Developing the Cyberinfrastructure for a National Archaeological Site Database

Project Summary

This project will for the first time develop protocols to integrate archaeological site file data from large areas of North America into a unified database structure, and make these capabilities readily available to a wide variety of user communities. These files are held by State Historic Preservation Offices (SHPOs) and allied federal and tribal agencies, and contain unique site identification numbers, cultural affiliations, descriptions of condition and use, and locational and management information. Many files also contain a wealth of ancillary information about site integrity and quality, diagnostic artifacts, associated collections and reports, and investigative methods. The resulting National Archaeological Site Database (NASD) will transform basic archaeological research, data and heritage management, and public education in the United States and beyond. NASD will be transformative because it will:

- be integrative and authoritative on a national scale;
- facilitate and enhance resource management and protection far beyond local levels;
- make protocols and, where appropriate, primary data readily available through open source formats, platforms, and services;
- allow for interoperability among multiple disparate datasets;
- be sustainable, flexible, adaptable, and capable of growth in a number of directions; and
- create frameworks for future "Linked Data" applications in North American archaeology.

The resulting data and research products will provide explicit examples of best practices and procedures for the integration of archaeological datasets in the exploration of research and management issues at regional, national, and international scales. While our goal is to ultimately integrate data from all 49 states on the North American continental mainland and eventually states and territories beyond, as well as from other countries, access issues and the project scale preclude developing comprehensive coverage at present. The current project has as a goal the integration of site file data from between 15 and 20 states, located in Eastern North America.

Intellectual Merit: The integration of site file data at continental scales in a new and unique informational infrastructure will allow, for the first time, the exploration of the North American archaeological record across multiple temporal periods and geographic regions. Using best data translation practices and distributed, goal-oriented testing, the project will demonstrate how important research, management, and educational results can be obtained from previously unconnected or incompatible databases. Using broadly shared ontological schema, data formats, and web services, the project will vastly expand the research and resource management value of both new and existing data. By creating translating routines rather than dictating procedures, individual partners/data contributors will be able to maintain their own data structures and conventions, regularly upload new data, and use the combined dataset to further their own interests as well as develop and explore new ones.

Broader Impacts: NASD will stand as the definitive gazetteer of archaeological sites in the United States. The integration standards and web architecture will bring together a range of datasets, "boot-strapping" informatics powered research across disciplinary boundaries. The project will help to achieve NSF's goal of fostering novel networking and data integration among multiple partners, as well as research and educational activities across multiple disciplines and geographic boundaries. The results of the research will be widely published and posted online.

1217240

Developing the Cyberinfrastructure for a National Archaeological Site Database

Overview of Research Plan and Justification

The proposed research involves developing protocols for collecting and integrating archaeological data from across the United States, specifically to combine information from state site file repositories into a series of datasets available to all users via Open Context. The capabilities of the product, the National Archaeological Site Database, will be evaluated through analyses by scholars and managers in participating institutions and organizations, with the example and results serving as a basis for further growth and development. The project examines the collective capacity of governmental site databases to serve as an informational infrastructure resource in order to explore questions about the North American archaeological record across multiple temporal periods and geographic regions. The purpose of this project is unique and new in American archaeology: to use data translation practices and distributed, goal-oriented testing to create numerous models of database interoperability for research purposes. The project addresses head-on a major challenge facing archaeological informatics: how to build information management practices that enable scientifically useful applications of currently unconnected and incompatible database systems. This project will greatly improve the scientific research value of both new and legacy archaeological site files (and, in the future, associated collections and other datasets) by broadly employing shared ontological schema, data formats, and Web services.

The linkage of locational, collections, and resource management data at spatial and temporal scales far beyond current best practices will permit, for the first time, the inclusion of archaeological data in the exploration of new multidisciplinary research topics. These include how human populations at continental scales responded to changes in climate, biota, and the occurrence of other groups over the course of the late Pleistocene and Holocene; how and why the occupation and abandonment of regions occurred; and human impacts on the landscape over time. It will facilitate resource protection and management far beyond the installation, locality, or state levels currently dominating practice, leading to enhanced preservation of national archaeological resources. The ready availability of output online in the form of maps and tables will enhance public awareness, use, and appreciation for scientific research in general and archaeology in particular. Finally, the demonstration that primary archaeological data can be integrated and used to address fundamental questions at such scales will stimulate similar efforts worldwide, as well as encourage participation by states currently reluctant to join in the effort.

Results from Prior NSF Support: Project PIs D. Anderson, E. Kansa, S. Kansa, J. Wells and S. Yerka don’t have prior NSF support, and hence have no results from prior NSF support to report.

Background

Attempts to examine sites and artifacts from across large areas have a long history in archaeology, dating back to early attempts to recognize cultures on the basis of the distribution of distinctive artifacts like Bell Beakers in Western Europe or shell tempered pottery in Eastern North America (e.g., Childe 1926, 1929; Holmes 1903). Typically, artifact occurrences or other cultural traits were plotted on maps, and the distributions used to infer such things as past cultures, migrations, and interaction networks. Similar procedures remain in common use to the present day in archaeology, albeit with far more sophisticated mapping and analysis procedures, incorporating Global Positioning System (GPS) technology, relational databases, and geographic information systems (GIS)(e.g., Chapman 2006; Conolly and Lake 2006; Hochstetter et al. 2011; Kvamme 1989, 1990, 1999, 2006; Mehrer and Wescott 2005; Wescott and Brandon 1999;

Wheatley and Gillings 2002). These recent efforts, while far easier to produce and much more precise, and for which a vast and growing literature exists, most commonly make use of archaeological information from fairly small areas, such as within areas slated for development, highway corridors, reservoir floodpools, or on military bases, forests, wildlife refuges, or parks. Only rarely have archaeological data been examined at larger scales, especially those crosscutting state lines, such as within or between major biomes, river drainages, or physiographic provinces. Site file data have never been compiled at the scale and for the research and management purposes discussed herein.

Efforts to collect archaeological data have a long history in American archaeology, however, and are exemplified by the site files that are maintained in every state. In the late 1920s and early 1930s the National Research Council issued guidelines for establishing state archaeological surveys and site files, which were widely adopted (O'Brien and Lyman 2001). With the passage of the National Historic Preservation Act of 1966 (NHPA), maintenance of state archaeological site files became not merely an esoteric enterprise carried on by few museum and university professionals, but a responsibility of state government mandated by federal law and partially supported through the Historic Preservation Fund implementing the act (National Park Service 2010). As a result of the NHPA requirements and other environmental legislation, archaeological survey has been conducted at a high level for decades throughout the United States. In 1993, the last time primary site file data was compiled nationally under the auspices of 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 likely grown appreciably in the nearly two decades since that time, given the increase in numbers observed in the southeast and midwest over this period (Anderson and Sassaman 2012), with the number now likely somewhere between 1.5 and 2.0 million.

Site files include precise locational information as well as details on periods of use, site types, associated artifact collections and documents, preservation condition, National Register of Historic Places eligibility status (i.e., eligible, not eligible, or undetermined), and a host of other variables useful for research and management purposes. Though these data have the potential to tell us revolutionary things about the past occupation of the North American continent, as well as how we can manage this record of human achievement, they have rarely been integrated and used beyond the state level. Indeed, no mechanism exists today to integrate and make collective use of this vast record. As we shall show, it is now possible to move beyond this state of affairs. This proposal offers a means to remedy this situation that will, in the process, transform archaeology in this country and beyond.

General Approach and Justification

The proposed research involves developing web-based workflows, or software protocols, for collecting and integrating archaeological data from across the United States, specifically information from state site file repositories into a single dataset available to all users via Open Context. The resulting data and research products will provide explicit examples of best practices and procedures that can be used in the integration of archaeological datasets in the exploration of research and management issues at larger regional, national, and international scales. Our long range goal is to integrate data from all 49 states on the North American continental mainland and eventually states and territories beyond, as well as from other countries. The scale of such a

project warrants incremental development, resolving potential issues with smaller datasets before moving to progressively larger scales, and in any event is likely beyond what the NSF archaeology division could support. A few states, unfortunately, do not yet have complete site file data in electronic format, and the managers in some states are reluctant to make their data available (although all we have contacted are very willing to share their file structure). The current project thus has as a goal the development of protocols to integrate site file data from between 15 and 20 states, located in a contiguous area encompassing the southeastern, eastern, and midwestern United States.

We know we can succeed at this effort because we have already done it through two proof of concept analyses in which we partially integrated site file data from nine states, including five in the southeast (Figure 1; encompassing Alabama, Florida, Georgia, South Carolina, and Tennessee) and four in the lower midwest (Figure 2; encompassing Missouri, Illinois, Indiana, and Kentucky). For these analyses we used data on Paleoindian components in the southeast and Mississippian components in the Midwest; full integration of all the data from these state files was not the objective of these analyses (e.g., O'Donoughue 2007; Wells 2011). We have also linked the site file data from the five southeastern states in question with Paleoindian artifact collections data available online on the Paleoindian Database of the Americas (PIDBA) (e.g., Figure 3). Comparisons of site file and collections datasets have shown the importance of information examined at this scale for both research and management purposes (Anderson et al. 2010; O'Donoughue 2007; Miller and Smallwood 2009). In some cases, for example, components are present in existing collections yet are not recorded in the site files, which may reflect knowledge at the time the site was recorded, and not what has subsequently been recovered or reported from that location. Linking disparate datasets can thus help identify and remedy gaps in our knowledge, and in the process improve our understanding of the past.

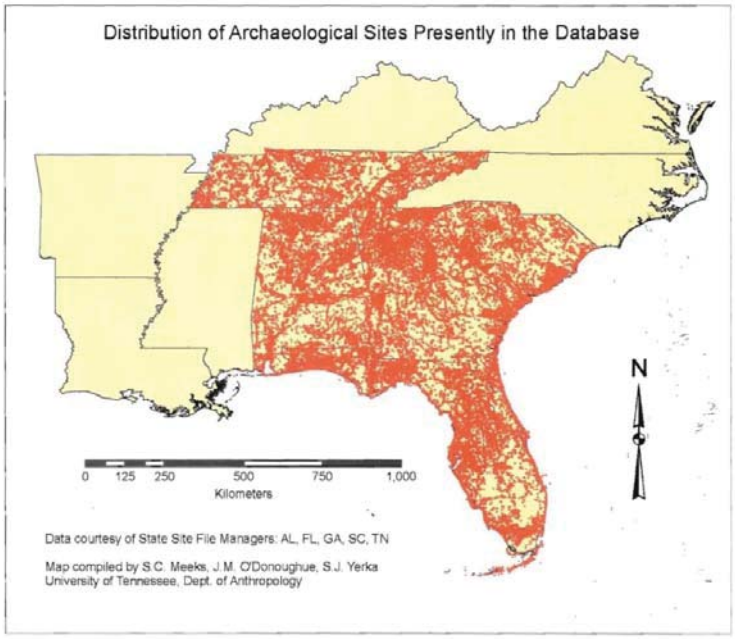

Figure 1. Distribution of all archaeological sites (n>150,000) in the state databases for Alabama, Florida, Georgia, South Carolina, and Tennessee as of 2007. Information provided by the site file managers in each state (after O'Donoughue 2007:127).

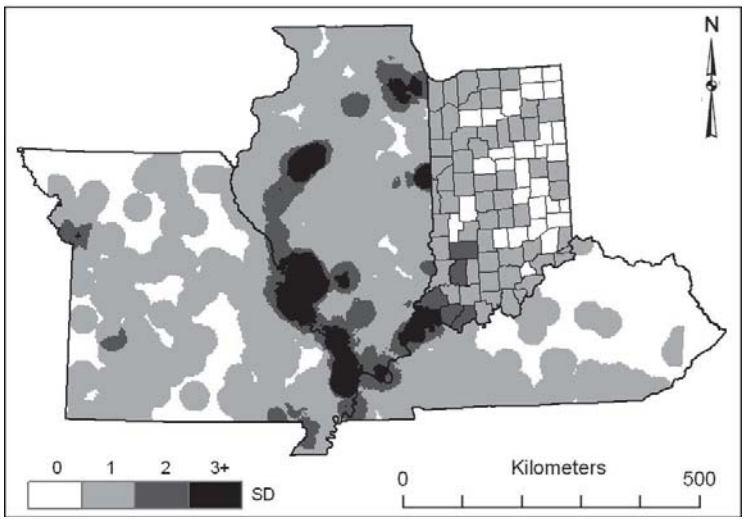

Figure 2. Density map of 4,591 Mississippian site locations defined by state archaeological site records. Continuous density for 3,922 locations across MO, IL, and KY with analytical radius of 2.5 km. County-level density presented for 671 locations in IN where site specific location data are incomplete. Standard deviation shading scale is the same for both sets. Information provided by the site file managers in each state (from Wells 2011).

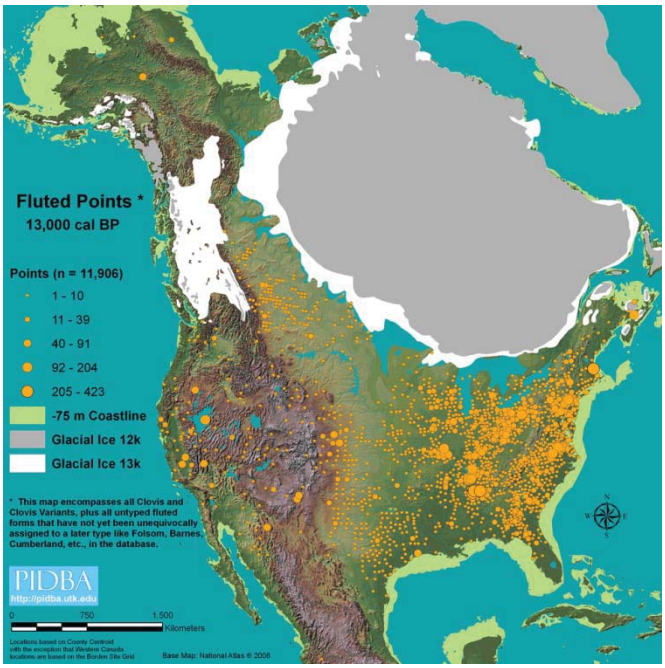

Figure 3. Distribution of Clovis culture projectile points, ca. 13,000 cal yr BP. The map shows the location of Clovis and Clovis Variants, plus points designated as ‘fluted’ in Paleoindian artifact recording projects but not yet assigned to a specific type in PIDBA, the Paleoindian Database of the Americas (from Anderson et al. 2010:71).

In an earlier proof of concept effort, in the mid-1990s one of the current PIs, David Anderson, brought together the site managers from across the southeast, resulting in a monograph length exploration of what could be done with southeastern site file data, and how to improve then-current practices (Anderson and Horak 1995). This effort led to the first maps of archaeological site distributions across the entire southeast during the Early, Middle, and Late Archaic periods, and the subsequent recognition that important information on settlement and land use was present in the combined and mapped data (Anderson 1996). Our project team thus has a long and well documented commitment to, and great familiarity with, the issues involved in integrating site file data at large geographic scales. Our recent feasibility studies have also made it clear that we have reached the point where readily available computer speeds, storage, and networking make site file integration and use at a large scale feasible. With it, the dream of earlier generations of archaeologists to map archaeological information at large geographic scales, and link it with other kinds of information, such as paleoenvironmental or physiographic dataset (i.e., glacial margins, shorelines, stream channel movements), is now within reach.

More specifically, we propose to aggregate archaeological databases from state site files, clean them, develop an ontological bridging mechanism that integrates aspects of their shared conceptual structures, and use this to merge the files into a larger, overarching National Archaeological Site Database (NASD). We will then use Open Context (Kansa and Bissell 2010; Kansa and Kansa 2011) to make the combined database and the linking metadata/ontology available online. Initial activity with NASD will include extended beta-testing to evaluate and debug the dataset and its presentation media. This will occur through a series of collaborations, including at least one face-to-face combined meeting with researchers and resource managers from those areas covered by the integrated data. In the process new and refined bridging mechanisms will be constructed and a host of new research questions investigated.

This project will proceed by taking disparate datasets and linking them together through translation programs that will permit new data to be uploaded routinely in whatever format the user employs. It is not a “top down” strategy, but instead entails a distributed networking approach, assuming that different users will want to maintain their current data management systems. It would, in fact, be naïve and unrealistic to expect otherwise, to impose a single data standard that would be broadly applicable for a continent full of archaeological sites collected by scores of different organizations for decades, and representing >13,500 years of differing cultures in varied and changing environmental conditions. Archaeological data can be highly heterogeneous, a situation further complicated by the legacies of individual federal, state and tribal government investments in separate database systems with unique constraints on data types and coding solutions. State site file managers are not going have the interest in nor, since many are overworked and understaffed, the time to restructure their datasets to meet standards that may not be appropriate for their needs. Most of the state systems currently in place, in fact, encompass tens of thousands of sites and have been in place for upwards of half a century. Rather than dictate protocols that will almost certainly be ignored, our project proceeds using methods we already know will work from our feasibility studies, by developing translation software to accommodate each user, that can be easily modified as necessary as data structures change. By showing how individual datasets can be integrated into a larger whole, the resulting structure will be perceived as something worthy of supporting.

Linked Open Data for North American Archaeology

This project will help develop and expand the technical infrastructure for American archaeology as advocated by many leaders in the field (Snow et al. 2006; Kintigh 2006; McManamon and Kintigh 2010). The key value of this project will be the minting and curation of Uniform Resource Identifiers (URIs) for specific North American archaeological sites, each of which is also uniquely identified with a site number. These URIs will be published with Open Context (see below) and will be critical infrastructure for future informatics projects, especially "Linked Data" applications. For example, publication of URIs for archaeological sites will make it feasible to unambiguously link description of archaeological collections (published anywhere on the Web) to places of origin. This will make it easier to analyze collections to discern regional and chronological patterning in material culture. Similarly, datasets describing zooarchaeological (archaeological animal remains) and paleoethnobotanical (archaeological plant remains) can be linked to places of origin, facilitating reconstructions of paleoenvironments and long-term human and ecological responses to climate change and other challenges. The same can be accomplished for reports providing information about specific sites.

Co-PI Eric Kansa's participation on the Google funded "Google Ancient Places" (GAP) project helps illustrate how a site-file gazetteer can be used in Linked Open Data programs (Barker et al. 2011). The GAP project used the openly licensed Pleiades Gazetteer and sophisticated text-mining and entity identification software to automate the identification of ancient places discussed in literature (books digitized by Google). The project results include novel information retrieval and book visualization and mapping tools, as well as quantitative data on the geographic coverage. Comparable data presentation efforts will be employed in the current project, including linkages with GIS datasets encompassing past and present topography and biotic resources, as has already demonstrated with PIDBA by Anderson and Yerka and collages at the University of Tennessee.

Technological Significance

The NASD will stand as the definitive gazetteer of archaeological sites in the United States and once the utility of the project is recognized and emulated, likely well beyond. The integration of standards and web architecture will become the cornerstone for innovative Linked Data / Semantic Web applications that can bring together artifact analyses, paleoecological data, geological data, collections information, and other datasets. Thus, this project will help build data management capacity in archaeology. Three primary systems will come together in this project, each with particular capabilities in digital dissemination and curation/preservation:

- **Open Context (http://opencontext.org):** Open Context will power open, password-free Application Programming Interfaces (APIs)/Webservices to enable third party access to machine-readable data in various representations (Atom, ArchaeoML-XML, KML, RDF, and JSON), using RESTful ("loosely coupled," see below) design patterns. Open Context also promotes interoperability by supporting the Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH), and Open Search protocols. To fully support Linked Data methods, this project will add a triple-store and SPARQL endpoint to Open Context to enable querying over RDF triples using the open source ARC2 library (for PHP/MySQL). These technical approaches will ensure that project data can be fully used

by other distributed systems supporting archaeological research, instruction, and public outreach.

- **PIDBA (http://pidba.utk.edu):** The Paleoindian Database of the Americas (PIDBA) serves as an example of the kind of dataset that can (and in this case will) be merged with site file data. PIDBA provides locational, attribute, and image data on Paleoindian materials >ca. 10,000 cal yr BP from all across the Americas. As of late 2011, PIDBA contains locational data on over 30,000 Paleoindian sites and artifacts from ca. 2000 discrete areas in North America. PIDBA grows through the voluntary user contribution of primary data, including artifactual, radiometric, and bibliographic information, much as we envision NASD will grow as its feasibility is demonstrated. The PIDBA dataset allows researchers to examine the distribution of artifacts at varying scales of resolution that are deliberately broad enough to ensure that site locations remain secure (Figure 3); similar course-grained scales will be used when distributing or presenting NASD information, to ensure sensitive or specific site locational or other data is not publically accessible. The PIDBA data have been used to document patterns of geographic dispersal, lithic raw material use, and demographic trends within the Paleoindian period (e.g., Anderson et al. 2010), and these and similar analyses can be used linking NASD data with other environmental datasets. PIDBA will be linked to URIs of sites published in Open Context as identifiers, enabling linking of PIDBA data with datasets published by other sources on the Web.
- **California Digital Library (CDL; http://www.cdlib.org):** The CDL will provide for data longevity and curation facilities. In addition, CDL offers a full range of digital library infrastructure to ensure that the datasets published in this project see indexing and citation tracking across scientific publications. These features help make published datasets “first class citizens” in researcher communications, and metrics can be used to track long-term impact and use of these data.

The overall design philosophy used to develop and maintain the NASD emphasizes “loose coupling” of systems. In loosely coupled systems, components share only a limited set of (usually simple) assumptions. This enables systems to evolve, because components that have very few dependencies on one another can evolve independently (Pautasso and Wilde 2009). The World Wide Web represents the best example of loosely coupled systems: very simple conventions can be used to hyperlink any page to any other. We propose to apply the principles and practices of loose coupling of dissemination (PIDBA, Open Context) and archiving (CDL) to encourage the growth of low-cost, evolvable, yet professionally archived and curated, archaeological cyberinfrastructure. This cyberinfrastructure will be easily extensible and open to new collections and archives that come online in the future. In addition, we will make the resulting data products available through other digital archaeological archives, such as tDAR (the Digital Archaeological Record) and the Federal Preservation Institute’s Preservation Portal to provide additional data longevity and specialized data integration tools. We have communicated our intent to managers of both archives (Dr. Francis P. McManamon, Dr. Richard Waldbauer), and they have expressed interest in cooperating with us in this effort. While the primary data will have much more precise locational characteristics, that level of resolution will not be made available publically at any time during this project, nor will specific data on site ownership, sacred status, or other sensitive information. Without explicit state and federal agency approval, in fact, such distribution would be inappropriate and potentially unlawful. We anticipate working

with the SHPOs (State Historic Preservation Officers) in each state, and with federal agency personnel, including the Federal Preservation Institute, to develop appropriate mechanisms for making data products available to the scholarly and resource management communities.

Publication of an authoritative gazetteer of North American sites requires research driven ontology development. The site ontology will reference MIDAS-XML, the UK based standard for describing sites (FISH 2005), used in conjunction with the National Monuments Record Thesauri (1999). The gazetteer will then be publicly viewable as KML data with two forms of representation, at the level of the US county and in a ~20km grid like the Borden grid employed for archaeological sites in Canada. These data (at the same scale, to maintain data security) will also be available for download as shapefiles, CSV data tables, and RDF and N3 triples for other Web and desktop applications, including desktop GIS investigation. The project will provide specific instructions for the open source GIS applications QGIS gvSIG, and uDIG, and the widely used proprietary application ArcGIS.

Referencing these standards will facilitate interoperability with other archaeological information systems. However, existing standards will need to be extended for suitability to North American contexts. For example, many of the terms in the English Heritage vocabulary are of little relevance to North American settings because they reflect different trajectories of historical and cultural development. In addition, the ontology developed for this project must take into account the limitations of data sources, some of which were developed with administrative, not research, purposes in mind. Any ontology developed for this project must accommodate the limitations of these source datasets. The project will develop its ontology, express it with RDF-SKOS, publish it with a Creative Commons Attribution license (a widely used standard for open copyright licensing), and archive it through the CDL.

Synergistic Collaboration between Open Context, PIDBA, and the CDL

The main technical goal of this project is to provide a foundation for distributed Linked Open Data initiatives in North American archaeology that builds on the model of securely shared, public data and facilitated research pioneered by databases like NADB and PIDBA. Open Context already participates in the Linked Open Data community by referencing URIs to entities in the Encyclopedia of Life, Pleiades Gazetteer, and GeoNames. It also publishes data with a high degree of granularity, minting a unique and stable URI for each location, object, media item, person, and other entities. This granularity, together with powerful APIs, enables efficient retrieval of precisely defined and unambiguously identified data. Furthermore, Open Context has a policy of only publishing data suitable for fully open and login-free access. These characteristics help make Open Context very well positioned for the Linked Open Data application proposed by this project.

However, some of the data managed by this project will be sensitive and will require restricted access for both ethical and legal reasons relating to site protection. To meet these needs, this project works synergistically to take advantage of the relative strengths of different parts of North America's emerging archaeological cyberinfrastructure, especially Open Context and tDAR. tDAR is a major, secure archaeological repository deployed by Digital Antiquity. While Open Context specializes in hosting structured data, tDAR hosts thousands of reports and other unstructured documents. Unlike Open Context, tDAR imposes a user login barrier to access data

and reports. A login barrier is necessary in some circumstances because of the sensitivity of certain forms of archaeological data. tDAR's greater emphasis on information security will permit future archiving of reports associated with sites. Even if precise site locations are redacted, these reports may contain other descriptive information that could enable site locations to be inferred. As such, some level of information security must be applied, as well as appropriate state and federal agency permissions obtained. This project takes advantage of the complementary strengths of Open Context and tDAR. Both systems are referenced by the NSF and NEH for archaeological data management. Open Context's login-free and highly granular data will facilitate access and use of site data (with locational information carefully redacted, see *Site Data Sensitivity and Security Measures* below) for museums, instruction, and many research applications built on the open Web. At the same time, the URIs minted for these sites can be used in relation to potentially more sensitive documents and data held in the more restricted-access tDAR repository (see *Data Management Plan* for additional details).

Data Quality and Consistency

Through funding from the Alfred P. Sloan Foundation, Open Context's editorial staff are currently developing editorial workflows to improve data quality and map contributed datasets to relevant shared taxonomies. Data contributors are often unwilling to invest much time and effort on improving the quality and consistency of their own data. They are also typically not conversant in relevant standards and lack the needed expertise to use standard vocabularies or ontologies. Open Context's editorial workflows aim to mitigate some of the social disincentives that currently inhibit data-sharing. Open Context uses Google Refine (2011), an open source tool to clean structured data. Google Refine logs each modification and edit as a machine-readable file in JSON format. This provides transparency to editorial choices and enables "roll-backs" to the former states of a dataset. Professionally cleaned and edited datasets can then be shared back to contributors. This may help motivate continued data sharing, since contributors will have access to cleaner, more usable versions of their own data.

Open Context's editors review and improve data quality and align datasets with appropriate domain standards while using Linked Data methods where feasible. These review processes not only enhance data quality and usability, but can help to elevate the prestige of data dissemination. For the latter, collaboration with the California Digital Library is critical. The CDL will assign permanent Digital Object Identifiers and Archival Resource Keys (DOIs and ARKs) to the dataset and maintain a "data paper" with full documentation of the dataset. The data paper will be citable, and citations can be tracked in academic literature through existing citation tracking services. Such services will further Linked Data applications also, since peer-review literature that references a project's data will be automatically linked back to the dataset (see *Data Management Plan*).

Information Retrieval and Services

Search strategies are an important determinant of the quality of meta-analyses in both the sciences (Stroup et al. 2000:2009) and humanities (Brown and Greengrass 2009). The "hit or miss" nature of Boolean keyword searches limits the effectiveness of information repositories for meeting scholarly needs, especially with regard to meta-analyses attempting to look across different datasets. Thus, in addition to keyword searching, this project will deploy faceted navigation systems for exploring and retrieving datasets and their metadata descriptions. Datasets

will be indexed using their native descriptive terminologies and the shared ontology developed by this project. The powerful open source application Apache-Solr (currently deployed by Open Context; http://lucene.apache.org/solr/) will be used to facilitate precision queries and retrieval of datasets using an innovative faceted search interface. In faceted navigation applications, users explore multiple dimensions of hierarchically-structured metadata with simple and intuitive "point and click" selections to progressively home in on specific information from a larger collection. Because filters are applied across an entire collection, users have greater certainty in the comprehensiveness of their results than with keyword searches. Feedback, in the form of subtotals for the numbers of items falling under each available facet, guides users in selecting additional filters. This feature offers users cues about the size and composition of the collection they are searching (Hearst 2006).

Open Context uses a highly abstracted and general global schema (ArchaeoML, see Schloen 2001). This permits Open Context to enable query and browse services on general and project/collection-specific metadata. Thus, Open Context will offer faceted search services for criteria defined by the general project ontology *and* the original descriptive terminologies defined by states.

Site Data Sensitivity and Security Measures

The security of archaeological sites must be protected for ethical as well as legal reasons. In the United States, the locations of archaeological sites represent highly sensitive data and their release could have grave repercussions. It is very 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 (i.e., permitting access to) 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, this project will only manage and store site location data at a very reduced level of geographic precision. The exact spatial resolution we will use for public data will be negotiated with SHPO and agency personnel; this resolution is expected to be at the county scale or at ca. 20 km resolution, which have been previously accepted in earlier efforts (e.g., Anderson and Horak 1995; NADB Maps 1993). This will still permit important research programs that examine regional and large-scale geographic patterning in archaeological data. It will also still permit innovative Linked Open Data applications, which mainly require URI identification of specific data resources (archaeological site records). The project will associate appropriate SHPO contact information with each data record to enable qualified researchers to directly obtain higher resolution spatial data from state officials. PIDBA has successfully implemented similar security strategies for 21 years and three project principals involved with data collection and selection are all RPA (Register of Professional Archaeologists) certified, with strong backgrounds in compliance and site protection.

Project Management and Timeline

The proposed project is planned to take 2 years (104 weeks) to complete. It will involve continuous interaction with local database originators, and rigorous, multi-institutional testing of the initial synthesized dataset in order to test usability and veracity of the primary product. Activities are to be broken down into four basic components, detailed below: 1. Organization (weeks 1-8), 2. Data Collection and Integration (weeks 9-52), 3. Data Analysis and Research Evaluation (weeks 53-80), 4. Public Demonstration of Best Practices for Cyberinfrastructure (weeks 81-104). All expenses and work will occur over a two year period, with final data products posted at the end of the second year.

Timeline

I. Organization	
Weeks 1-8	**Solicitation of Data Managers and Testers, Initial Assessment of Datasets**
II. Data Collection and Integration	
Weeks 9-52	**Get SHPO/THPO/NPS databases**
Weeks 12-52	**Clean databases and link to online Open Access database/GIS platforms**
Weeks 12-52	**Develop bridging structures for site archaeological and management data fields**
Week 53	**Post the integrated database**
III. Data Analysis/Research Evaluation	
Weeks 53-79	**Project collaborators explore database potential**
Week 80	**Face to Face Workshop of project participants**
IV. Public Presentation/Demonstration of Best Practices/Data as Cyberinfrastructure	
Weeks 81-104	**Summarize and synthesize lessons learned**
Week 104	**Post complete description of the project effort, protocols, and data products**

Principals of the Research Team

The senior investigators are all well-qualified to conduct the proposed research, and all have important roles in archaeological data management. They are:

David G. Anderson, PhD, RPA, University of Tennessee
Eric C. Kansa, PhD, School of Information, University of California, Berkeley
Sarah Whitcher Kansa, PhD, Alexandria Archive Institute, San Francisco, California
Joshua J. Wells, PhD, RPA, Indiana University, South Bend
Stephen J. Yerka, MA, RPA, Archaeological Research laboratory, University of Tennessee

E. and S. Kansa work with the Alexandria Archive Institute and they and Wells work with Open Context; Anderson and Yerka work with PIDBA (Project Director and Co-Director, respectively), and Anderson led an effort in the mid-1990s to integrate southeastern site file data, which included the publication of an edited monograph released through the National Park Service entitled “Archaeological Site File Management: A Southeastern Perspective" (Anderson and Horak 1995). Anderson and E. Kansa have both successfully facilitated the interoperation of

data on several large archaeological projects, and all four have significant experience with numerous state and federal site file systems through compliance and research activities. All five are published in American archaeology and have successful experience with large grant or contract funded activities with defined deliverables.

Phase 1. Organization (Weeks 1-8)
During the organizational period, project staff will initially solicit and assess the capacities of the datasets which will together form the NASD through detailed communication with their state or agency managers. Project principals will also solicit volunteer researchers to test prototype NASD datasets. All of these interactions will serve as bases from which to coordinate and grow networks of researchers who will ultimately create new source data for the NASD and share its usage products to grow the cyberinfrastructure of American archaeology.

Solicitation of Testers. Project principals will solicit professional involvement to test prototype NASD structures. These volunteers will meet RPA qualifications and be specialists in particular archaeological research areas that are specific to the scope of the constituent datasets that form the NASD. Up to three researchers will be identified to work with each of the project prototypes related to the Paleoindian, Archaic, Woodland, Late Prehistoric and Historic periods. Because of their particular research histories, project principals will also conduct their own tests of the Paleoindian and Late Prehistoric datasets. In addition, state and agency site file managers participating in the project will have access to the integrated datasets and will be urged to explore their utility, and provide feedback to the project principals.

Phase 2. Data Collection and Integration (weeks 9-52)
Primary File Structure/Data Collection. Collection of site file data structures and primary information will begin in earnest after week eight in order to produce a set of data that spans no less than 15 contiguous states. Databases will be gathered directly from SHPO, THPO, and federal agency managers and each individually analyzed to create metadata reports, also informed by discussions in the organization phase, detailing the structure and contents of each set. Exhaustive analysis will be conducted for each database to identify ontological structures of key archaeological terms and concepts (including colloquial and/or erroneous variants), forms of quantitative data representation, variation in expression of geolocational coordinates, and purposes and forms of ancillary data (site quality, previous investigations, ownership, etc.).

Data Security. At all times site locations and other sensitive information will only be available to authorized parties. No data will be publically posted or distributed.

Data Integration. An integrated database structure will be developed through the production of (1) an ontological bridge for qualitative concepts, and (2) a combined site file data table with cleaned and standardized quantitative data and original qualitative data. The ontological bridge will be a series of data tables that is primarily keyed on global representations for qualitative concepts as determined by project principals, and contains data fields for the original representations that relate to each global type. The combined site file data table will be primarily keyed with the unique Smithsonian trinomial identifier for each site; it will be populated with qualitative information from the ontological bridge that are pertinent to the salient aspects of an archaeological site, including culture history affiliations, site type, and assessed informational

quality of archaeological deposits; quantitative data such as geographic coordinates, spatial measures, and specific calendrical dates will be modified to standard formats; original qualitative data, to the extent they are represented in the ontological bridge, will be maintained as metadata for each site.

Tasks for creating the database structure will be divided among project principals and contracted employees. Initial data collection will be instigated by Anderson and Wells and conducted by Anderson, Wells, and Yerka. Assessment of scope and variation within original qualitative and quantitative data will be conducted by Yerka with hourly/graduate student employees at UTK, and by Wells with hourly graduate student employees in the IU system. Determination of global representations for the ontological bridge and quantitative data types will be conducted by Anderson and Wells. Construction of the ontological bridge and the combined site file data table will be conducted by Yerka with hourly employees at UTK, and Wells with hourly graduate student employees in the IU system. During this time E. Kansa will integrate the open source Arc2 RDF/SPARQL library on Open Context. Both Open Context and Arc2 use PHP/MySQL, facilitating integration.

Open source issue tracking software (http://issues.opencontext.org) will help this distributed team coordinate data cleanup, documentation, and referencing to the ontology. The issue tracker integrates with the Google Refine (deployed at http://refine.opencontext.org) API enabling easy communication of dataset edits and cleanup. At the end of week 52 the initial version of the NASD will be published online with Open Context, with all security measures described herein followed to ensure no precise locational or sensitive data is released.

Phase 3. Networked Data Analysis and Research Evaluation (weeks 53-80)
During this phase, analysis and evaluation will be achieved at four levels of networked community and data abstraction: (1) by collaborative researchers at a highly precise level of analysis with an expected error of less than 30m; and (2) by project principals at the same precise level of analysis. (3) by the general archaeological community and interested public at the county and ca. 20 km level of analysis, promoted and facilitated through public electronic forums; and (4) by means of a face-to-face meeting among project researchers and data contributors. Project principals, collaborative researchers, and representatives of originating data sources will also engage in periodic and ideally bimonthly teleconferences to discuss their efforts (note: methodological and research findings from these activities will be made publicly available by week 104). The four levels of analysis and integration are discussed in turn.

Collaborative Research Activity (1). For collaborative researchers (identified in weeks 1-8), tabular site description data and precise site location information will be made available to test questions potentially related to archaeological settlement, cultural resource management, or environmental change. Final reports about the capabilities of the database for research purposes will be required of all collaborators at the end of 12 months of activity, and will be presented at the workshop to take place toward the end of the second year of the project, as described above. Collaborators will also be asked to submit bi-monthly summary reports of activity to be published through OpenContext and available for public comment. Principals E. Kansa, S. Kansa, and Wells will facilitate this process.

Principal Research Activity (2). Principals Anderson, Wells, and Yerka will test the capacity for the Paleoindian and Mississippian portions of the integrated database structure to answer research questions about site distribution related to human behavior at the two endpoints of the North American prehistoric timeline. Project principals will be held to the same reporting standards as collaborating researchers. Comparable analyses will be undertaken with other portions of the dataset by participating researchers.

Public Activity (3). The integrated database structure will be published through Open Context by E. Kansa so that it is viewable in tabular, geospatial and other machine-readable formats (XML, RDF) with site locations abstracted to the county level (or similar spatial resolution, as approved by SHPOs). RDF outputs of the Pleiades Gazetteer will be used as a model for Open Context, thus facilitating semantic and technical interoperability with another key, related project. Documentation and metadata will be created by project principals (E. Kansa, S. Kansa, Wells, Yerka) so that the data can be easily accessed through a virtual globe utility such as Google Earth or GIS software that can be accessed by members of the professional community for research purposes. The broader use of the data for research (and also public education and higher education) will be facilitated by instructions for use of the data in open source GIS applications.

Face-to-Face Workshop (4). At a time deemed appropriate by the project principals, probably toward the end of Year 2 (ca. project week 80), a three day face-to-face meeting between NASD project team members and active data contributors, prototype testers researchers, and knowledgeable representatives of all cooperating entities that manage site files (sees supplementary materials) will take place in Knoxville at the University of Tennessee. The purpose of this gathering will be to evaluate the effectiveness of the integrated database for research and management purposes, and to propose ways in which the overall effort could be improved. Presentations of applications by users will be followed by discussions. The proceedings will be recorded with audio and video, and used to provide a basis for subsequent summary publications resulting from the effort. All participants will meet RPA qualifications for education and professional practice. This will be a physical meeting supplemented with internet videoconference inclusion of those unable to attend in person. These scales of networked research interaction will test the capacity for the American archaeological community and interested public to use these scientific data; these results will guide project principals as they create a stable, publicly accessible, research-quality database, supported by networked community collaboration and documentation.

Participants will be guided through discussions about the goals of the NASD project. All will be given ample opportunity to discuss their perceptions about the capacities of their datasets relative to the NASD and those of other states regarding data types, data structures, and metadata. Participants will be frequently encouraged to discuss their medium and long-term expectations for the NASD as a professional resource in full consideration of federal ARPA protections, along with state and municipal statutes. These records will augment the researched legal guidelines of project principals with examples of how legal and ethical codes are put into practice in myriad offices. In addition, participants will gain exposure to Linked Data and other Web data methods and approaches, so that they can better align work processes toward more efficient Web dissemination and scientific research outcomes. A summary of the proceedings will be created and hosted at Open Context's weblog (Heritage Bytes) within ten days of the event. This meeting

will serve NASD goals, but will also serve as a touchstone for future efforts to coordinate archaeological cyberinfrastructure at regional and national scales.

Phase 4: Public Demonstration of Best Practices for Cyberinfrastructure (weeks 81-104)
Summation and Synthesis. Lessons about the capabilities of the data and best practices gleaned from all four levels of project participation will be organized and refined in a set of public documents. The purpose of these documents will be (1) to describe the methods and results employed by all participants to investigate research questions, (2) to encourage continued research though exemplary instructions about use the integrated database structure and technical instructions to facilitate regular updates with new data, and (3) to serve as a base for continued growth in professional development of comprehensive archaeological data management standards throughout the cultural resource management sector. Further publication of work by project participants will be strongly encouraged in a variety of media, including peer-reviewed journals, and open access release of non-sensitive data. This summation will be posted through OpenContext, PIDBA, and other potential cooperating outlets such as tDAR and the Federal Preservation Institute's Preservation Portal.

At the end of week 104 the final version of the NASD and the public documents, data products, and protocols described above will be posted online, with all security measures described herein followed to ensure no precise locational or sensitive data is released.

Enabling Future Public Outreach Projects
The project will develop significant data resources in manner that greatly facilitates application and reuse by the wider community. These data and services can lay the foundation for compelling public outreach efforts (including mobile apps) that can enable the public to engage with the rich cultural and historical depth that lies all around them. During the course of this project we will promote various 3rd-party "spin-off" projects by consulting with the SAA's Public Education Committee and other bodies.

Conclusions
The integration of site file data at continental scales in a new and unique informational infrastructure will allow, for the first time, the exploration of the North American archaeological record across multiple temporal periods and geographic regions. The geographic scale and extent of data integration proposed is currently unprecedented in American archaeology yet, we believe we have demonstrated that it is readily achievable. Site file data integrated in a manner like that described herein will have the potential to continue to grow, and as the resulting datasets become more inclusive, transform the practice of our profession.

The NASD project protocols and technical products will be available online and shared with the public and the scientific community. The project team members are preparing technical papers describing the goals, procedures, and technical results of our research as described herein, and as the project proceeds, for presentation at professional conferences and submission to peer reviewed technical journals encompassing archaeology and informatics.

e. References Cited

Anderson, David G.
1990 A North American Paleoindian Projectile Point Database. *Current Research in the Pleistocene* 7:67–69.

1996 Modeling Regional Settlement in the Archaic Period Southeast. In *Archaeology of the Mid-Holocene Southeast*, edited by Kenneth E. Sassaman and David G. Anderson, pp. 157-176. University Presses of Florida, Gainesville.

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

Anderson, David G., and Kenneth E. Sassaman
2012 *Recent Advances in Southeastern Archaeology: From Colonization to Complexity*. Society for American Archaeology Press, Washington, D.C. In Press.

Anderson, David G., D. Shane Miller, Stephen J. Yerka, J. Christopher Gillam, Erik N. Johanson, Derek T. Anderson, Albert C. Goodyear, and Ashley M. Smallwood.
2010 PIDBA (Paleoindian Database of the Americas) 2010: Current Status and Findings. *Archaeology of Eastern North America* 38:63-90.

Barker, Elton, Kate Byrne, Leif Isaksen, and Eric 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

Brown, Stephen, and M. Greengrass
2009 Research Portals in the Arts and Humanities. *Literary and Linguistic Computing*. doi: 10.1093/llc/fqp032 Online ISSN 1477-4615 - Print ISSN 0268-1145

Chapman, Henry
2006 *Landscape Archaeology and GIS*. Tempus Press, Letchworth Garden City, UK

Childe, V. Gordon
1926 *The Dawn of European Civilization*. Kegan Paul, Trench, Trübner & Company, London, U.K.

1929 *The Danube in Prehistory*. Clarendon Press, Oxford, U.K.

Conolly, James, and Mark Lake
2006 *Geographical Information Systems in Archaeology*. Cambridge Manuals in Archaeology, Cambridge University Press, Cambridge, U.K.

FISH (Forum on Information Standards in Heritage)

2005 FISH Interoperability Toolkit. English Heritage and the Council for British Archaeology. Available online at http://www.heritage-standards.org.uk/

Google Refine
2011 A Power Tool for Working with Messy Data (formerly Freebase Gridworks). Available online at http://code.google.com/p/google-refine/

Hochstetter, Francisco Torres, Sergio Rapu Haoa, Carl P. Lipo, and Terry L. Hunt
2011 A Public Database of Archaeological Resources on Easter Island (Rapa Nui) Using Google Earth. *Latin American Antiquity* 22:385-397

Holmes, William Henry
1903 Aboriginal Pottery of the Eastern United States. *Twentieth Annual Report of the Bureau of American Ethnology to the Secretary of the Smithsonian Institution 1898-99*, edited by J.W. Powell, pp. 1-237. Government Printing Office, Washington, D.C.

Kansa, Eric C.
2010 Open Context in Context: Cyberinfrastructure and Distributed Approaches to Publish and Preserve Archaeological Data. *SAA Archaeological Record* 10(5):12-16.

Kansa, Eric C. and Ahrash Bissell
2010 Web Syndication Approaches for Sharing Primary Data in "Small Science" Domains. *Data Science Journal,* Vol 9 pp.42-53

Kansa, Eric C. and Sarah Whitcher Kansa
2011 Towards a Do-it-yourself Cyberinfrastructure: Open Data, Incentives, and Reducing Costs and Complexities of Data Sharing. In *Archaeology 2.0: New Approaches to Communication and Collaboration*, pp. 57-91, edited by E.C. Kansa, S.W. Kansa, and E. Watrall. Cotsen Institute of Archaeology Press: Los Angeles, CA.

Kansa, Eric C., Sarah Whitcher Kansa, Margie M. Burton, and Cindy Stankowski.
2010 Googling the Grey: Open Data, Web Services, and Semantics. *Archaeologies, Journal of the World Archaeological Congress,* http://www.springerlink.com/content/c259133070q276vg/

Kintigh, Keith, W.
2006 The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71(3):567-578.

Kvamme, Kenneth W.
1989 Geographic Information Systems in Regional Archaeological Research and Data Management. *Archaeological Method and Theory*, Volume 1. edited by Michael B. Schiffer, pp. 139-202. University of Arizona Press, Tucson.

1990 The Fundamental Principles and Practice of Predictive Archaeological Modeling. In *Mathematics and Information Science in Archaeology: A Flexible Framework*, edited by A. Voorrips, pp. 257-295. Studies in Modern Archaeology, Volume 3, Holos-Verlag Press, Bonn, Germany,

1999 Recent Directions and Developments in Geographical Information Systems. *Journal of Archaeological Research* 7(2):153-201.

2006 There and Back Again: Revisiting Archaeological Location Modeling, In *GIS and Archaeological Predictive Modeling*, edited by Mark W. Mehrer and Konnie L. Wescott, pp. 3-38. CRC-Taylor and Francis Press, Boca Raton, Florida.

McManamon, Francis P., and Keith Kintigh W.
2010 Digital Antiquity: Transforming Archaeological Data into Knowledge. *The SAA Archaeological Record* 10(2):37-40.

Mehrer, Mark W., and Konnie L. Wescott, editors
2005 *GIS and Archaeological Site Location Modeling*. CRC-Taylor and Francis Press, Boca Raton, Florida.

Miller, D. Shane and Ashley M. Smallwood
2009 Paleoindian Settlement in the American Southeast: A GIS-based Approach to Identify Potential Aggregation Loci. Paper presented at the Society for American Archaeology Annual meeting in Atlanta, Georgia, April 2009.

NADB-Maps, National Archaeological Database
1993 Archeological Site Densities. State Historic Preservation Officers, National Park Service. http://cast.uark.edu/other/nps/maplib/USsitdens.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/

National Park Service
2010 The Historic Preservation Fund Annual Report. National Park Service, Washington, D.C. Also available at http://www.nps.gov/history/hps/hpg/downloads/2010_HPF_Report.pdf

O'Brien, Michael J., and R. Lee Lyman
2001 *Setting the Agenda for American Archaeology: The National Research Council Archaeological Conferences of 1929, 1932, and 1935*. University of Alabama Press, Tuscaloosa.

O'Donoughue, Jason
2007 Paleoindian Settlement in the Southeastern United States: Insights from Examining Regional Databases. *Current Research in the Pleistocene* 24:126-129.

Pautasso and 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), Madrid, Spain, April 2009

Schloen, David
2001 Archaeological Data Models and Web Publication using XML. *Computers and the Humanities* 35:123–152

Snow, Dean R., Mark Gahegan, C. Lee Giles, Kenneth G. Hirth, George R. Milner, Prasenjit Mitra, and James Z. Wang
2006 Cybertools and Archaeology. *Science* 311(5763):958-959.

Stroup D. F., J. A. Berlin, S. C. Morton, I. Olkin, G. D. Williamson, D. Rennie, D. Moher, B. J. Becker, T. A. Sipe, and S. B. Thacker
2000-2009 Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) Group. JAMA 2000 Apr 19;283(15):2008-12. Available online at http://www.ncbi.nlm.nih.gov/pubmed/10789670 .

Waldbauer, Richard C., Constance Werner Ramirez, and Dan Buan
2007 Historic Preservation Learning Portal: A Performance Support Project for Cultural Resource Managers. *SAA Archaeological Record* 7(4):33-34.

Wells, Joshua 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/archives/947, accessed May 2011.

Wescott, Konnie L., and R. Joe Brandon, editors
1999 *Practical Applications of GIS for Archaeologists: A Predictive Modeling Toolkit.* CRC Press, Boca Raton, Florida.

Wheatley, David, and Mark Gillings, editors
2002 *Spatial Technology and Archaeology: The Archaeological Applications of GIS.* CRC Press, Boca Raton, Florida.

DAVID GEORGE ANDERSON, PH. D. R.P.A.

Professional Preparation

Case Western Reserve	BA	1972	Anthropology
University of Arkansas	MA	1979	Anthropology
University of Michigan	PhD	1990	Anthropology

Appointments

2009-Present 2007-Present	*Professor* *Associate Head*	Department of Anthropology, University of Tennessee, Knoxville
2004-2009	*Associate Professor*	Department of Anthropology, University of Tennessee, Knoxville
1996-2003	*Archaeologist*	Southeast Archeological Center, National Park Service
1990-1996	*Archaeologist*	Interagency Archaeological Services Division, National Park Service

Related Publications

1995 *Archaeological Site File Management: A Southeastern Perspective*. (David G. Anderson and Virginia Horak, editors). Interagency Archeological Services Division, National Park Service, Southeast Regional Office, Atlanta, Georgia. 140pp.

1996 Modeling Regional Settlement in the Archaic Period Southeast. In *Archaeology of the Mid-Holocene Southeast*, edited by Kenneth E. Sassaman and David G. Anderson, pp. 157-176. University Presses of Florida, Gainesville.

2003 *Archaeology, History, and Predictive Modeling: Research on Fort Polk 1972–2002*. (David G. Anderson and Steven D. Smith). University of Alabama Press, Tuscaloosa.

2007 A Continental-scale Perspective on the Peopling of the Americas: Modeling Geographic Distributions and Ecological Niches of Pleistocene Populations*. Current Research in the Pleistocene* 24:86-90. (J. Christopher Gillam, David G. Anderson, and A. Town Peterson).

2010 PIDBA (Paleoindian Database of the Americas) 2010: Current Status and Findings. (David G. Anderson, D. Shane Miller, Stephen J. Yerka, J. Christopher Gillam, Erik N. Johanson, Derek T. Anderson, Albert C. Goodyear, and Ashley M. Smallwood). *Archaeology of Eastern North America* 38:63-90.

Other Publications

1990 A North American Paleoindian Projectile Point Database. *Current Research in the Pleistocene* 7:67–69.

1996 *The Paleoindian and Early Archaic Southeast*. (David G. Anderson and Kenneth E. Sassaman, editors). University of Alabama Press. 526 pp.

2002 *The Woodland Southeast*. (David G. Anderson and Robert C. Mainfort, Jr., editors) University of Alabama Press, Tuscaloosa. 696 pp.

2007 *Climate Change and Cultural Dynamics: A Global Perspective on Mid-Holocene Transitions* (David G. Anderson, Kirk A. Maasch, and Daniel H. Sandweiss, editors). Academic Press/Elsevier, Amsterdam, The Netherlands. 604 pp.

Anderson, David G., and Kenneth E. Sassaman
2012 *Recent Advances in Southeastern Archaeology: From Colonization to Complexity*. Society for American Archaeology Press, Washington, D.C. In Press.

.

Synergistic Activities

(1) 1990-present: Director, Paleoindian Database of the Americas http://pidba.utk.edu/ This research group has developed new and innovative ways of exploring spatial distributions and prehistoric landscape for Paleoindian and Archaic Periods over the Americas using legacy archaeological survey data and environmental proxy data from multiple sources.

(2) 2009-present: Director, Cumberland River/Midsouth Paleoindian Project with D. Shane Miller (University of Arizona) and Tom Pertierra SEPAS, Inc) http://bellsbend.pidba.org/

(3) Examining possible migration pathways of early peoples in the Americas with J. Christopher Gillam, Stephen Yerka. Recent activity has included attending and presenting at a workshop on "Peopling of the Americas" organized by Murray Gell-Mann, Ilia Peiros, and Henry T. Wright at the Sante Fe Institute, Sante Fe, New Mexico, 24-26 September 2010. Also attended and presented at the Great Migrations II UNESCO Workshop, New York City, December 1-2, 2011.

(4) 1999-present: Shiloh Indian Mounds Excavation and Reporting Effort in collaboration with the Southeast Archeological Center, National Park Service, Tallahassee, Florida.

Collaborators and Other Affiliations

Collaborators and Co-Authors

Eric Kansa, School of Information, UC Berkeley
Sarah Kansa, Alexandria Archive Institute
Stephen J Yerka, University of Tennessee
Joshua Wells, University of Indiana South Bend
John C. Cornelison, Southeast Archeological Center, National Park Service
J. Christopher Gillam, SC Institute of Archaeology and Anthropology, University of South Carolina

Paleoindian Database of the Americas
Approximately 20 BA, MA, and PhD level colleagues, listed at http://pidba.utk.edu/contact.htm
D. Shane Miller, University of Arizona

ERIC C. KANSA, PH.D.

Professional Preparation

University of California, San Diego	BA	1990-1994	Anthropology
Harvard University	PhD	1995-2001	Anthropology

Appointments

2010-present	*Researcher*	University of California, Berkeley, *School of Information*
2010-present	*Technology Director*	Open Context
2008-2010	*Adjunct Associate Professor*	University of California, Berkeley, *School of Information*
2007-2008	*Executive Director*	University of California, Berkeley, *School of Information, Information and Service Design Program*
2003-2007	*Executive Director*	The Alexandria Archive Institute (nonprofit executive position, organizational co-founder)
2001-2003	*Lecturer on Anthropology*	Harvard University, *Dept. of Anthropology*

Related Publications

Kansa, Eric C.

2011 New Directions for the Digital Past. In *Archaeology 2.0: New Tools for Communication and Collaboration*, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Los Angeles: Cotsen Institute of Archaeology Press, pp. 1-25.

Kansa, Eric C. and Sarah Whitcher Kansa

2011 Towards a Do-it-yourself Cyberinfrastructure: Open Data, Incentives, and Reducing Costs and Complexities of Data Sharing. In *Archaeology 2.0: New Tools for Communication and Collaboration*, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Los Angeles: Cotsen Institute of Archaeology Press, pp. 57-91.

Kansa, Eric

2011 Open Government: Beyond Black-box Transparency. In *Government Information Management in the 21st Century, International Perspectives*, edited by Peggy Garvin. Surrey, UK: Ashgate Publishing, Ltd., pp. 157-170.

Kansa, Eric C., Sarah Whitcher Kansa, Margie M. Burton, and Cindy Stankowski.

2010 Googling the Grey: Open Data, Web Services, and Semantics. Archaeologies, *Journal of the World Archaeological Congress* 6(2):301-326.

Kansa, Eric C. and Ahrash Bissell

2010 Web Syndication Approaches for Sharing Primary Data in "Small Science" Domains. *Data Science Journal* 9:42-53.

Other Publications

Sorin Adam Matei, **Eric Kansa**, and Nicholas Rauh

2011 The Visible Past / Open Context Loosely Coupled Model for Digital Humanities Ubiquitous Collaboration and Publishing: Collaborating Across Print, Mobile, and Online Media. *Spaces and Flows: An International Journal of Urban and ExtraUrban Studies* 1(3): 33-48.

Kansa, Eric C

2009 “Opening Archaeology to Mash-ups: Field Data and an Incremental Approach to Semantics”, *Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content Conference Proceedings,* Metaxia Tsipopoulou (Editor). Hellenic Ministry of Culture, Athens, Greece.

Kansa, Eric C

2009 "Indigenous Heritage and the Digital Commons. In *Traditional Knowledge, Traditional Cultural Expressions and Intellectual Property Law in the Asia Pacific Region*. Christoph Antons, Editor. Alphen aan den Rijn: Kluwer Law International, pp. 219-244.

Kansa, Eric C and Erik Wilde
2008 "Tourism, Peer Production, and Location-Based Service Design". *IEEE International Conference on Services Computing*. Volume 2, Issue 1: 629-636

Kansa, Eric C, Jason Schultz, and Ahrash N. Bissell
2005 Protecting Traditional Knowledge and Expanding Access to Scientific Data*. International Journal of Cultural Property* 12(3): 285-314.

Synergistic Activities

(1) I co-founded the Alexandria Archive Institute, a nonprofit organization dedicated to enhancing research and education through online publication of primary social science data. In continuing this work, I lead development of *Open Context* (opencontext.org), an open access, web-based data sharing system for publishing and exploring diverse research content (archaeological and other field science datasets and related media). My oversight of *Open Context* includes data modeling, web architecture and service design, information retrieval, and user interface design and collaboration with the California Digital Library for data citation and data curation and preservation.

(2) I led a program at the UC Berkeley Information School to provide a focus for teaching and research on the skills and concepts required by a service-led and information-powered economy. The program developed a framework for the study of service by leveraging the School's distinctive configuration of competencies in information models, systems analysis and design methods, implementation of Web-based services and information-intensive applications, and Internet business architecture. I also managed and taught at the Information Systems Clinic, which gave students opportunities to work with faculty on research projects and on hands-on design and implementation activities.

(3) I am a Scientific Advisor to the Digital Antiquity cyberinfrastructure effort (at Arizona State University), as well as the Digital Atlas of the Holy Land initiative centered at UC San Diego. I was the founding convener (2006-2009) of the Digital Data Interest Group for the Society for American Archaeology.

(4) I am currently working to enhance social science computing infrastructure by deploying innovative textual analysis tools (place-entity identification) in a services ecosystem, with Project Bamboo (funded by the Mellon Foundation). In addition, I explore clashes in social norms and expectations regarding data privacy and intellectual property rights in the context of field research, particularly within researcher communities and between research communities and members of indigenous and tribal communities. I recently completed efforts to guide in depth transparency architectures for the 2009 *American Recovery and Reinvestment Act*.

Collaborators and Other Affiliations

Collaborators and Co-Editors

Erik Wilde, UC Berkeley
Raymond Yee, UC Berkeley
Sarah Kansa, Alexandria Archive Inst.
Stuart Campbell, U. Manchester, UK
Elizabeth Carter, UCLA
Tom Elliot, New York University
Patricia Cruse, California Digital Library
David G. Anderson, U. Tenn, Knoxville
Josh Wells, Indiana State South Bend
Ray Larson, UC Berkeley
Lew Lancaster, UC Berkeley
Thomas E. Levy, UC San Diego
Sorin Matei, Purdue University
Ethan Watrall, Michigan State
Jenna Burrell, UC Berkeley
Stephen Yerka, U. Tenn, Knoxville

Graduate Advisors and Postdoctoral Sponsors

CC Lamberg-Karlovsky, Harvard University
Lawrence Stager, Harvard University

Thesis Advisor and Postgraduate Scholar Sponsor (None)

SARAH WHITCHER KANSA, PH.D.

Professional Preparation

University of California, San Diego	Anthropological Archaeology	BA, 1993
University of Edinburgh, UK	Near East & East Mediterranean Archaeology	Ph.D. 2000

Appointments

The Alexandria Archive Institute	Executive Director	2007 – present
Domuztepe Excavations, Turkey	Project Zooarchaeologist	1996 – present
The Alexandria Archive Institute	Assistant Director	2001 – 2007
URS Corporation, Oakland, CA	Project Osteologist	2002 – 2003
Harvard University	Research Assistant	2000 – 2001

Publications- Related

Kansa, Sarah Whitcher and Francis Deblauwe
2011 User-Generated Content in Zooarchaeology: Exploring the "Middle Space" of Scholarly Communication. In Archaeology 2.0: New Tools for Communication and Collaboration, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Cotsen Institute of Archaeology Press: Los Angeles, CA. Online at: http://escholarship.org/uc/item/1r6137tb#page-201

Kansa, Eric C. and **Sarah Whitcher Kansa**
2011 Towards a Do-it-yourself Cyberinfrastructure: Open Data, Incentives, and Reducing Costs and Complexities of Data Sharing. In Archaeology 2.0: New Tools for Communication and Collaboration, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Cotsen Institute of Archaeology Press: Los Angeles, CA. Online at: http://escholarship.org/uc/item/1r6137tb#page-73

Kansa, Sarah Whitcher and Eric C. Kansa
2011 Beyond BoneCommons: Recent Developments in Zooarchaeological Data Sharing. The SAA Archaeological Record 11(1): 26-29.

Kansa, Sarah Whitcher and Eric C. Kansa
2009 Yes, it is all about you: User needs, archaeology and digital data. *CSA Newsletter*. Vol. XXII, No. 1 (April 2009). Bryn Mawr: Center for the Study of Architecture. http://csanet.org/newsletter/spring09/nls0902.html

Kansa, Sarah Whitcher, Eric C Kansa and Jason Schultz
2007 An Open Context for Near Eastern Archaeology. *Near Eastern Archaeology* 70(4): 187-193.

Publications- Other

Buckley, Mike, **S.Whitcher Kansa**, S. Howard and S. Campbell
2010 Distinguishing archaeological sheep from goat bones using a single collagen biomarker. *Journal of Archaeological Science* 37(1): 13-20.

Kansa, Eric C., **Sarah Whitcher Kansa**, Margie M. Burton, and Cindy Stankowski
2010 Googling the Grey: Open Data, Web Services, and Semantics. Archaeologies. Online at: http://www.springerlink.com/content/c259133070q276vg/

Kansa, Sarah Whitcher, Amanda Kennedy, Stuart Campbell and Elizabeth Carter
2009 Resource Exploitation at Late Neolithic Domuztepe: Faunal and Botanical Evidence. *Current Anthropology* 50(6): 897-914.

Kansa, Sarah Whitcher and Eric C Kansa
2007 Open Content in Open Context. *Educational Technology Magazine*. Vol. XLVII. (Nov-Dec 2007): 26-31.

Kansa, E., and **S. Whitcher Kansa**

2007 Open Context: Collaborative Data Publication to Bridge Field Research and Museum Collections, in *International Cultural Heritage Informatics Meeting (ICHIM07): Proceedings*, J. Trant and D. Bearman (eds). Toronto: Archives & Museum Informatics. 2007. Published September 30, 2007 at http://www.archimuse.com/ichim07/papers/kansa/kansa.html.

Synergistic Activities

(1) Development of methods for the online preservation and dissemination of archaeological content: I co-founded and currently lead a nonprofit organization dedicated to enhancing research and education through online sharing of primary social science data. In this capacity, I work with the archaeological and museum communities to encourage open access frameworks for publishing cultural heritage content. Specifically, I promote development and use of *Open Context* (www.opencontext.org), an open access, web-based data publishing system, whose development has been supported by the Hewlett Foundation, NEH, the Alfred P. Sloan Foundation, and others. I also mange a Digital Humanities Start Up Grant from the NEH for the project *The Gazetteer of the Ancient Near East* to expand temporal and spatial coverage of ancient places in the Pleiades system and improve research through the use of Linked Open Data.

(2) Documenting advances in the field of digital archaeology: I am editor on a volume entitled *Archaeology 2.0: New Tools for Communication and Collaboration* (co-edited with Eric C. Kansa and Ethan Watrall), published in autumn of 2011. This volume presents a variety of perspectives on the conceptual, theoretical and practical approaches to communicating archaeological knowledge through the use of new technologies and platforms. (Available online at: http://escholarship.org/uc/item/1r6137tb)

(3) Research on user experience around data sharing in archaeology: My organization recently concluded a project funded by the National Endowment for the Humanities and the Institute of Museum and Library Services (Advancing Knowledge: The IMLS/NEH Digital Partnership grant) entitled *Enhancing Humanities Research Productivity in a Collaborative Data Sharing Environment.* This 2-year project aimed to better understand the users and producers of digital cultural heritage content. Using case studies from archaeology, we produced a set of recommendations for individuals and organizations to help maximize the reusability and reach of their digital heritage content. This study included two intensive workshops, led with my colleague Eric Kansa at UC Berkeley's School of Information, as well as small focus groups, interviews and surveys.

(4) Service to the scientific community: I am a long-standing member of the International Council for Archaeozoology (ICAZ), where I stand on the Executive Committee and the International Committee (2010-2014). I am the Editor of BoneCommons (www.bonecommons.org), an online forum for the ICAZ community and beyond, which leverages digital technologies to improve communication and collaboration in zooarchaeology. I also serve on the Committee for Archaeological Research and Policy (class of 2013) for the American Schools of Oriental Research (ASOR), where I have worked to promote policies around the dissemination of research data. As a service to ASOR, my organization has sponsored an annual Open Archaeology Prize competition for best open-access, open-licensed, digital contributions to Near Eastern archaeology by ASOR members.

Collaborators and other Affiliations

Collaborators and Co-Editors

David Anderson, *University of Tennessee*
Michael Buckley, *University of York*
Margie Burton, *San Diego Archaeology Center*
Stuart Campbell, *University of Manchester, UK*
Elizabeth Carter, *UCLA*
Francis Deblauwe, *Alexandria Archive Institute*
Suellen Gauld, *Santa Monica Community College*
Sarah Howard, *University of Sheffield*
Eric Kansa, *UC Berkeley*
Amanda Kennedy, *University of Queensland*
Thomas E. Levy, *UC San Diego*
Jason Schultz, *UC Berkeley*
Cindy Stankowsky, *San Diego Archaeology Center*
Ethan Watrall, *University of Michigan*
Stephen Yerka, *University of Tennessee*

Graduate Advisors and Postdoctoral Sponsors

Nicola Murray, *University of Edinburgh*
Trevor Watkins, *University of Edinburgh*

Thesis Advisor and Postgraduate-Scholar Sponsor: Hannah Lau, *UCLA* (thesis advisor)

STEPHEN J. YERKA, M.A., R.P.A.

Professional Preparation

Middle Tennessee State University	BS	2002	Anthropology, Computer Information Systems
University of Tennessee, Knoxville	MA	2010	Anthropology
University of Tennessee, Knoxville	PhD	In progress	Anthropology

Appointments

2008-present	*IT Manager/ Geophysical Archaeologist*	Archaeological Research Laboratory, University of Tennessee, Knoxville
2006-2008	*IT Specialist*	Archaeological Research Laboratory, University of Tennessee, Knoxville
2005-2006	*Research Assistant*	Department of Anthropology, University of Tennessee, Knoxville
2002-2005	*Interpretive Ranger*	State Parks, Tennessee Department of Environment and Conservation

Related Publications

Stephen J. Yerka
2010 *Geophysical Study at Old Stone Fort State Archaeological Park, Manchester, Tennessee.* Master's Thesis, University of Tennessee, 2010.http://trace.tennessee.edu/utk_gradthes/845.

Anderson, David G., Stephen J. Yerka, and J. Christopher Gillam
2010 Employing High Resolution Bathymetric Data to Infer Possible Migration Routes of Pleistocene Populations. Current Research in the Pleistocene 27:60-64.

David G. Anderson, D. Shane Miller, Stephen J. Yerka, J. Christopher Gillam, Erik N. Johanson, Derek T. Anderson, Albert C. Goodyear, and Ashley M. Smallwood
2010 PIDBA (Paleoindian Database of the Americas) 2010: Current Status and Findings. *Archaeology of Eastern North America* 38:63-90.

J. Christopher Gillam, David G. Anderson, Stephen J. Yerka, and Shane Miller
2006 Estimating Pleistocene Shorelines and Land Elevations for North America. Current Research in the Pleistocene . 23: 185-187.

David G. Anderson, D. Shane Miller, Stephen J. Yerka, and Michael K. Faught
2005 Paleoindian Database of the Americas: 2005 Status Report. Current Research in the Pleistocene.Vol. 22.

Synergistic Activities

(1) 2007—present Co-Director, Paleoindian Database of the Americas http://pidba.utk.edu/. This research group has developed new and innovative ways of exploring spatial distributions and prehistoric landscape for Paleoindian and Archaic Periods over the Americas using legacy archaeological survey data and environmental proxy data from multiple sources.

(2) As an IT-minded contract archaeologist, I have over the last several years participated regularly in initiatives to facilitate data sharing and delivery for numerous federal agencies including NPS, DOD, TVA, as well as numerous state and government agencies.

(3) I currently serve as a special unit instructor for the National Forensic Academy and the University of Tennessee's Forensic Anthropology Center presenting on the application of technology and archaeological methodology for forensic investigation. The intent in my participation is to find innovative ways for law enforcement to apply technology for forensic searches and reconnaissance, including digital mapping techniques and geophysical survey.

(4) I have been an information systems consultant for several database design projects, most recently for the National Institute of Justice with Nicholas Herrmann and Bruce Ralston entitled Development of a Web Based Geographic Information System for the National Missing and Unidentified Persons System (NamUS), Award No. 2008-IJ-CX-K406. Additionally I am in consultation with the Forensic Anthropology Center developing arelational database system that organizes the information from Anthropology Forensic Collection, the Forensic Research Facility Donation Program, and interfaces with multiple faculty and student researchers.

Collaborators and Other Affiliations

Collaborators and Co-Authors

David G. Anderson, University of Tennessee
Daniel WH Brock, University of Tennessee
Howard Cyr, University of Tennessee
Mathew D. Gage, OAR, University of Alabama
Nicholas P. Herrmann, Cobb Inst., MS State
Joshua Wells, University of Indiana South Bend
Eric Kansa, School of Information, UC Berkeley
Sarah Kansa, Alexandria Archive Institute
Paleoindian Database of the Americas:
Approximately 20 BA, MA, and PhD level colleagues, http://pidba.utk.edu/contact.htm

Graduate Advisors

Gerald F. Schroedl, University of Tennessee
David G. Anderson, University of Tennessee
Boyce N Driskell, University of Tennessee
Kandace D Hollenbach, University of Tennessee

JOSHUA J. WELLS, PH.D., R.P.A.

Professional Preparation

University of Wyoming	BA	1997	Anthropology
Indiana University, Bloomington	PhD	2008	Anthropology

Appointments

2009-present	*Assistant Professor*	Indiana University South Bend, Department of Sociology & Anthropology, and Department of Informatics
2008-2009	*Visiting Assistant Professor*	Indiana University South Bend, Department of Sociology & Anthropology, and Department of Informatics
2006-2008	*Visiting Research Faculty*	Indiana University-Purdue University Fort Wayne, Archaeological Survey, and Department of Anthropology

Related Publications

Wells, Joshua J.

2009 A Space Syntax Analysis of Dohack and Range Phase Villages in the American Bottom. In *Tools of the Trade: Methods, Techniques and Innovative Approaches in Archaeology* (Proceedings of the 2005 Chacmool Archaeological Conference). J. Wilkins and K. Anderson, eds. University of Calgary Press.

Wells, Joshua J., and Robert G. McCullough

2009 Multiple Scales of Data on Falls Mississippian Settlement Practices. *Indiana Archaeology* 4(1):56-80

Wells, Joshua J., Craig R. Arnold, and Robert G. McCullough

2008 *Multiple Methods of Landscape and Site Specific Survey in an Archaeological Assessment of Clark County, Indiana*. IPFW Archaeological Survey Report of Investigations 802. 353 pp.

White, Andrew A. (with spatial analytical contributions by **Joshua J. Wells**)

2007 *The Late Pleistocene and Early Holocene Occupation of the Kankakee Region of Northwest Indiana*. IPFW Archaeological Survey Report of Investigations 702. 295 pp. http://new.ipfw.edu/centers/archaeology/reports

Other Publications

VanderVeen, James, and **Joshua J. Wells**

2011 Group Work Online. In *Quick Hits: Teaching with Technology*. R. Morgan, et al., eds. Indiana University Press.

Garver, Lydia, April Sievert, **Joshua J. Wells**, and Alicia Ebbitt

2008 Pedagogical Experiences of Anthropology Graduate Students. *Anthropology News* 49(6):33.

Wells, Joshua J., and Carlina de la Cova

2007 Stable Isotopic Relationships Between Age, Sex, and Maize Consumption in the Mississippian Vincennes Phase of Indiana. *Society for Archaeological Sciences Bulletin* 30(4):12-16.

Strezewski, Michael, Robert McCullough, Dorothea McCullough, Craig Arnold, and **Joshua J. Wells**

2007 *Report of the 2006 Archaeological Investigations at Kethtippecanunk (12-T-59), Tippecanoe County, Indiana*. IPFW Archaeological Survey Report of Investigations 703. 342 pp. http://new.ipfw.edu/centers/archaeology/reports

Synergistic Activities

(1) Since 2010, I have engaged in the compilation and analysis of a data set of 4,591 archaeological site records from four state SHPO databases (Kentucky, Illinois, Indiana, and Missouri), sites generally categorized with the keyword *Mississippian*, to develop a more comprehensive model of how these sites are defined and represented in historic preservation contexts, and to suggest ways to enhance the information value of such records to researchers. This received internal funding at my institution for 2010 ($8,477) and I have presented findings at three professional meetings in 2010 and 2011.

(2) I am the current, elected convener of the Digital Data Interest Group for the Society for American Archaeology (2010-present). I have led the organization of topical symposia and an invited panel discussion on issues related to data management and professional training in data management techniques. I have also represented the opinions of interest group members in a panel discussion on the professional effects of the NSF's requirement for data management plans.

(3) I am a member of the editorial board for *Open Context* (opencontext.org), an open access, web-based data sharing system for publishing and exploring diverse research content (archaeological and other field science datasets and related media). My duties include assistance in the development of editorial policies and ethical guidelines, identification of potential data sources for publication in Open Context, and recommendation of potential reviewers for items that require peer review.

(4) I am in my third semester of a two-year project to examine the effectiveness of technology-enabled active learning (TEAL) methods in an introductory archaeology classroom. This IRB-approved pedagogical project is based on concepts from the Massachusetts Institute of Technology TEAL initiative, and tests learning outcomes related to collaborative networked exercises, topical data acquisition, and ubiquitous computing practices that have been shown to benefit students in physics courses. This received internal funding at my institution ($2,993).

(5) I am planning my first year of activity in cooperation with the National Park Service and the University of Notre Dame to investigate the integrity and research potential of archaeological deposits at the Bailly Homestead, a National Historic Landmark at the Indiana Dunes National Lakeshore. Fieldwork will start in May 2012.

Collaborators and Other Affiliations

Collaborators and Co-Authors

Christopher Andres, Indiana–Purdue U. Fort Wayne
Cameron Griffith, Central Michigan University
Lawrence Kuznar, Indiana–Purdue U. Fort Wayne
Sorin Matei, Purdue University
Robert McCullough, Indiana–Purdue U. Fort Wayne
Daniel Osborne, University of Nebraska
Mark Schurr, University of Notre Dame
April Sievert, Indiana University, Bloomington
Jay Sturtevant, National Park Service MWAC
James VanderVeen, Indiana University South Bend
Ethan Watrall, Michigan State University

Graduate Advisors and Postdoctoral Sponsor

Della Collins Cook, Indiana University, Bloomington
Christopher S. Peebles, Indiana University, Bloomington
K. Anne Pyburn, Indiana University, Bloomington

Thesis Advisor and Postgraduate Scholar Sponsor (None)

h. Data Management Plan

This project will be an exemplar of archaeological data management and preservation. In keeping with the archival principle of LOCKSS ("Lots of Copies Keeps Stuff Safe"), project data will be accessioned by two digital repositories, one providing broad digital curation support for a host of disciplines (the University of California's California Digital Library via Open Context), and one dedicated specifically to archaeological data preservation, namely tDAR. tDAR is a digital repository using Fedora, and Digital Antiquity is currently in discussion with the Arizona State University system for additional digital library support for tDAR. Collaboration between Open Context and tDAR helps maximize the longevity of this project's data through archiving in multiple digital repositories.

Open Context has full digital archival support provided by the University of California's California Digital Library (CDL). The CDL's preservation micro-services continually crawl Open Context's Atom feeds to accession new data into the CDL's institutional digital repository. The micro-services archive Open Context data in a variety of open and non-proprietary formats, including XHTML+RDFa, ArchaeoML (XML), RDF, JSON, and CSV. The preservation micro-services mint special archival identifiers ("Archival Resources Keys" or ARKs) and assign these to Open Context's URIs.

Open Context (backed by CDL) supports versioning of individual data records and aggregate datasets. Open Context will maintain information and query retrieval services for current data, and reference archival copies of prior versions managed by the CDL. This version control will facilitate continued accessibility of current and comprehensive data while at the same time enabling researchers to go back to prior versions as needed.

By binding archival identifiers (maintained by CDL) to Open Context's URIs, we help ensure the permanence of identifiers created in this project. In addition, each dataset will be assigned a DOI and be published and archived as a "data paper" (analogous to a scholarly paper) in the CDL. This data page will have rich digital library metadata (METS, DataCite) for citation tracking and enhanced interoperability. Since citation plays a fundamental role in research and scholarly communications, adopting widely used data citations standards will help maximize the research value of these data.

Comprehensive release of non-sensitive data under the CC-Zero public domain dedication will ensure its usability in all future applications and cyberinfrastructure. Use of open data formats and open licensing will help ensure that the data outcomes of this project will be freely and openly available to all. This openness will help ensure that the project outcomes will encourage and fuel further development of archaeological cyberinfrastructure.