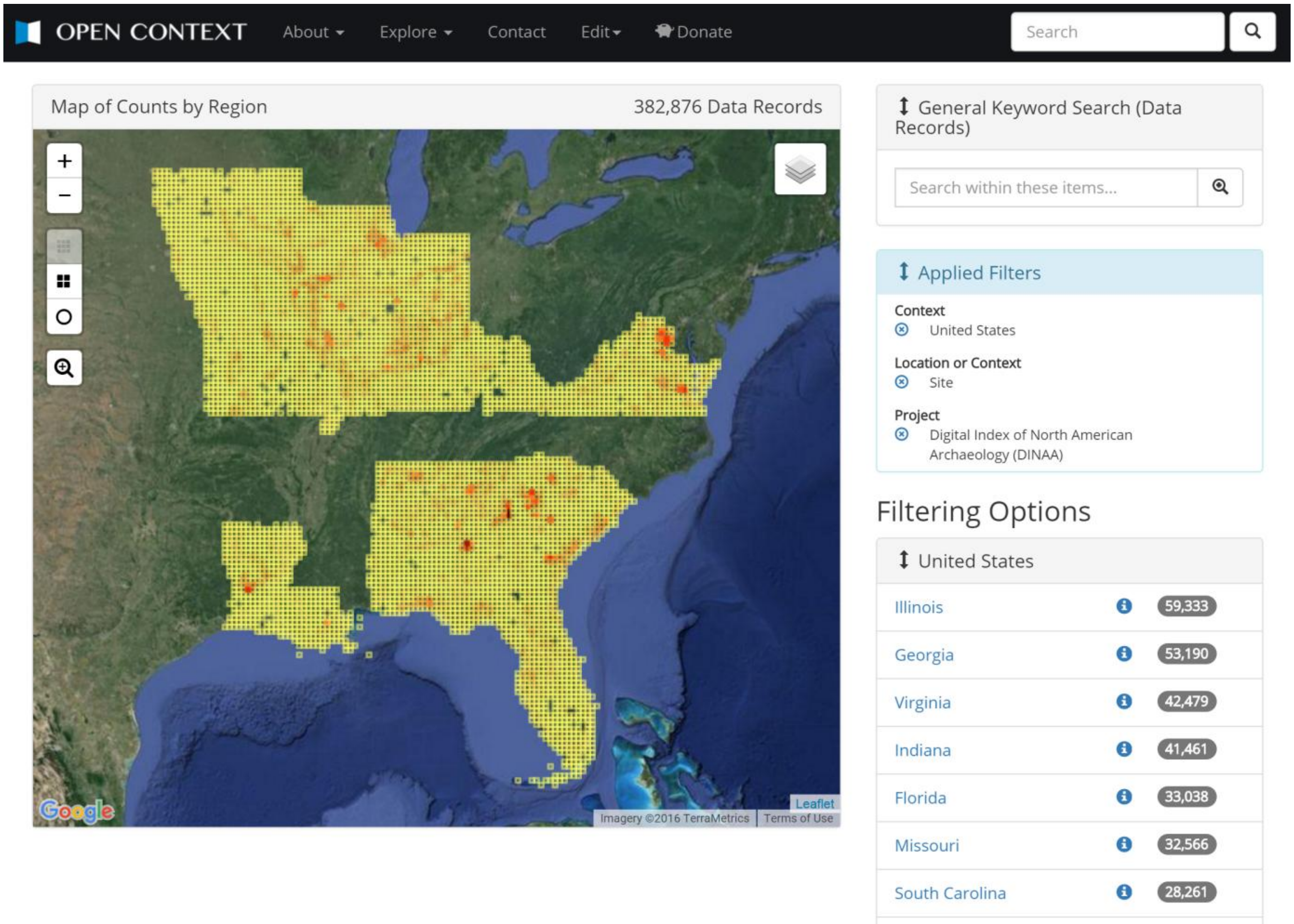
Wells, J.J., S.J. Yerka, and C.J. Parr. 2015. Archaeological Experiences with Free and Open Source Geographic Information Systems and Geospatial Freeware: Implementation and Usage Examples in the Compliance, Education, and Research Sectors. In *Open Source Archaeology – Ethics and Practice*, edited by Wilson and Edwards, pp. 130-146. DeGruyter Open Press, Warsaw, Poland.

Wells, J.J., E.C. Kansa, S.W. Kansa, S.J. Yerka, D.G. Anderson, T.G. Bissett, K.N. Myers, and R.C. DeMuth. 2014. Web-Based Discovery and Integration of Archaeological Historic Properties Inventory Data: The Digital Index of North American Archaeology (DINAA). *Literary and Linguistic Computing* 29(3):349–360. [http://dx.doi.org:10.1093/llc/fqu028]

Wiethaus, U (editor). 2007. *Foundations of First Peoples' Sovereignty: History, Education & Culture*. Peter Lang, New York.

Wildcat, M., M. McDonald, S. Irlbacher-Fox, and G. Coulthard. 2014. Learning from the Land: Indigenous Land Based Pedagogy and Decolonization. *Decolonization: Indigeneity, Education & Society* 3(3):I-XV.

Figure 1: Current Map Summarizing DINAA Coverage



Map of Counts by Region 382,876 Data Records

General Keyword Search (Data Records)

Applied Filters

Context
United States

Location or Context
Site

Project
Digital Index of North American Archaeology (DINAA)

Filtering Options

United States

Illinois	59,333
Georgia	53,190
Virginia	42,479
Indiana	41,461
Florida	33,038
Missouri	32,566
South Carolina	28,261

Figure 2: Map Summarizing Progress in Obtaining SHPO Dataset

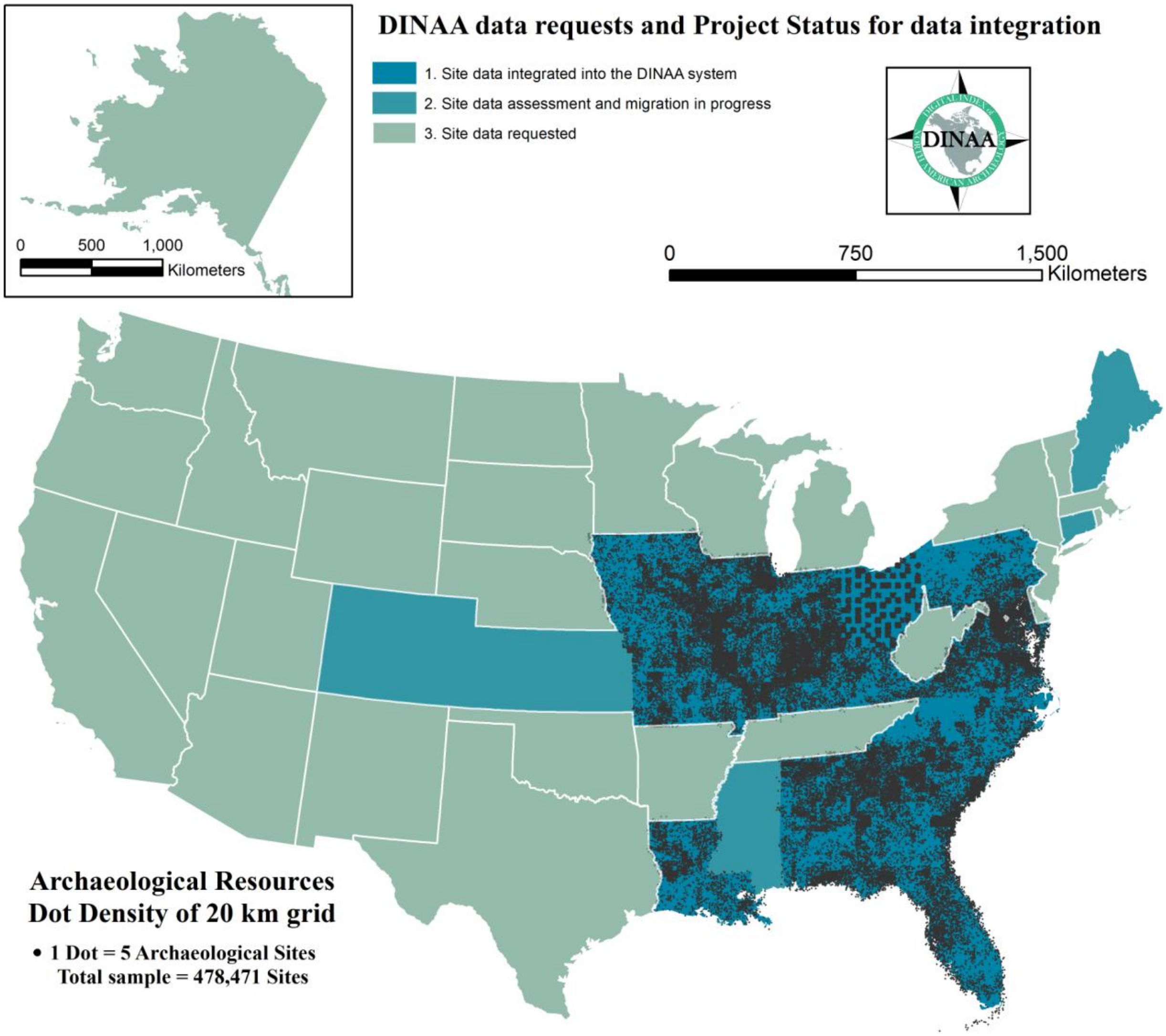

DINAA data requests and Project Status for data integration
1. Site data integrated into the DINAA system
2. Site data assessment and migration in progress
3. Site data requested
DINAA
DIGITAL INDEX OF NORTH AMERICAN ARCHAEOLOGY
0
500
1,000
Kilometers
0
750
1,500
Kilometers
Archaeological Resources
Dot Density of 20 km grid
• 1 Dot = 5 Archaeological Sites
Total sample = 478,471 Sites

Figure 3: Illustration of DINAA Cross-references with Related Content

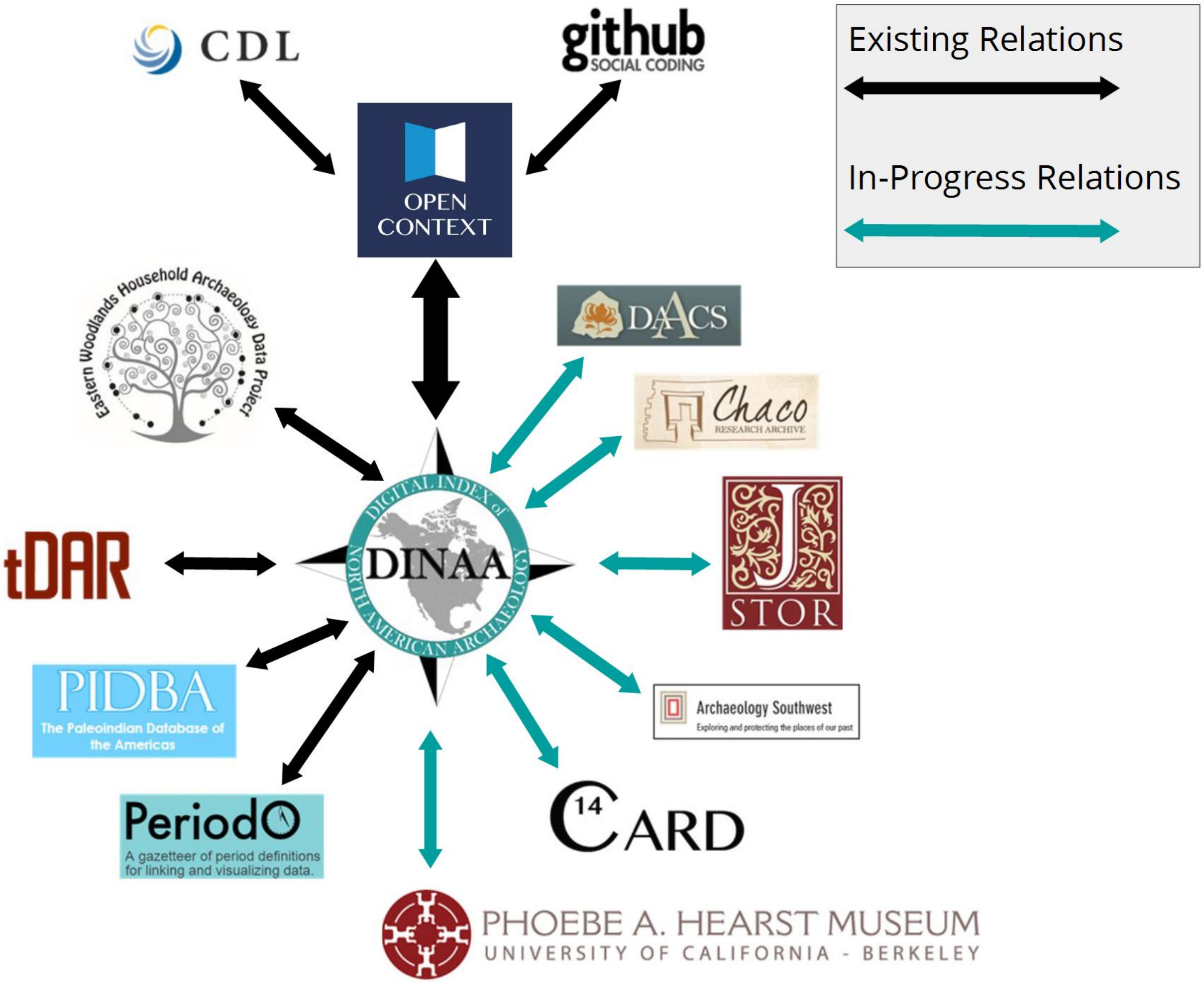
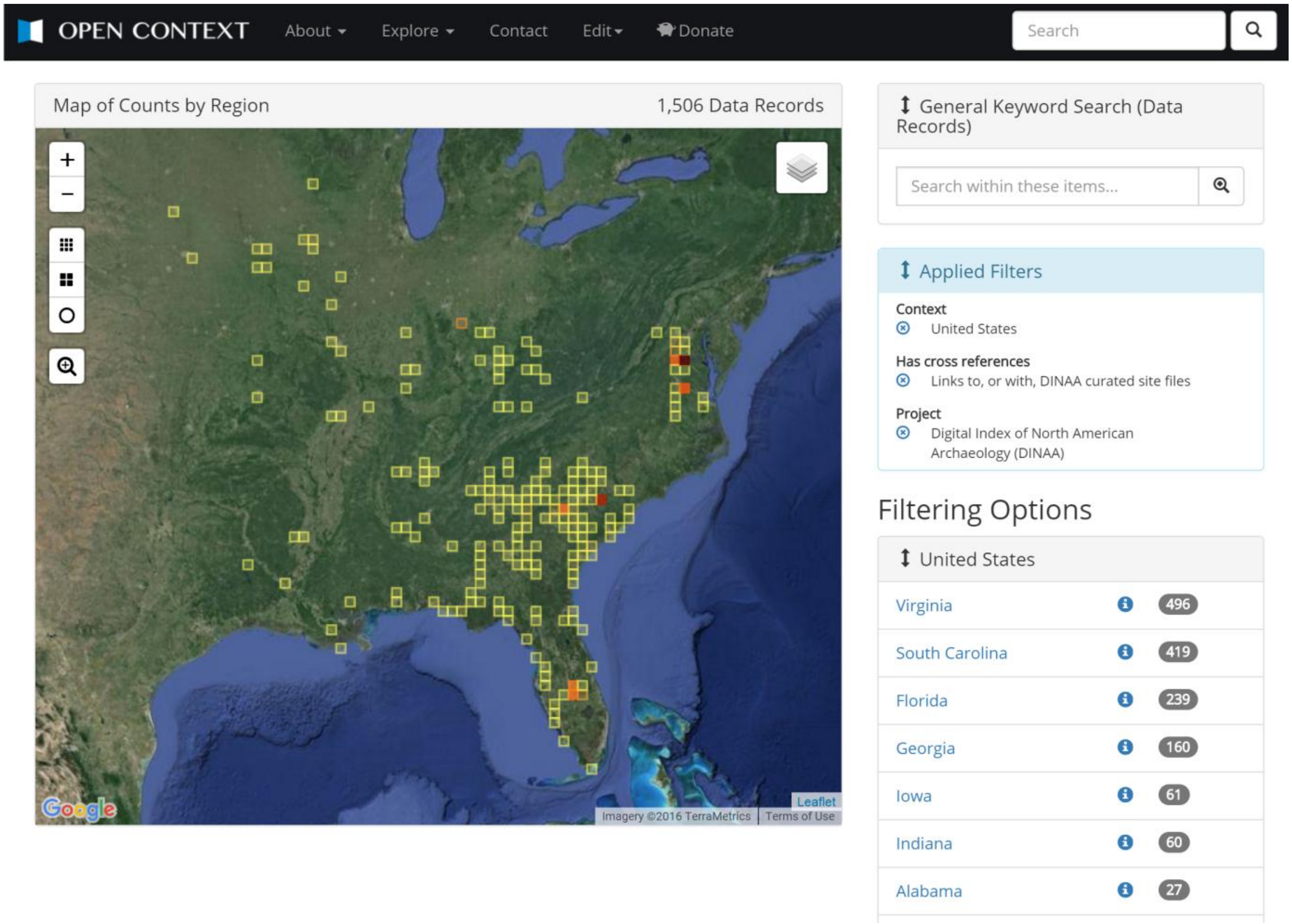

CDL
github SOCIAL CODING
Existing Relations
In-Progress Relations
OPEN CONTEXT
Eastern Woodlands Household Archaeology Data Project
DAACS
Chaco RESEARCH ARCHIVE
DIGITAL INDEX of NORTH AMERICAN ARCHAEOLOGY
DINAA
tDAR
JSTOR
PIDBA The Paleoindian Database of the Americas
Archaeology Southwest Exploring and protecting the places of our past
PeriodO A gazetteer of period definitions for linking and visualizing data.
C14 ARD
PHOEBE A. HEARST MUSEUM UNIVERSITY OF CALIFORNIA - BERKELEY

Figure 4: DINAA Map with Sites Cross-referenced by other Collections



Map of Counts by Region

1,506 Data Records

General Keyword Search (Data Records)

Search within these items...

Applied Filters

Context
⊗ United States

Has cross references
⊗ Links to, or with, DINAA curated site files

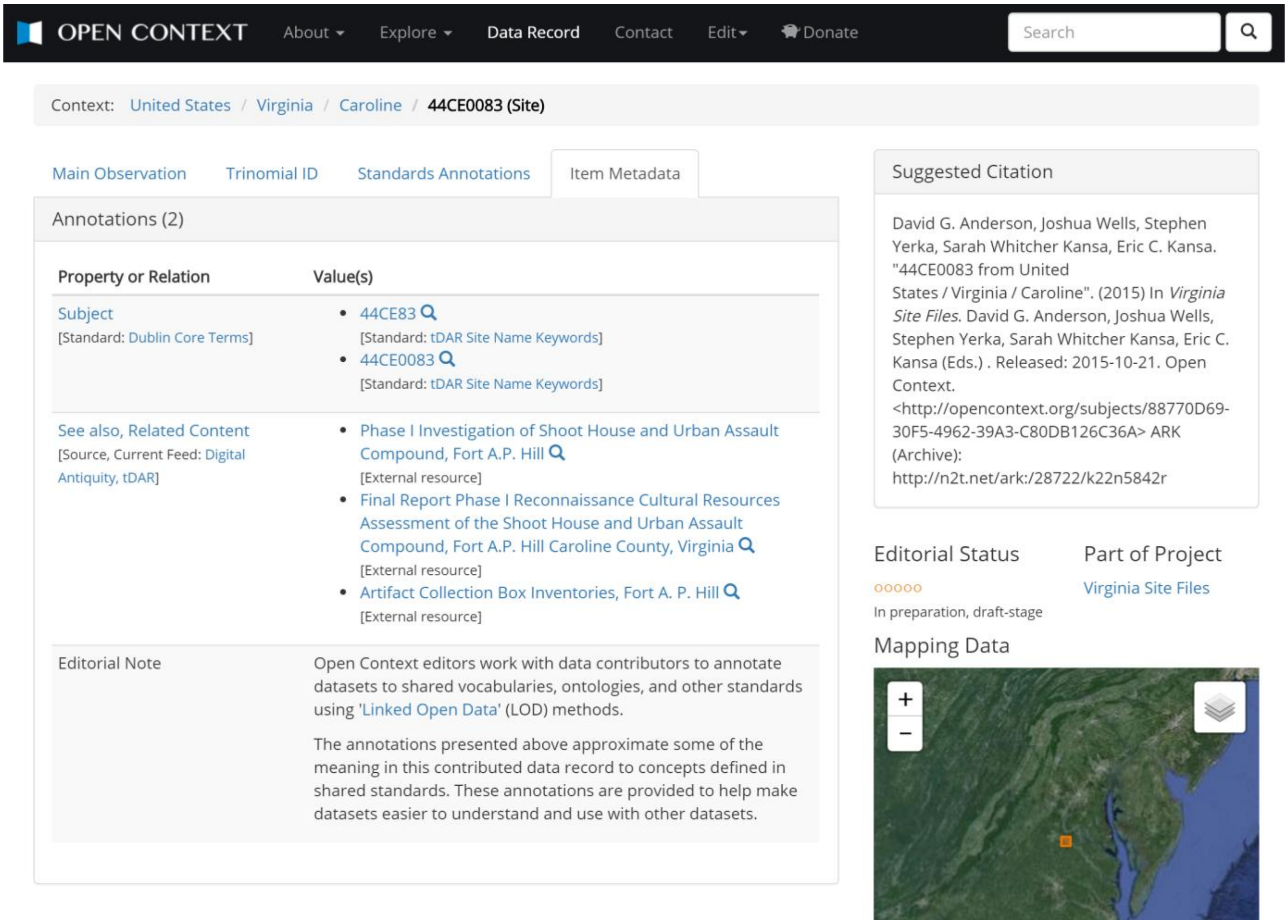
Project
⊗ Digital Index of North American Archaeology (DINAA)

Filtering Options

United States

Virginia	496
South Carolina	419
Florida	239
Georgia	160
Iowa	61
Indiana	60
Alabama	27

Figure 5: DINAA Record with Cross-reference to tDAR, Displaying tDAR Content via API Call



Context: United States / Virginia / Caroline / **44CE0083 (Site)**

Main Observation | Trinomial ID | Standards Annotations | Item Metadata

Annotations (2)

Property or Relation	**Value(s)**
Subject [Standard: Dublin Core Terms]	- 44CE83 [Standard: tDAR Site Name Keywords] - 44CE0083 [Standard: tDAR Site Name Keywords]
See also, Related Content [Source, Current Feed: Digital Antiquity, tDAR]	- Phase I Investigation of Shoot House and Urban Assault Compound, Fort A.P. Hill [External resource] - Final Report Phase I Reconnaissance Cultural Resources Assessment of the Shoot House and Urban Assault Compound, Fort A.P. Hill Caroline County, Virginia [External resource] - Artifact Collection Box Inventories, Fort A. P. Hill [External resource]
Editorial Note	Open Context editors work with data contributors to annotate datasets to shared vocabularies, ontologies, and other standards using 'Linked Open Data' (LOD) methods. The annotations presented above approximate some of the meaning in this contributed data record to concepts defined in shared standards. These annotations are provided to help make datasets easier to understand and use with other datasets.

Suggested Citation

David G. Anderson, Joshua Wells, Stephen Yerka, Sarah Whitcher Kansa, Eric C. Kansa. "44CE0083 from United States / Virginia / Caroline". (2015) In *Virginia Site Files*. David G. Anderson, Joshua Wells, Stephen Yerka, Sarah Whitcher Kansa, Eric C. Kansa (Eds.) . Released: 2015-10-21. Open Context. <http://opencontext.org/subjects/88770D69-30F5-4962-39A3-C80DB126C36A> ARK (Archive): http://n2t.net/ark:/28722/k22n5842r

Editorial Status

○○○○○

In preparation, draft-stage

Part of Project

Virginia Site Files

Mapping Data

Building a Gazetteer of Anthropocene North America (The Alexandria Archive Institute)

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

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.

(1) Structured data contributed by state agencies (SHPO offices) will be published by Open Context in a variety of human and machine readable formats under the Creative Commons Zero (CC0) public domain dedication.
(2) DINAA primarily focuses on disseminating datasets created by US states. Tribal nations (THPO offices) will be welcome to publish "open data" with Open Context if they see it in their interests, but the project has no such expectation for THPO collaborators (see response A.3).
(3) Leaning and instructional media created by the project will be made available under a Creative Commons Attribution (CC-By) license.
(4) Open Context is GNU-GPL licensed open source software. Code enhancements to Open Context will have the same GNU-GPL license. CollectionSpace is freely distributed open source under the ECLv2 license. Enhancements to CollectionSpace will be made available under the same license.

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 project will make all materials available either as public domain data, Creative Commons Attribution licensed media, or open sourced licensed software (see above). Open Context has no login barrier to access, and Open Context's only terms of use require users and their software agents to "rate limit" requests to Open Context so as not to overwhelm the service.

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.

Our team strongly advocates for open access, open data, open science and reproducible research while recognizing the need to safe-guard sensitive information and that different communities, with different histories and cultural norms, have different expectations for privacy and different needs to protect certain kinds of information.

Federal law and archaeological ethics require that we take strong measures to protect sensitive information, particularly specific site location data. The project will use Indiana University's (IU), SLASHTMP system (see letter from Wells) for secure and encrypted transfer of data files from SHPO offices to Register of Public Archaeology (RPA) credentialed DINAA team members. SLASHTMP is used by IU to meet federal data security needs when transferring sensitive medical and financial data. RPA certified team members will temporarily maintain encrypted copies of SHPO data in locked offices while they redact sensitive data. To guard against accidental release and malicious hacking, only redacted datasets will be transfered to Open Context for repository archiving and public dissemination. Files containing sensitive data will then be deleted. We will reduce the precision of site locations stored and made available to the public, assigning sites to an arbitrary ~20 KM grid as already agreed to by SHPO officials in several states.

With regard to indigenous intellectual property issues, Eric Kansa has published research papers (2005, 2009), participated in advocacy and policy development (with Creative Commons -iCommons, and the American Library Association) and collaborative research with the Intellectual Property issues in Cultural Heritage Project (http://www.sfu.ca/ipinch/project-components/working-groups/digital-information-systems-cultural-heritage-working-group). This background guides our collaborative orientation, shaping our efforts so that "open data" serves the interests of tribal communities, and does not merely represent a new form of cultural appropriation.

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.

The project will create supplemental instructional and learning media in the form of HTML5 web pages (~ 5 pages), that include images (JPEG, GIF, PNG, SVG) and perhaps some video (WebM, but also versions to be hosted by YouTube). These will be published either on Open Context itself, or its weblog, Heritage Bytes (http://ux.opencontext.org/), a Wordpress instance.

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 project team will create the content using standard office suites, image and video editors, and code editors.

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).

The project team will create the content using standard office suites, image and video editors, and code editors. Media content created by this project will help guide use of DINAA data, the primary focus. Images and videos need to meet needs of simple and speedy Web delivery and cost-effective creation; exact quality specifications will be determined by Phoebe France, the project Web designer.

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).

SHPO offices are typically understaffed and have poor technical support. Data managed by SHPO offices often have a variety of data quality problems. These include inconsistent use of controlled vocabularies, missing values, repeated identifiers, improperly typed data (string values in numeric fields), and other problems associated with manual data entry. Our work flow for improving data quality, used for all Open Context data publications is as follows:

(1) Redaction of sensitive data transfered by SHPOs by RPA-certified team members. This step also includes reduction of precision of geospatial data by assigning sites to an arbitrary ~20x20 km grid.
(2) Validation and clean-up using Open Refine. This step checks integrity and "uniqueness" of identifiers, spell checks and consolidates controlled vocabulary terms, checks date and numeric fields, etc.
(3) Using Open Refine, county names are added and cleaned. Fields for FIPS codes and Geonames URIs added and appropriate values are assigned to all counties (by using Open Refine to call the Geonames API).
(4) Schema mapping and import to Open Context. The process involves entity identification and description according to Open Context's own internal global schema (defined by this ontology: https://github.com/ekansa/oc-ontologies/blob/master/vocabularies/oc-general.owl). Open Context mints UUIDs (part of Open Context URIs) for each entity. Open Context occasionally catches errors where primary key identifiers in the contributed source dataset were repeated, resulting in message to check and resolve identifier integrity issues.
(5) Editorial review of imported data. Open Context site administers can log in and review imports, checking that all data and fields mapped property from the source schema to the Open Context schema and that all records have been imported. Indexing by Solr provides numeric counts of metadata facets, and Web mapping visualizations provide visual feedback. Facets and map visualizations may reveal errors in the data that require troubleshooting. Wells and Yerka serve on Open Context's editorial board and have familiarity and experience with these processes.
(6) Publication and indexing. Once the DINAA team (and relevant SHPO officials) are satisfied a dataset is adequately cleaned and ready for publication, Open Context allows public access to the data and faceted index. Public access provides a wider community the ability to use the data, and potentially discover errors that need correcting.
(7) Versioning. Open Context uses GitHub for public version control of published data (see: https://github.com/opencontext). Github's issue tracking features, Git version control, and other collaborative features provide the public with multiple means to note errors and provide corrections that can then be merged back into Open Context.

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).

Open Context is an open access data publishing venue for archaeology. Open Context facilitates long term data preservation with metadata documentation, publishing structured data in widely used nonproriety text-based open formats (JSON-LD, CSV) and only accepting media that adheres to the Digital Antiquity / Archaeology Data Service guide to best practice (http://guides.archaeologydataservice.ac.uk/).

While Open Context is not a long term digital repository, we accession data published by Open Context into the repositories of institutional partners. The University of California's California Digital Library runs the Merritt digital repository, a system that continually crawls Atom feeds generated by Open Context to discover new content to accession (see Open Context in Merritt: https://merritt.cdlib.org/m/ucb_open_context). Merritt then accessions multiple representations of data from Open Context, including HTML and "machine-readable" XML or JSON representations, and associated media files. The Atom feed also notifies Merritt of new versions of records, so Merritt can obtain updated data and build a version history of URI-identified resources in Open Context. Merritt mints persistent ARK identifiers for content it archives, and using Merritt's APIs, Open Context then associates these ARK identifiers to its own URIs. The ARK identifiers can then be used in persistent URIs for DINAA site file records. The California Digital Library is a "co-owner" of DOI and ARK identifiers created for Open Context, and can redirect resolution services to archived content should Open Context cease operations. Finally, in the spirit of "Lots of Copies Keeps Stuff Safe" (LOCKSS), we will also accession dumps of the DINAA dataset into the Indiana University ScholarWorks digital repository (see letter of commitment from Prof. Joshua Wells).

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).

Open Context publishes structured data as JSON-LD (an RDF serialization) as follows:

GeoJSON-LD: GeoJSON is an extensively-used open data format for expressing geospatial data. Major commercial and open source GIS and Web mapping systems support this standard. GeoJSON- LD builds upon standard GeoJSON (while maintaining full backward compatibility) to express more precise semantic information as Linked Data using the W3C endorsed JSON-LD standard. Using GeoJSON-LD has many advantages for DINAA. As a lightweight and simple open standard, it will ensure that DINAA data can be easily used with platforms popular on the mainstream Web and open source or commercial desktop GIS software, and enables us to express more formal semantics as Linked Data, thus offering the semantic precision needed for scientific computing applications.

Chronological Alignments and the DINAA Ontology: GeoJSON-LD will also be used to express annotations that situate site file data in chronological periodizations expressed using the PeriodO standard. Each US state has its own site file management system and these systems typically have idiosyncratic chronological periodization schemes. When published with DINAA, each site file record has its original state-specific metadata (redacted of sensitive information, especial that which may compromise site security). To promote interoperability, we supplement state-specific metadata with annotations to a more general PeriodO ontology (which itself is expressed in JSON-LD, Turtle and CSV formats, using Github version control). PeriodO models different calendar date ranges for different geographic regions, encoding dating systems from current published regional chronologies. Calendar date ranges provide additional chronological controls that are needed because a named time period (such as “Early Archaic”) may have some regional variation in time spans.

Ontologies and Digital Library Standards: Via Linked Data, we will use widely accepted vocabularies and ontologies including: (1) Dublin Core Terms (for basic metadata); (2) CIDOC-CRM + Archaeology extensions (an ISO standard ontology for cultural heritage data, used in this project to model archaeological sites and occupation events); and (3) various supplemental standards (RDF, RDFS, SKOS, and FOAF) to express more generic semantics. In addition to these geospatial data standards and domain specific semantic standards, Open Context also supports a variety of other digital library and Web standards that promote interoperability. These standards include FDGC geospatial standards (where applicable, given uncertain data collection and handling approaches prior to publication with Open Context), OAI-PMH (currently in development for metadata and Web services for metadata harvesting to also support the DataCite standard), and the Atom Syndication Format (a format for sharing updates of new or revised data). Open Context also references Library of Congress Subjects Heading URIs (subject metadata), GeoNames URIs (spatial metadata), ORCID URIs (creator, contributor or related person + organization metadata), and URIs provided by publishers and literature repositories (DOIs,, JSTOR URIs, WorldCat URIs).

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

In the case of Open Context, there is not much of a distinction between the data it publishes and metadata that it publishes. Both "data" and "metadata" are expressed in the same resources, with HTML, JSON-LD, and sometimes CSV representations. The data archiving and repository processes described in response B.2 equally apply to Open Context metadata.

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).

Open Context offers a variety of APIs and services and it interfaces with a number of different systems to promote discovery. Open Context "data publications" are now indexed by Google Scholar, facilitating discovery by a high-traffic service. In addition, we currently developing an OAI-PHM service that will also support DataCite metadata. Open Context also offers very flexible and powerful APIs that offer GeoJSON-LD (a widely supported geospatial format that can also be parsed as RDF). The APIs have extensive documentation (http://opencontext.org/about/services) and demonstrations. An open-source R-stats client for our API was developed by rOpenScience (see: https://github.com/ropensci/opencontext). Finally, Open Context provides a simple, old-school, paged Atom feed to facilitate crawling of all of its content (http://opencontext.org/manifest/.atom). Finally, Open Context content is not only in Open Context. Open Context uses GitHub for (short term) dissemination and public version control. GitHub has a huge user community and good search services that can enable discovery. In addition, because the CDL archives Open Context content, the CDL Merritt repository provides its own set of discovery services, APIs, and metadata harvesting services.

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).

Open Context makes all data available open access and in a variety of human and machine-readable formats. In the backed, Open Context is a Django-Python (3+) application that uses a Postgres data store. It runs on the Google Cloud infrastructure to enable easy replication and scaling. Open Context is also replicated and fully mirrored on the cloud computing infrastucture built by the German Archaeological Institute (a part of the German Foreign Ministry) at: http://opencontext.dainst.org/

Open Context uses Apache Solr for faceted search and provides geospatial visualization, querying and browse interfaces. Open Context can be used with standard browsers (without any plug-ins) and it scales well for smaller screens on mobile devices. Open Context's API can enable 3rd parties to develop mobile apps, alternative visualizations, or alternative information services.

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

http://opencontext.org
Examples:
(1) Excavations in Murlo, an Etruscan site: http://dx.doi.org/10.6078/M77P8W98
(2) Excavations at Kenan Tepe, a Bronze Age site in Turkey: http://dx.doi.org/10.6078/M7H41PBJ
(3) Excavations on Kodiak Island, Alaska: http://dx.doi.org/10.6078/M7VD6WC7

Also, a small National Endowment for the Humanities "startup" grant project: http://artiraq.org/maia/

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.

The project will make interface improvements to Open Context, so that search, mapping and visualization features are more straightforward, aesthetically pleasing, and usable for serving DINAA data.

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.

Open Context is an existing, functional and deployed software system for publishing structured data (and associated media) contributed by cultural heritage researchers and organizations. While Open Context is unique in having schema mapping tools to import diversely structured datasets, other software have similar capabilities for Web dissemination of structured archaeological and cultural heritage data. These include Ark (http://ark.lparchaeology.com/), which focuses on archaeological field recording and dissemination, and Arches (http://www.archesproject.org/) a heritage management database system. While Open Context is not a digital repository (it partners with the CDL for repository services), Open Context is referenced by the NSF and NEH for grant data management. Another system referenced by these agencies for archaeological data management is tDAR (https://www.tdar.org/), a digital repository run by Digital Antiquity.

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.

Django-Python (for Python 3+ environments) , javascript (especially Leaflet for Web mapping), and Bootstrap for responsive Web design. We use these because of their wide use and support, open licensing, and wide developer community (especially with regard to Python's active community developing open source libraries for linked data, geospatial data, and scientific computing).

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

The main point of software development for the IMLS request focuses on iterative development of user interface features for people using browsers. Open Context provides ample opportunities for interoperability with APIs, an (in progress) OAI-PMH service, and adoption of widely used standards expressed as JSON-LD for easy RDF serialization. Open Context also integrates outside APIs (ORCID, Arachne, tDAR, Encyclopedia of Life, Geonames, and others) to relate data it publishes with a wider context of information.

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

Open Context requires a Linux server with a Python 3 virtual environment and Postgres (MySQL should work but it is untested). Because Open Context uses Apache Solr for indexing, Java must also be installed (or installed on another server hosting the Solr service).

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

We use GitHub for software version control, issue tracking, and documentation. We have several documentation files that describe the API and provide deployment instructions and trouble-shooting. Git commit messages provide a detailed account of software development progress.

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

Open Context repository: https://github.com/ekansa/open-context-py

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.

To enable inspection and reuse, we release Open Context source code, under the GNU-GPL open source license at repository:
https://github.com/ekansa/open-context-py

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

GitHub, with documentation:
https://github.com/ekansa/open-context-py

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

GitHub

Name of publicly accessible source code repository: GitHub
URL: https://github.com/ekansa/open-context-py

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.

The data are created by state historical preservation offices (SHPOs) charged with implementing federal historical preservation laws that protect archaeological and historical sites. The data come from public archaeologists (government employees), contract archaeologists (commercial investigators working under contract to provide compliance services), and some academic archaeologists. These data are organized in tables, generally with relational database systems, GIS systems or even spreadsheets. SHPOs submit these data to Open Context for schema alignment, metadata documentation, and linked data annotation (to reference controlled vocabularies and ontologies).

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?

No IRB oversight is required. These are government administrative data about geographic features in the landscape.

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).

State agencies generally redact sensitive data before transfer to the DINAA team. However, sometimes we need to manage sensitive site-location data governed by Federal law and archaeological ethics. The project will use Indiana University's (IU), SLASHTMP system for secure and encrypted transfer of data files from SHPO offices to Register of Public Archaeology (RPA) credentialed DINAA team members. SLASHTMP is used by IU to meet federal data security needs when transferring sensitive medical and financial data. RPA certified team members will temporarily maintain encrypted copies of SHPO data in locked offices while they redact sensitive data. To guard against accidental release and malicious hacking, only redacted datasets will be transfered to Open Context for repository archiving and public dissemination. Files containing sensitive data will then be deleted. We will reduce the precision of site locations stored and made available to the public, assigning sites to an arbitrary ~20 KM grid as already agreed to by SHPO officials in several states.

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.

We are not collecting these kinds of documentation.

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).

Datasets collected by the project will be tabular dumps from SHPO agency databases. They will be in CSV or Excel file formats, and will usually contain up to 50,000 records for each state (and sometimes contain some related supplemental information summarizing finds, reports, and administrative actions). Given the relatively small scale of these datasets, standard office suite software, desktop GIS, and Open Refine provide the needed data processing capabilities.

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?

Data documentation quality, varies state-by-state. Some states have supplemental documents we can use for documentation, but in others, the state site file records are understood through the "tacit knowledge" of database users. In every case, we document data mainly with annotation to controlled vocabularies using Linked Open Data methods. For example, we link chronological period terms in individual datasets to a common periodization scheme (we will update to use the PeriodO scheme) using SKOS relations. We also consult archaeological literature and experts to provide additional documentation, when state site file administrators lack the time to dedicate to provide us with rich data documentation.

In general, every descriptive property (predicate) and controlled vocabulary concept originating from a source dataset has its own URI in Open Context and can take text, image, and other documentation. SKOS relations can link these dataset-specific descriptions and concepts to more widely used community curated controlled vocabularies like PeriodO, Geonames, etc.

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

Open Context will continue to serve the DINAA dataset as part of its general data dissemination services. The data will also be available through GitHub, as version controlled data "dumps", and thought the California Digital Library's Merritt repository.

8. Identify where you will be publicly depositing dataset(s):Open Context, the CDL Merritt repository

Name of repository: Merritt (California Digital Library, University of California)
URL: https://merritt.cdlib.org/m/ucb_open_context

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

Data dissemination and preservation are core goals of this project. We will use Merritt's API to verify accession of each record of DINAA data into the repository.

Original Preliminary Proposal

Building a Gazetteer of Anthropocene North America

We seek IMLS support to extend the Digital Index of North American Archaeology (DINAA) project. DINAA aggregates archaeological and historical data from state and tribal governmental authorities that manage United States cultural resources.[1] DINAA contributes to the national digital platform by providing the most comprehensive and detailed database documenting human settlement in North America currently available. Open Context (http://opencontext.org), an open access data publishing service for archaeology, hosts DINAA. Currently, researchers and the public can download over 340,000 site file records (with precise location and other sensitive data redacted) free of charge, and free of intellectual property restrictions.

DINAA has already successfully integrated archaeological site data from 15 states[2], encompassing the rich chronological, legal, and environmental metadata used by government officials and the research community alike. In this project we propose to continue this work to encompass the remainder of the United States, which based on our efforts with DINAA to date, is estimated to contain between two and three million archaeological sites. In doing so, DINAA will provide researchers, museums, libraries, government offices, and members of the public with a powerful Linked Open Data gazetteer of all known historical and archaeological sites in the United States.

Field-wide need addressed by the project

Linked Data increasingly plays a key role in library science, museum informatics and archaeological data curation and integration. Open Context, like other Linked Data systems, emphasizes the use of stable Web URIs to identify concepts and other entities so they can be easily and precisely referenced and related across different data collections on the Web. In archaeology and historical geography, the "site" is a key organizational entity. Minting stable Web URIs and offering rich temporal, geographic, and cultural metadata about sites will therefore create significant Linked Open Data resources essential for broadly integrating museum, library, and scientific datasets. This project builds capacity in the following ways:

1. *Open data, reproducible research:* Open Context, referenced by both NSF and NEH for archaeology grant data management, provides open access data publication services for archaeology and related fields and hosts the massive DINAA dataset. Open Context publishes with open licenses to advance the "open science" goals articulated in the 2013 White House Office of Science and Technology Policy memorandum requiring open access and open data for federally funded research. Our research on integrating editorial practices and public version control (with GitHub) won the "Best Paper" award at the 2014 Digital Curation Conference.

2. *Open services*: Open Context's services offer powerful and publicly-available RESTful (a simple, best practice for Web architecture) APIs (application program interfaces) to enable interoperability, extensibility, and alternate forms of visualization and user interfaces. This project will use Open Context's APIs for a variety of purposes aimed at maximizing interoperability and extensibility by using widely supported open standards and conventions.

3. *Integrating existing infrastructure*: This project also capitalizes on prior NSF, NEH, and IMLS investments by promoting data integration and cross referencing across scientific data repositories. Our current request to the IMLS will support project workshops and training sessions to promote wider use of computational methods needed meaningfully aggregate these powerful repositories. Building a wider community will help multiply the impact of this project.

4. *Outreach:* Our project will work in close collaboration with Tribal Historic Preservation Offices (THPOs) to exchange knowledge, develop technical capacity by THPOs to evaluate, use, and

[1] For more about DINAA, visit: http://ux.opencontext.org/blog/archaeology-site-data/

[2] See map of current DINAA data in Open Context: http://opencontext.dainst.org/sets/United+States?proj=52-digital-index-of-north-american-archaeology-dinaa&geodeep=11

participate in Linked Data, and ensure that Native American historical perspectives have greater representation in the Web of (cultural heritage) Data.

Projected performance goals and outcomes

The DINAA project is well poised to achieve significant positive impacts for libraries, museums and researchers with interests in North American prehistory and history. With IMLS support, we can further expand the coverage of DINAA and pilot the following key applications:

1. *Mapping publications*: Since the 1960s, many researchers have published scholarly papers and books identifying historical and archaeological sites with "Smithsonian Trinomials." Since DINAA curates Smithsonian Trinomial identifiers, with IMLS support we will text-mine literature in JSTOR and the Hathi Trust collections to find trinomials and associate these with DINAA records. This will power map-based search and browse interfaces to discover scholarly literature. We can also display "heat maps" showing where academic scholarship has focused, helping to illustrate the history of research, and gaps in scholarly attention.
2. *Cross-reference with other data sources*: By matching Smithsonian Trinomial identifiers, we have established links between DINAA site file records and metadata records in other datasets and repositories. These include the Paleoindian Database of the Americas, the Eastern Woodlands Household Archaeological Data Project, and tDAR (a major digital repository for North American archaeology, managed by Digital Antiquity). We have developed powerful entity reconciliation services to enable others to find DINAA URIs (and other metadata) for Smithsonian Trinomials. With IMLS support we will host workshops and develop online training materials to help others, especially libraries and museums, use DINAA as linked data to enhance their metadata and broaden the impact and reach of their collections. We will measure the success of our efforts to promote DINAA adoption by tracking the number of digital collections that cross-reference with DINAA.

Potential impact and longer-term goals

Our current IMLS request focuses on expanding DINAA's coverage and to promote community adoption. Beyond the period of IMLS support, we expect this work to have a long-term and significant impact on cross-disciplinary research. The term "anthropocene" helps capture an emerging scientific consensus about the key role human societies played in shaping the natural world. In North America, this started with initial settlement during the late Pleistocene, some 15,000 years ago. Comparing the trajectories of complex coupled systems of the deep past with the dynamics of today's global industrial civilization can improve the explanatory power of anthropocene studies. Developing a longitudinal science of the anthropocene represents a key strategic need in order for political and economic institutions to manage rapid changes in coupled social and ecological systems. DINAA can play a key role in integrating archaeological and paleoenvironmental datasets, opening powerful new research opportunities to explore dynamic relationships between human and natural systems in North America and globally.

Project director and partners

The project team has successfully collaborated on conceptualizing and developing the DINAA project for the past three years. The team includes Project Director Eric Kansa, Program Director for Open Context at the Alexandria Archive Institute; Kelsey Noack Myers, Tribal Archaeologist for the Chippewa Cree Tribe; Sarah Whitcher Kansa, Executive Editor for Open Context; Joshua Wells, University of Indiana, South Bend; and David Anderson and Stephen Yerka, University of Tennessee, Knoxville.

Budget

The total cost of this work over two years is estimated at $250,000 (salaries: $132,000; fringe benefits: $25,000; supplies: $1000; equipment: $3000; travel: $12,000; workshops: $25,000; consulting (user experience): $15,000, other costs (servers, hosting, archiving): $14,000; indirect costs: $23,000).