Developing the Cyberinfrastructure for a National Archaeological Site Database

Annual Report 2012-2013

Award Number: 1217240
Project Title: Collaborative Research: Developing the Cyberinfrastructure for a National Archaeological Site Database
Report Type: Annual Project Report
PI: David Anderson, Eric C. Kansa, Sarah Whitcher Kansa, Stephen J. Yerka
In collaboration with Joshua J. Wells, Indiana University, South Bend

Awardee: University of Tennessee Knoxville
Program Officer Name: John E. Yellen

Reporting Period: 9/1/2012 – 8/31/2013

Accomplishments

Project goals

The Digital Index of North American Archaeology (DINAA) project addresses head-on a major challenge facing research communities worldwide: how to link disconnected and incompatible data systems in such a way that the combined data are useful for important scientific research. Efforts to collect and compile archaeological data have a long history, and information about archaeological sites and collections is maintained by every state and territory. Only rarely, however, have these data been compiled and examined at large geographic scales, especially those crosscutting state lines, and never to the extent and for the research and management purposes proposed in this project. The DINAA project works with the databases held by State Historic Preservation offices and allied federal and tribal agencies in Eastern North America, with the goal of developing protocols for their linkage across state lines for research and management purposes. The primary goal is to integrate and link data from 15- 20 states (more than a half million sites) using an informatics framework that promotes extensions and reuse by government personnel in state and federal agencies, and domestic and international researchers. 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 availability of output online in the form of maps and data tables (at significantly reduced spatial resolution, to protect sensitive locations) will enhance public awareness, education, and appreciation for scientific research in general and archaeology in particular. The demonstration

that primary archaeological data can be integrated and used to address fundamental questions at such scales will stimulate similar efforts worldwide. Finally, by creating translating routines rather than dictating procedures, this project will foster archaeological cooperation through cyberinfrastructure with a high ratio of benefits to costs.

Major activities

Activities during the first year of the project included establishing the technical and organizational framework from which to coordinate and grow the networks of researchers who will provide and ultimately create new source data for the DINAA. We are also developing this professional network in order to promote use of DINAA information products to further expand the capabilities of American archaeology's cyberinfrastructure. Much of the work during this period involved communications with potential participants, presenting the project to the community at meetings, and developing tools for obtaining and integrating a subset of data in order to test the capabilities of the DINAA database and refine a workflow for the project. These activities are described in the following sections.

Secondly, we started software development on a variety of enhancements and new features to prepare Open Context (http://opencontext.org) for DINAA data dissemination. These include facilities for the publication of ontologies needed for semantic data integration of site file datasets and new map-based browsing and visualization interfaces and Web services.

Progress

Our progress during the first twelve months of the project has been as follows:

1. **Solicitation of Datasets:** During the first year of the project, project principals solicited and assessed the capacities of the datasets which will together form the DINAA through detailed communication with their state or agency managers. (Status: Complete)
2. **Solicitation of Testers:** Over the course of the first year of the project, project principals have solicited RPA-qualified professionals to test prototype DINAA structures. To date, the project has identified several researchers who have agreed to work with each of the project prototypes related to the Paleoindian, Archaic, Woodland, Late Prehistoric and Historic periods. These individuals are in addition to the project principals who, because of their particular research histories, are also qualified to conduct their own tests of the Paleoindian and Late Prehistoric datasets. We are developing additional contact with other researchers, with the goal of having analyses completed by early 2014, in time for presentation at the project workshop in the winter of 2014 (see below) (Status: Ongoing)
3. **Data Collection:** The majority of the first year of the project has involved collection of site file data structures and primary information. The aim was to produce a set of data that spans 15 ideally contiguous states in eastern North America. Databases have been 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. Project principals analyzed each dataset to determined variations in ontological structures of key archaeological terms and concepts, forms of quantitative data representation, expression of geolocational coordinates, and purposes and forms of ancillary data (site quality, previous investigations, ownership, etc.). Data collection will continue during the second year of the project. (Status: 50% complete)

4. **Data Integration.** Based on analysis of the first contributed datasets, an integrated database structure is being 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. (Status: Ongoing)

Significant results

This project develops, for the first time, 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. Results from the first twelve months of the project include:

- **Site sensitivity:** In the first few weeks of the project, we developed and publicly posted a policy pertaining to site sensitivity (see: http://ux.opencontext.org/blog/archaeology-site-data/dinaa-sensitive-data-security-measures-and-shpo-collaboration/). A project announcement was posted to the National Association of State Archaeologists list, which resulted in extensive on-line discussion among the state archaeologists, many of whom subsequently became partners in the project, and all of whom had questions and suggestions we addressed when we contacted officials in individual states directly. Maintaining data security was a big concern for many state site file managers and developing protocols to allay their concerns was the first major hurdle to cross before approaching them about their state's data.
- **Data transfer:** We developed and posted procedures for the secure transfer of data from states to the DINAA database (http://ux.opencontext.org/blog/wp-content/uploads/2012/09/DINAA-data-transfer-instructions-Oct12.pdf)
- **Development of ontologies:** We are progressing in the development on an ontology required for semantic integration of the site file data, basically permitting information from separate states to be combined for use in analysis and report/map generation. The ontology is expressed in OWL (the Web Ontology Language), an open-standard widely supported in semantic Web applications. We chose OWL instead of SKOS (as proposed) because tDAR supports OWL. Thus, we have aligned our efforts in a way that complements other information systems serving the archaeological community. In addition, we are using GitHub as a version control and collaboration tool to facilitate ontology development and promote transparency in our efforts. All revisions, changes, and discussion related to the ontology are publicly tracked and available for inspection and comment by the wider community. The ontology is publicly accessible here: https://github.com/ekansa/oc-ontologies/blob/master/vocabularies/dinaa.owl.

- **Data import and integration:** We have begun the process of importing into the DINAA database site files held by State Historic Preservation Offices (SHPOs) and allied federal and tribal agencies. These files contain unique site identification numbers, cultural affiliations, descriptions of condition and use, and location and management information, as well as ancillary information about site integrity and quality, diagnostic artifacts, associated collections and reports, and investigative methods.
- **Other Software Development and Testing:** We made great progress in developing features to support implementations of OWL ontologies (discussed above) and innovative geospatial browsing and Web-services of data. The DINAA dataset will obviously have an important geospatial component, even though we are legally and ethically barred from sharing specific site location coordinates. To enable map-based browsing and visualization of geospatial data, we made significant progress on new features for Open Context (http://opencontext.org), the Web publication venue that will host the DINAA data (see Figure 1). The new geographic browsing feature uses a GeoJSON web-service. GeoJSON is a popular and widely supported open format for expressing geospatial data on the Web. By using GeoJSON, Open Context will further enhance the interoperability of the DINAA.

Perhaps most significant accomplishment and simultaneously the biggest challenge during the first year of the project was in establishing working relationships with research partners and site file managers in many states in Eastern North America. Basically, this proved to be more time consuming than we had anticipated, and required more direct, personal contact either via conference calls or actual visits to state agencies, even those who had agreed to work with us previously, and had been so identified in our initial research proposal. Essentially, a lot of time has been spent in communicating project procedures and objectives, and establishing the trust and understanding essential to good working relationships. This is an ongoing process, and we now have good relationships with many colleagues across the region, providing an essential foundation for the project.

Key outcomes and other achievements

While the goal of the DINAA project is ultimately 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, we set a goal of 15-20 states for the project period. To date we have 13 partners and five others pending. We have completed data transfers from six states [eight states as of 20 September 2013], and have transfer agreements for seven additional states with data gathering in progress for transfer (Figure 2). We have also contacted five other states with positive outcomes expected.

Notes on individual state activities related to the DINAA project: *1. Requested state data submitted to DINAA; 2. Data transfer agreement with DINAA, data migration in progress; 3. Database structure and metadata submitted to DINAA; 4. Data transfer agreement under negotiation*

Alabama[2]: DINAA has negotiated partnership with the Alabama Office of Archaeological Research (OAR), Director Matthew D. Gage. Alabama archaeological data are used virtually in the OAR office from an off-site location and OAR staff members are working with their state data repository managers to acquire our requested materials in a transferable format.
Florida[1]: DINAA has negotiated partnership with the Florida Division of Historical Resources, Master Site File Supervisor Vincent Birdsong, and taken secure delivery of requested data from the Florida archaeological database.
Georgia[1]: DINAA has negotiated partnership with the Georgia Archaeological Site Files, Director Mark Williams, and taken secure delivery of requested data from the Georgia archaeological database.
Illinois[2]: DINAA has negotiated partnership with the Illinois State Museum, database coordinator Michael Wiant. IAS staff members are currently processing our data request.
Indiana[1]: DINAA has negotiated partnership with the Indiana Department of Natural Resources (DNR) Division of Historic Preservation and Archaeology (State Archaeologist Richard Jones) and Office of Legal Counsel (Counselor Andrew J. Wells). DINAA staff has been granted complete access to the Indiana online archaeological database for professional archaeologists. After submission of a request for data, DINAA has taken secure delivery of requested data from the Indiana archaeological database.
Iowa[2, 3]: DINAA has negotiated partnership with the Office of the Iowa State Archaeologist (OISA), directed by John Doershuk. DINAA staff has been granted complete access to the Iowa online archaeological database for professional archaeologists, and have submitted a data request that is being processed by OISA site records staff.
Kentucky[2]: DINAA has negotiated partnership with the Kentucky Office of State Archaeology (KOSA), Director George M. Crothers. KOSA staff members are currently processing our data request.
Louisiana[3, 4]: DINAA is in a process of negotiation with the Louisiana Division of Archaeology (LDA), Director Charles McGimsey, to transfer and utilize data in a manner that is consistent with Louisiana state guidelines. This process is expected to culminate in DINAA partnership and a complete data transfer per project parameters. LDA staff has provided DINAA with their database and metadata structures to coordinate Louisiana data with other sets.
Maryland[2, 3]: DINAA has negotiated partnership with the Maryland Historical Trust (MHT), State Terrestrial Archaeologist Charles L. Hall. DINAA staff has been provided with metadata for the state archaeological database structure and have submitted a data request that is being processed by MHT site records staff. [DINAA took secure delivery of requested data in September].
Mississippi[4]: DINAA has contacted the Mississippi Department of Archives and History, Chief Archaeologist Pamela Edwards Lieb, to determine the extent to which Mississippi data structures are compatible with the DINAA project. This process will hopefully culminate in DINAA partnership and a complete data transfer per project parameters.
Missouri[1]: DINAA has negotiated partnership with the Missouri State Historic Preservation Office, Cultural Resource Inventory Coordinator Kerry Nichols, and taken secure delivery of requested data from the Missouri archaeological database.
North Carolina[2]: DINAA has negotiated partnership with the North Carolina Office of State Archaeology (NCOSA), State Archaeologist Stephen R. Claggett. NCOSA staff members are currently processing our data request. [DINAA took secure delivery of requested data in September]

Ohio[3, 4]**:** DINAA is in a process of negotiation with the Ohio Historical Society – Ohio Historic Preservation Office (OHS-OHPO), Archaeology Survey and Data Manager Brent A. Eberhard, to transfer and utilize data in a manner that is consistent with Ohio state guidelines. This process will hopefully culminate in DINAA partnership and a complete data transfer per project parameters. OHS-OHPO staff has provided DINAA with their database and metadata structures to coordinate Ohio data with other sets.
Pennsylvania[1]**:** DINAA has negotiated partnership with the Pennsylvania Bureau for Historic Preservation, GIS Database Coordinator Noel Strattan, and taken secure delivery of requested data from the Pennsylvania archaeological database.
South Carolina[1]**:** DINAA has negotiated partnership with the South Carolina Institute of Archaeology and Anthropology, Director and State Archaeologist Jonathan Leader, and taken secure delivery of requested data from the South Carolina archaeological database.
Vermont[4]**:** DINAA is in a process of negotiation with the Vermont Division for Historic Preservation, State Archaeologist and Director Giovanna Peebles, to determine the extent to which Vermont data structures are compatible with the DINAA project. This process will hopefully culminate in DINAA partnership and a complete data transfer per project parameters.
Virginia[2, 3]**:** DINAA has negotiated partnership with the Virginia Department of Historic Resources (DHR), Archaeology Inventory Manager Jolene L.U. Smith. DINAA staff has been granted complete access to the Virginia online archaeological database for professional archaeologists, and have submitted a data request that is being processed by DHR site records staff.
West Virginia[4]**:** DINAA has contacted the West Virginia Division of Culture and History, Deputy State Historic Preservation Officer Susan M. Pierce, to determine the extent to which West Virginia data structures and state guidelines are compatible with the DINAA project. This process will hopefully culminate in DINAA partnership and a complete data transfer per project parameters.

The database structures and data are processed, documented, and separated from spatial information and attributes that would reveal sensitive site locations separately as they are incorporated into DINAA. This process creates the opportunity to evaluate and re-evaluate protocol and workflow. These protocols and workflows will serve as the basis for standard operating procedures and best practices particularly in the handling of sensitive site information when interfacing with the web.

Spatial analysis, pattern recognition, cluster analysis, etc. will provide exciting new ways of investigating the cultural resource record of North America when applied to the DINAA ontology. Figure 3 shows a dot density distribution of site counts across the six states incorporated into DINAA thus far. Dots do not refer to exact site locations, but to groups of five sites whose position has been randomly distributed within 20x20km grid cells. Even at this small scale (1:12,000,000), regional patterning is observable with minimal data transformation hinting at the potential utility of the Digital Index of North American Archaeology.

Opportunities for training and professional development

The project will fund graduate students, and will assist their training in critical information management skills for the 21st century. At the University of Tennessee one graduate PhD

student, Thaddeus G. Bissett, M.A., R.P.A., will be working on the project throughout the second year, helping with both integration of the various site files, as well as developing research products from them. Mr. Bissett has extensive GIS and database management experience, and will be working closely with the other PIs, particularly Wells and Yerka, who also have extensive experience in these areas.

At Indiana University, two graduate students on the Bloomington campus are directed in activities by Dr. Wells from the South Bend campus. Ms. Kelsey Noack Myers, M.A., R.P.A., and Mr. R. Carl DeMuth, have been employed as part-time hourly members of the DINAA staff. As a registered professional archaeologist, Ms. Noack Myers is capable of dealing with sensitive site location data if necessary. These two employees have undergone professional development with a variety of digital resources, including secure data transfer standards using the Indiana University Slashtmp system, MS-Access database standards and other database systems, Protégé ontology editor and OWL web ontology language, and the software development web-hosting service GitHub. These two employees have created individual and comparative prospectuses for state archaeological data structures based on data transfers to DINAA, metadata transfers to DINAA, and online guidebooks and other public metadata made available by individual state offices. The work of these employees has been integral to the assessment of various state standards in data management for project planning purposes, to the development of "pick lists" of data types communicated to state partners, to the inspection of data shared with DINAA, and to the production of all DINAA conference presentations (see below) to date, and for which these employees were listed as co-authors.

Dissemination

Because the DINAA project involves individuals in over a dozen states, and is of interest to site file managers in all states, as well as other countries around the world where similar datasets are maintained, clear and timely communication is critical to the project's success. Since the project's inception, we have communicated with active and potential project participants, as well as with any others interested in following the project's developments, by:

1. Posting messages to the NASA listserv (National Association of State Archaeologists)
2. Posting all updates and general communications on the project website: http://ux.opencontext.org/blog/archaeology-site-data/. Key communications during the first twelve months of the project included:
    a. Initial project announcement: http://ux.opencontext.org/blog/2012/08/07/nsf-support-for-linking-state-site-files/
    b. Discussion of sensitive data: http://ux.opencontext.org/blog/archaeology-site-data/dinaa-sensitive-data-security-measures-and-shpo-collaboration/
3. Presenting papers and posters about the DINAA project at major conferences. Conferences this year included:
    a. Society for American Archaeology (Honolulu, April 2013), where we presented two posters:
        - "Archaeology is Data Coding or it is Nothing: Relationships between State and Governmental Archaeological Site File Structures and Professional Definitions through the Digital Index of North American Archaeology (DINAA)" (Figure 4)

- "Web-based Discovery and Integration of Site File Data: The Digital Index of North American Archaeology (DINAA)" (Figure 5)

b. The 2013 Digital Humanities (DH) conference (Lincoln, Nebraska, July 2013), where we presented a paper "An Introduction to the Practices and Initial Findings of the Digital Index of North American Archaeology (DINAA)" in the session *Current Research & Practice in Archaeology*.

Beginning in 2013, email and on-line posting of research results to our research partners and the larger profession will be provided on a bimonthly basis. A workshop of project partners will be held in Knoxville in the winter of 2014, bringing together site file managers and researchers, to highlight what has been accomplished, the state of site file management in the region, and how to proceed in the future. The conference proceedings will form the basis for professional papers and a technical report summarizing the results, to be produced in the spring and summer of 2014, with a final project report to be submitted to NSF by September 2014.

Next steps

Over the course of the project period, we will continue to accept datasets for upload and integration. In September 2013, we will begin analysis and evaluation of the existing data. This will occur at various levels of networked community and data abstraction, ranging from precise levels of analysis (at ca. 30m spatial resolution, with strict data security) by collaborative researchers and project principals to much broader levels (ca. 20km) by the general archaeological community and interested public. This work will be facilitated through public electronic forums. Also during this phase, we will hold a face-to-face meeting among project researchers and data contributors. The three-day meeting will bring together DINAA project team members and active data contributors, prototype testers, researchers, and knowledgeable representatives of all cooperating entities that manage site files. The purpose of the meeting, which will take place in Knoxville at the University of Tennessee in the winter of 2014, is 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. Results of the meeting will be presented at the Society for American Archaeology conference in Austin in April 2014), and subsequently through more formal journal articles and a monograph-length technical report comparable to the report of an earlier research project involving site file managers from across the southeast (Anderson and Horak 1995).

From the incipience of the DINAA project, we have as a goal to make the project protocols and technical products available online and shared with the public and the scientific community. During the next phase of the project, the team members will continue to prepare presentations and publications describing the goals, procedures, and technical results of this work, which we will disseminate in the archaeological community and beyond (informatics, digital humanities).

Supporting Files

PDF with Figures 1-9.

PRODUCTS

Publications:

Joshua S. Wells, Eric C. Kansa, Sarah Whitcher Kansa, Stephen J. Yerka, and David G. Anderson. "Developing the Cyberinfrastructure for a National Archaeological Site Database: The Digital Index of North American Archaeology (DINAA) Project." To be submitted to *Advances in Archaeological Practice*.

Technologies:

Map-based browsing: In the past twelve months, Open Context has been enhanced to support map-based browsing of data. This will be an important feature for navigating and visualizing data compiled by the DINAA project. Below are two examples of these map-based browsing features (currently in a testing phase):

- Example 1: Counts of items classified as cattle (*Bos taurus*), See Figure 1 and on the web: http://tinyurl.com/p88xvo8
- Example 2: Percentage of sheep/goat at sites in Turkey compared to all items with a biological classification, See Figure 6 and on the web: http://tinyurl.com/pvokmpl

DINAA has the goal of building freely and publically accessible information resources for American Archaeology. A necessary concern for the DINAA project lies in protecting sensitive site location data while still offering mapping capabilities for the public and the research community. Thus, we are working toward methods that enable map-based interfaces and visualizations of the data without endangering sensitive cultural heritage resources in the landscape.

We have considered the issue in some depth and are developing methods and data processing protocols to reduce exposure of sensitive geospatial data both directly and through inference. Some risk mitigation steps include:

(1) Reducing risks of Accidental Disclosure: In order to reduce the risks of accidental exposure of specific site location data, we have "compartmentalized" information-handling within our team. Eric Kansa and Sarah Whitcher Kansa, responsible for public dissemination, do not have access to the source data files that have site location information. Only data files stripped of sensitive data are transferred to project personnel responsible for public dissemination.

(2) Encyrption: Data transfers between SHPOs and the project staff use secured, encrypted methods.

One of the most challenging issues in managing geospatial data risks lies in the possibilities of combining multiple datasets and reasoning to infer and closely approximate site locations. For

example, we could naively assign sites to certain arbitrarily defined geospatial regions (as used in the Open Context map interface). At first glance, assigning site locations to large geospatial regions reduces risks, because a site can be anywhere with a region (see Figure 7).

However, other clues in the SHPO data can be used to infer locations with greater precision. One can sometimes derive fairly close site locations by combining even imprecise location data with publicly available state and county boundaries (county information is generally associated with each site record). See examples in Figures 7, 8, and 9.

With respect to mitigating such inference risks, we are exploring ways metadata about sites can be combined with other locational cues. Most SHPO datasets provide information about county locations. In these datasets, counties are generally the most specific spatial reference in site records (in addition to coordinates). To reduce the inference possibilities of combining site locations (even approximated to large regions, as in the examples above) with county boundaries we are first approximating the locations of sites to an arbitrary point *within* a county boundary. This step ensures that combining publicly disclosed location information with county boundaries provides no opportunities for inference, since the publicly disclosed location information is already based on county boundaries.

The DINAA project team is dedicated to developing new and innovative ways to help protect vulnerable resources that contribute significantly to the understanding of our shared North American cultural heritage while providing frameworks for access to the archaeological research community as well as the general public.

Ontology development: As discussed, we are using GitHub as a version control system to facilitate collaboration, track document changes, and monitor issues associated with the ontology development. We are also using Protege, an open source ontology editor for creating, documenting, and visualizing ontology concepts. The ontology is publicly accessible via GitHub at the following link: https://github.com/ekansa/oc-ontologies/blob/master/vocabularies/dinaa.owl

Websites:

The project website: http://ux.opencontext.org/blog/archaeology-site-data/

The Digital Index of North American Archaeology (DINAA) is a multi-institutional undertaking to create interoperability models for archaeological site databases in the eastern United States. The 2-year project, which is funded by the National Science Foundation, commenced September 1, 2012. View the technical proposal.

This page provides ongoing information about the progress of the project. Visit the links below to learn more about this project. We also broadcast updates to the community via the NASA (National Association of State Archaeologists) listserv. We welcome your questions, comments, and other participation in this project, so please comment on the blog posts or contact us directly by emailing project PIs David Anderson and Josh Wells.

References Cited

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.

Supporting files [These are attached at the end of the report figures, section]

UPLOAD (5MB max each):
(1) Society for American Archaeology April 2013 poster "Archaeology is Data Coding or it is Nothing: Relationships between State and Governmental Archaeological Site File Structures and Professional Definitions through the Digital Index of North American Archaeology (DINAA)" Joshua J. Wells, Stephen J. Yerka, David G. Anderson, Kelsey Noack, R. Carl DeMuth, Eric C. Kansa and Sarah Whitcher Kansa.

(2) Society for American Archaeology April 2013 poster "Web-based Discovery and Integration of Site File Data: The Digital Index of North American Archaeology (DINAA)" Eric C. Kansa, Sarah Whitcher Kansa, David G. Anderson, Stephen Yerka, and Joshua G. Wells.

(3) Digital Humanities July 2013 Conference An Introduction to the Practices and Initial Findings of the Digital Index of North American Archaeology (DINAA) Joshua J. Wells, David G. Anderson, Stephen J. Yerka, Eric C. Kansa, Sarah Whitcher Kansa, Kelsey Noack Myers, and R. Carl DeMuth.

Participants / Collaborators

David Anderson: 1.5 months in 2012/2013
Assisted in the collection and integration of state archaeological site files with the other project PIs.
Funding Support: University of Tennessee
No foreign travel on international collaborations on this project.

Joshua Wells: 1.9 months in 2012/2013

Stephen Yerka: 2 months in 2013
Contribution to the Project: Assisted in the collection and integration of state archaeological site files with the other project PIs.

Funding Support: NSF
No foreign travel on international collaborations on this project.

Eric Kansa: 2 months in 2012/2013
Contribution to the Project: Assisted in the collection and integration of state archaeological site files with the other project PIs. Developed project website and linkages with Open Context for hosting DINAA products.
Funding Support: NSF
No foreign travel on international collaborations on this project.

Sarah W. Kansa: 1 month in 2012/2013
Contribution to the Project: Assisted in the collection and integration of state archaeological site files with the other project PIs. Developed project website and linkages with Open Context for hosting DINAA products.
Funding Support: NSF
No foreign travel on international collaborations on this project.

Kelsey Noack Myers: 0.9 months in 2012/2013
R. Carl DeMuth: 0.9 months in 2012/2013

Project Impacts

Impact in archaeology

The DINAA project serves as a public demonstration of best practices for cyberinfrastructure. It also fosters novel networking and data integration among multiple partners across disciplinary and geographic boundaries. This project will share lessons about the capabilities of the data and best practices 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.

Impact on other disciplines

The DINAA project will foster novel networking and data integration among multiple partners, as well as research and educational activities across multiple disciplines and geopolitical boundaries.

Impact on human resources

The project has facilitated communication between site file managers across eastern North America, and helped them become aware of operations in states other than their own. This communication is expected to grow in the second year of the project as research projects are developed using combined site file data, and as researcher and site file managers prepare for and come together for the project workshop in the winter of 2014. The resulting technical papers and monograph resulting from this project and workshop, furthermore, will provide a baseline summary of the current state of archaeological data management across much of the eastern United States, and offer guidance about how to proceed in the future.

Impact on institutional resources that form infrastructure

The DINAA, upon completion of this project, will be web-based and, thus, will be global, accessible to researchers across multiple disciplines and institutions. As such the project should be able to grow and develop independently, past this initial effort. As noted, our ultimate objective is to have the DINAA, or something like it, eventually operate at multiple levels, from the local level (i.e., at the state or portion thereof in scale), regional (multiple states), to continental or even global in scope. Ideally, we envision the same kind of project we are attempting here linking data between states to be adopted in the years to come to link information from national datasets, to provide analytical potential at a truly global scale. While this will take time and resources to develop, the procedures by which it could be done are being developed as part of the DINAA project.

Impact on information resources that form infrastructure

The DINAA provides mechanisms to link information from state and national datasets, to provide analytical potential at local, regional, and national scales.

Impact on technology transfer

The project will develop significant data resources in a 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.

Impact on society beyond science and technology

The DINAA project is demonstrating how opening up data can benefit research. This is something that has yet to be successfully demonstrated in archaeology at such a large geographic scale. Many resources exist, and many new ones are being created, but the potential and feasibility of linking those resources is not well understood. A major goal of the DINAA project is to demonstrate the feasibility of linking numerous disparate data sources and their potential to address large-scale research questions.

Project Changes

The original project name in the approved proposal was the National Archaeological Site Database (NASD). At the start of the project, the project principals changed the name to the Digital Index of North American Archaeology (DINAA). The change was done in response to comments from partners, who thought the original acronym was too limiting, and did not encompass the full potential of the project. And, frankly, DINAA, besides being more accurate, looks and sounds better.

Developing the Cyberinfrastructure for a National Archaeological Site Database
Award Number: 1217240

Annual Report 2012-2013, Figures
August 31, 2013

Figure 1. Map-based browsing (Old World data example) on Open Context, the web publication venue that will host DINAA data products.

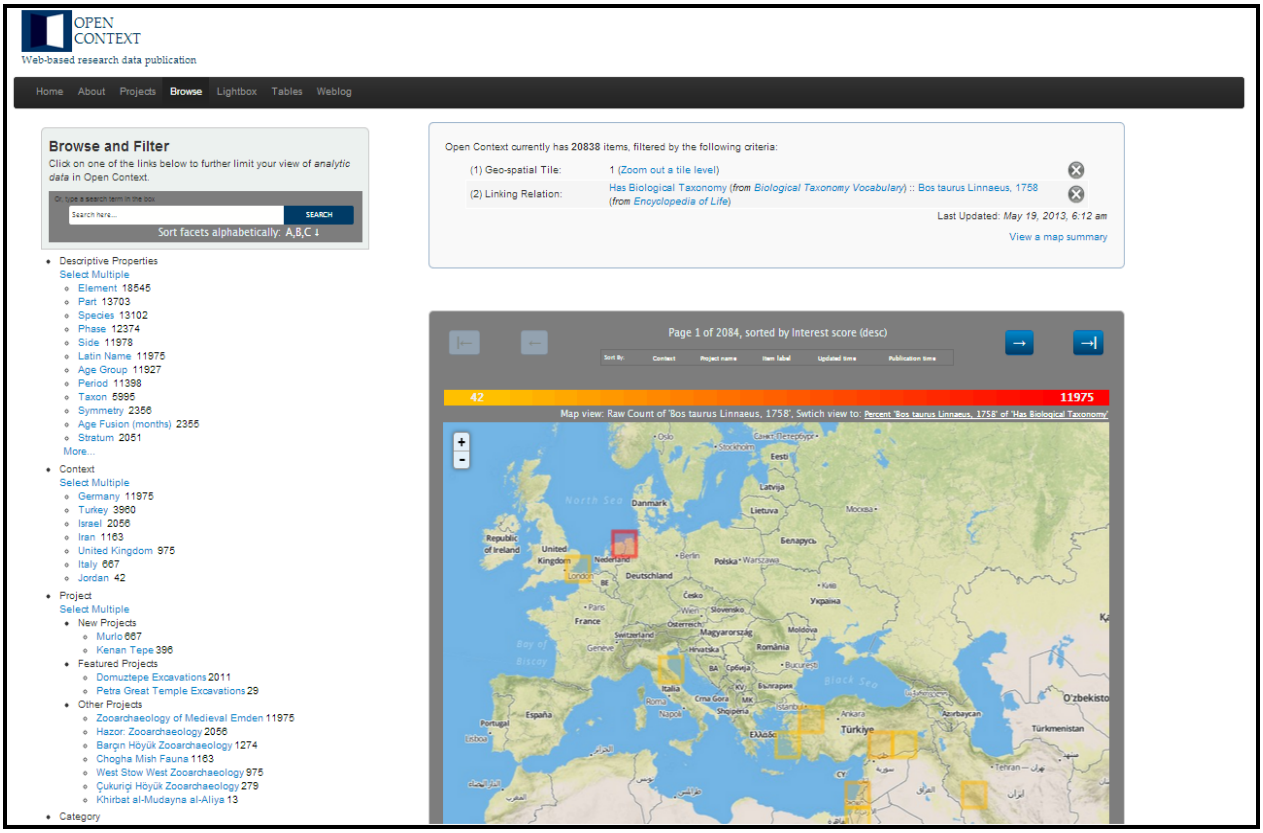

Figure 2. DINAA partnerships as of 30 September 2013. Colors in the figure represent the phase of interaction with partners.

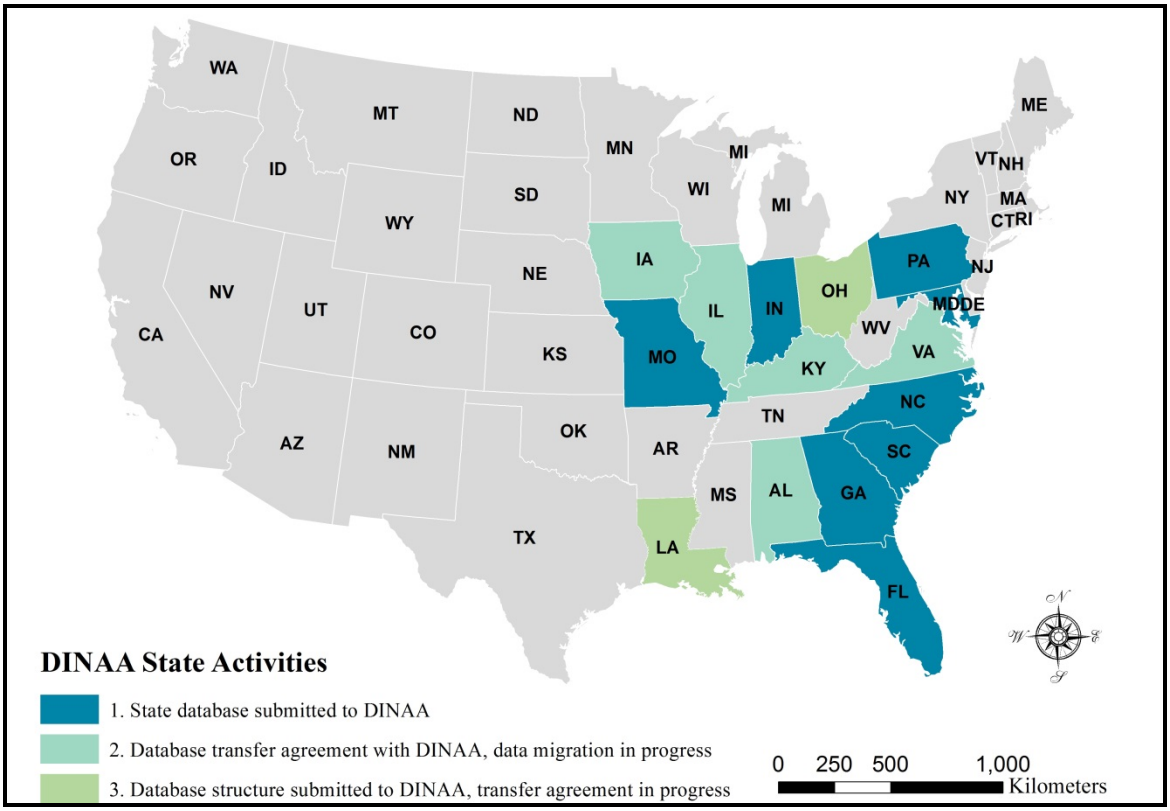

Figure 3. DINAA partnerships as of 30 September 2013 with dot density plot showing the distribution of cultural resources at low resolution within states whose data has been received thus far. Dots do not refer to exact site locations, but to groups of five sites whose position has been randomly assigned within arbitrary 10x10km grid cells (about the size of a USGS 7.5` Quadrangle).

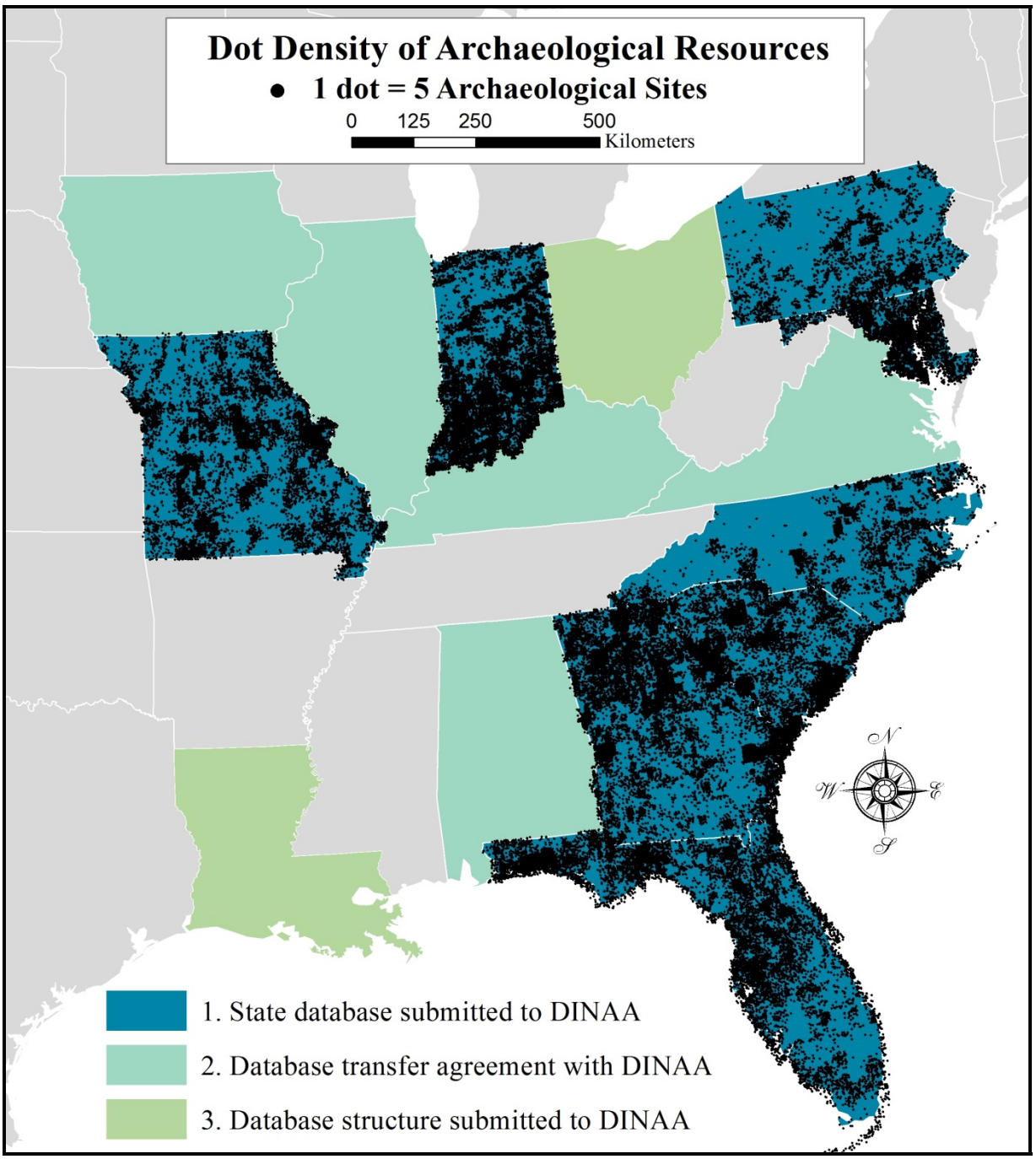

Figure 4. Society for American Archaeology April 2013 poster "Archaeology is Data Coding or it is Nothing: Relationships between State and Governmental Archaeological Site File Structures and Professional Definitions through the Digital Index of North American Archaeology (DINAA)" Joshua J. Wells, Stephen J. Yerka, David G. Anderson, Kelsey Noack, R. Carl DeMuth, Eric C. Kansa and Sarah Whitcher Kansa

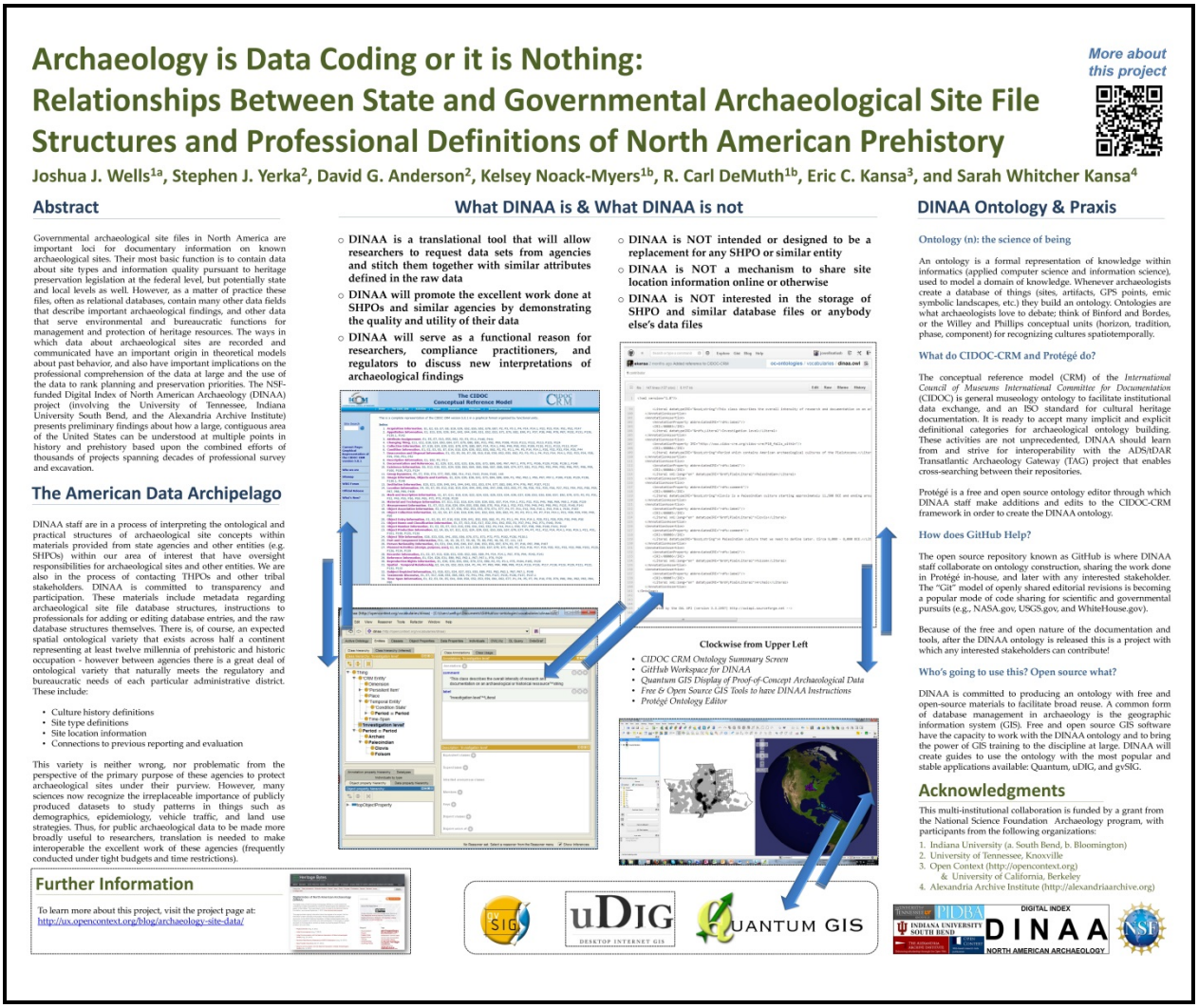

Figure 5. Society for American Archaeology April 2013 poster "Web-based Discovery and Integration of Site File Data: The Digital Index of North American Archaeology (DINAA)" Eric C. Kansa, Sarah Whitcher Kansa, David G. Anderson, Stephen Yerka, and Joshua G. Wells

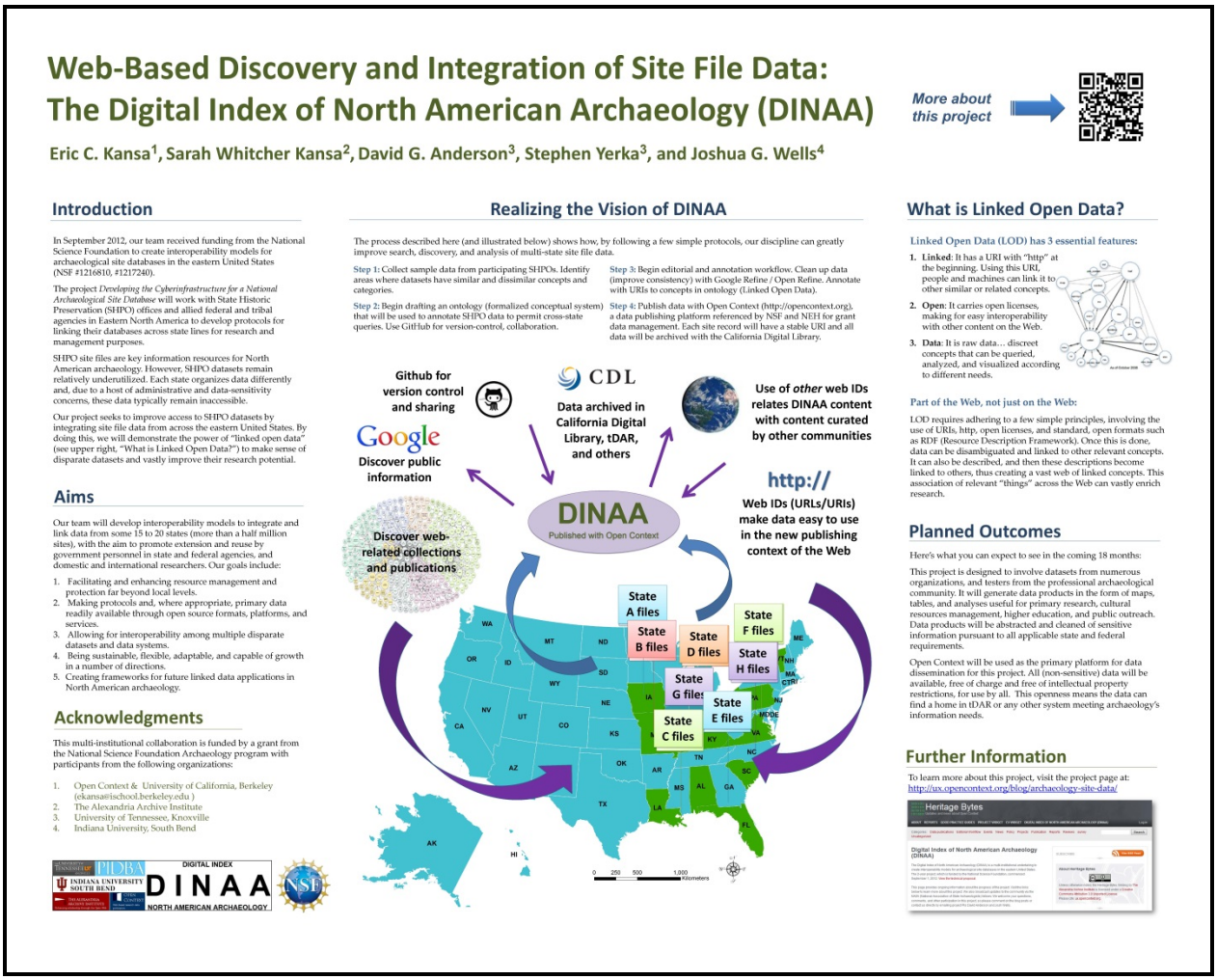

Figure 6. Sheep/goat as a percentage of all biological taxa

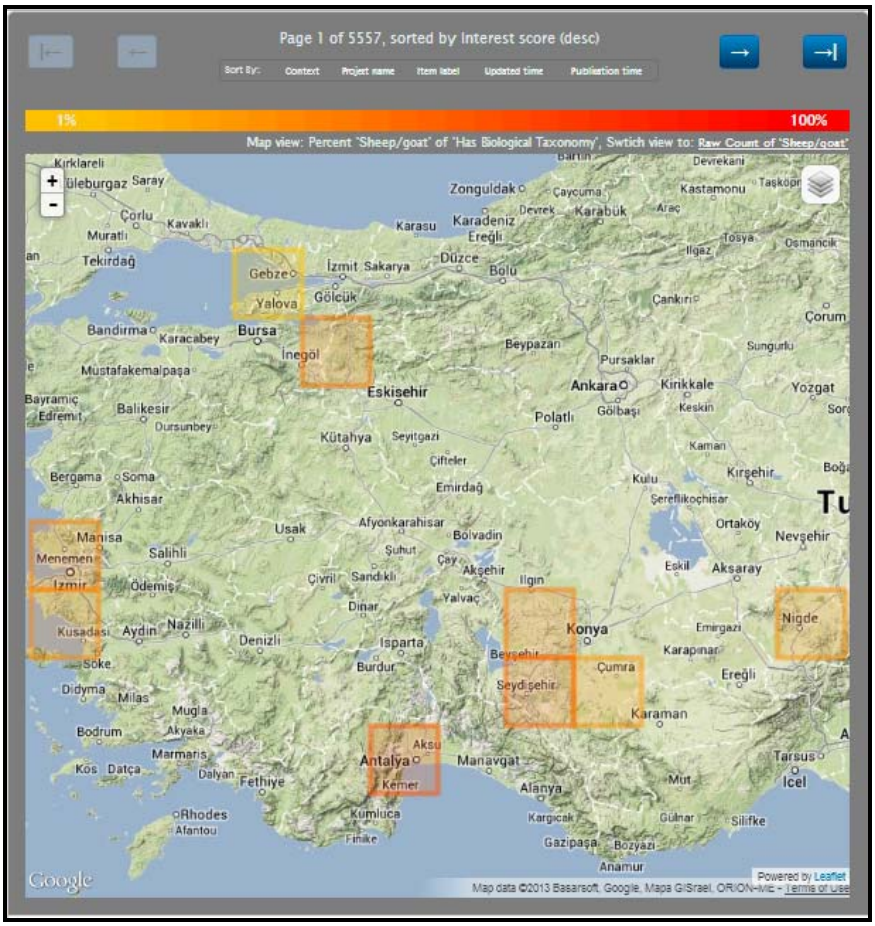

Figure 7. Louisiana, Mississippi and Alabama Gulf Coast with arbitrary regions

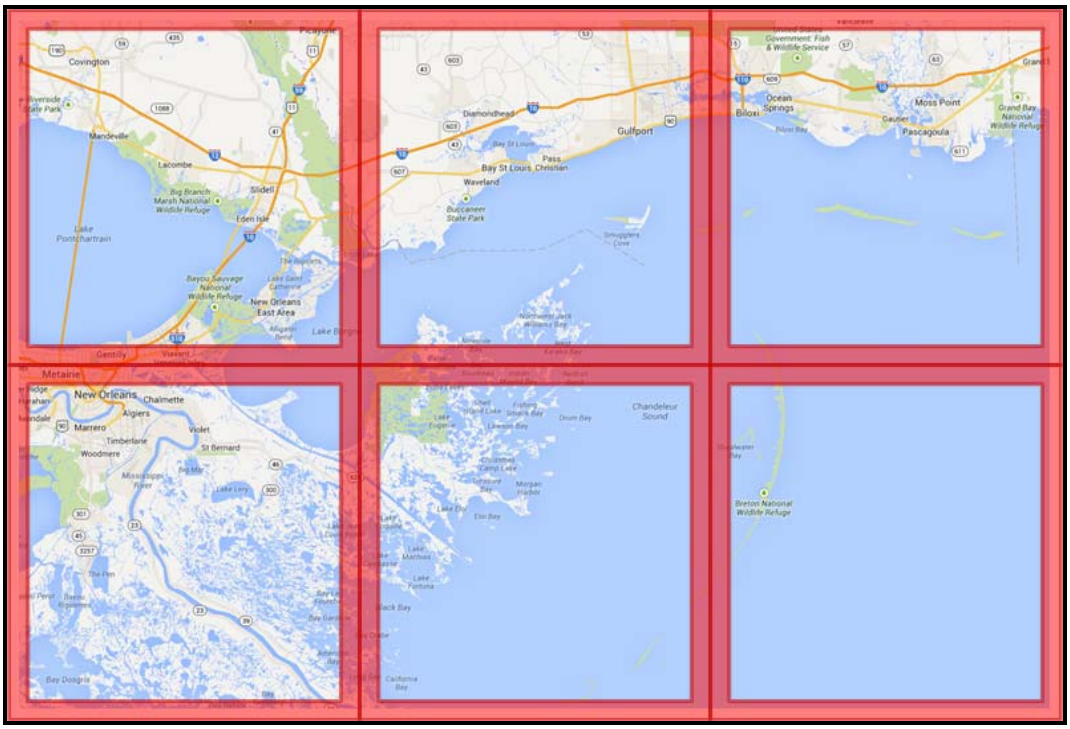

Figure 8. A hypothetical Louisiana site located in a specified region

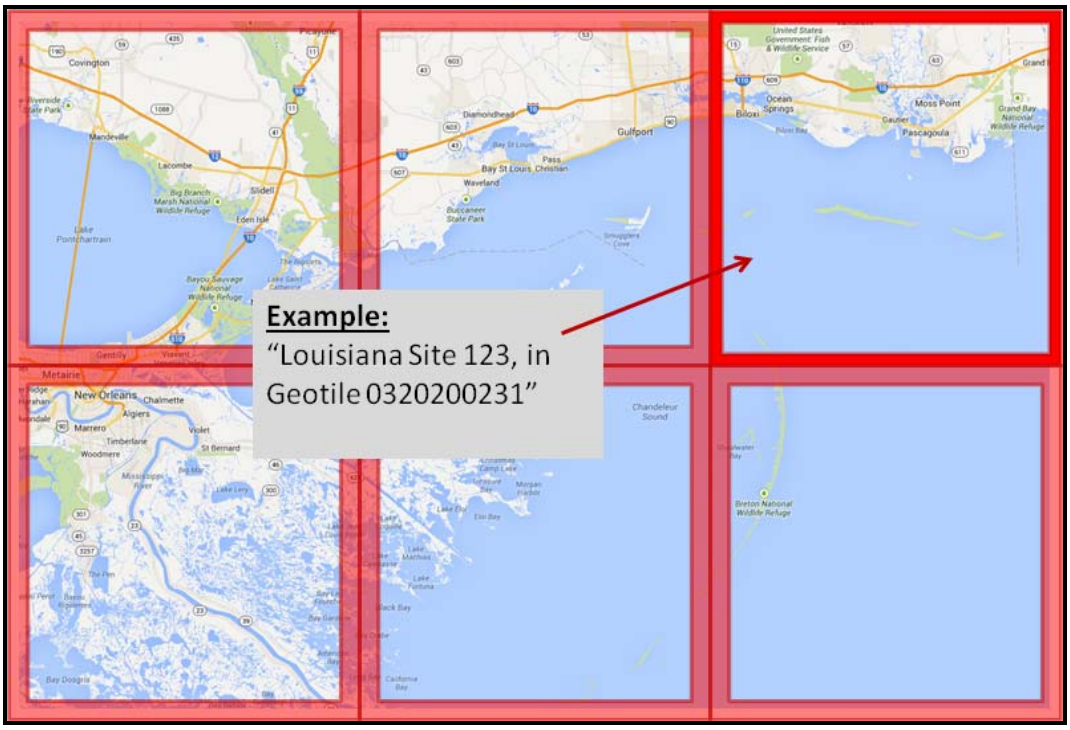
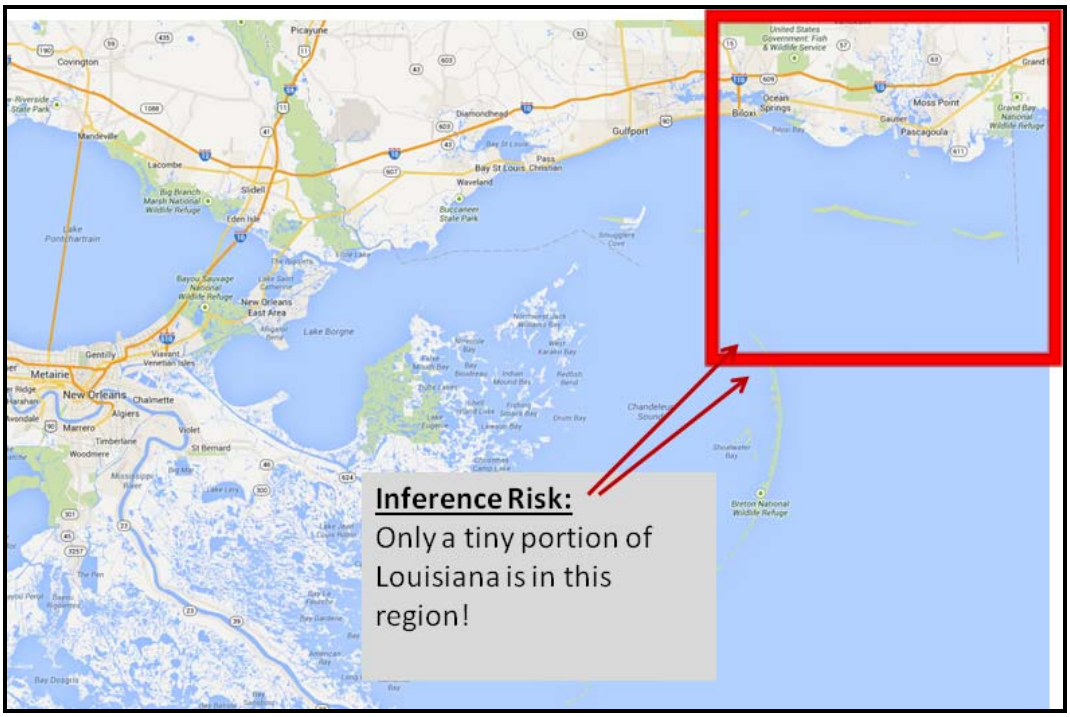
Figure 9. A hypothetical inference: Site location on tiny portion of Louisiana in this region