

**Final Report**

**for**

**Digital Antiquity:**

**Planning a Digital Information Structure for Archaeology**

**Grant Period: July 1, 2007 to June 30, 2008**

**Submitted to**

**The Andrew W. Mellon Foundation**

**Scholarly Communications Program**

**by**

**W. Fredrick Limp**

**University of Arkansas**

**Fayetteville AR 72701**

## Executive Summary

In July 2007 The Mellon Foundation Scholarly Communications program awarded \$152,000 (and an additional \$3,861 in interest accrued) to the Archaeoinformatics Consortium with the University of Arkansas serving as prime contractor. The mission of the consortium was to create a cyberinfrastructure to preserve digital archaeological data in all their forms, along with the metadata that make them meaningful and to provide scholars and the public access to archaeological digital data. In addition to the University of Arkansas (Uark), other initial participants were Arizona State University (ASU), the Penn State University (PSU), Statistical Research Incorporated (SRI), and Washington State University (WSU); in February 2008, the University of York was added to the consortium. The consortium steering committee was: Jeff Altschul (SRI), Keith Kintigh (ASU), Tim Kohler (WSU), Fred Limp (Uark), Julian Richards (York) and Dean Snow (PSU). Over the project period (July 2007-June 2008) the Steering Committee met bi-weekly utilizing the AccessGRID video conferencing system. They developed an extensive web site (<http://archaeoinformatics.org>) and hosted an AccessGRID-based public lecture series with 10 public lectures. Video of each lecture and associated presentation materials are archived at the web site. A disciplinary survey to define the current “state of the art” and to identify needs was conducted; it had 270 respondents. A Technical and Disciplinary Advisory Board was formed with prominent individuals from the fields of computer science, archaeology, libraries and physical anthropology. Board members represented consulting archaeology, university and museum researchers, various professional organizations and federal agencies. A three day meeting was held in Santa Fe NM in February 2008 that lead to a formal set of recommendations and an assessment of work performed to date. A preliminary plan was presented at the Mellon Foundation All-Projects meeting in March 2008. Steering Committee members under the director of Keith Kintigh worked closely with Mellon staff to prepare a series of increasingly structured documents defining the technical and organizational elements needed to execute the consortium’s defined mission. A final implementation proposal formalizing the results of the planning process was submitted to the Foundation in October 2008.

Individual and summary fiscal reports are provided as appendices. As documented in the appendices the funds allocated were expended consistent with the proposed tasks. For the University of Arkansas distribution a positive balance of \$398.09 remains after all expenditures were paid. Minor (positive) variances also exist between budgeted and allocated subcontracts for WSU (\$992.73), PSU (\$173.51) and Arizona State (\$679.66). One substantial deviation between the proposed and expended funds is in the SRI budget. Because of staffing changes during the project at SRI, Dr. Clay Mathers was replaced by Dr. Jeff Altschul. Dr Altschul is the founder and Chairman of SRI and the president of the SRI Foundation. Due to the fiscal structure of SRI, Dr. Altschul’s salary cannot be charged and as a result only \$8,289.41 of the total \$21,801 initially requested by SRI could be expended, leaving \$13,511.59 unexpended. The total remaining amount, from all the sources (\$15,755.58), is being held by the University of Arkansas to be returned to the Mellon Foundation or otherwise distributed as directed.

## Organizational History and Structure

Over the last few years, a number of previous reports and studies had identified the lack of a cyberinfrastructure as a key constraint on archaeology, particularly in the Americas. One key report on the problem was Kintigh et al. (2006) who stated:

Archaeological insights have enormous potential to contribute to the understanding of long term social and socio-ecological dynamics. However, the complexities of archaeological datasets, the lack of data comparability across projects, and limited access to primary data have crippled our efforts to understand phenomena operating on large spatial and temporal scales. The fundamental challenge is to enable scientifically meaningful integration and use of the expanding corpus of systematically collected archaeological data.

Snow and associates, in an article in *Science*, (2006):

Archaeology often depends on archived data acquired by other researchers for other purposes, often long ago. Differences in recording protocols, terms, measurement units, and language are commonplace. Data are often obscurely archived and difficult to access, and policies regarding confidentiality vary considerably. Even when databases are accessible, they often differ in size, format, structure, and semantics and seem to defy fusion. In archaeology, research on the most important issues in today's society—the evolution of culture, the growth in population, and the long-term interaction of cultures with their physical and biological environments—will remain impoverished in the absence of a new generation of cybertools.

And Waters emphasizes (2007) that the process is complex:

... discovery is only the beginning of the scholarly process. Scholars then must zero in on the subsets they find—the primary and secondary source objects of interest to their work. They need to pull together these selected subsets for deeper analysis. The process of aggregation at this stage is more difficult and complicated because data need to be reviewed for anomalies, normalized, and prepared in a more rigorous fashion than is likely to be necessary or affordable for the commodity search engines. Provenance and authenticity of the information need to be established; rights need to be cleared; databases and database schemas need to be created; textual objects may need to be translated and marked up for grammatical and structural features, as well as semantically according to certain knowledge structures; numeric data may need conversion to common measures; assumptions and guesswork need to be carefully documented; and provision needs to be made to ensure that the data are maintained and can be reliably cited over time. The maintenance and preservation functions compose what is coming to be known as *data curation*, but the broader set of computationally-based research practices define the domain of *informatics*, which transformed the field of biology beginning in the 1980s and which is gradually being applied in other fields of study today.

In Europe generally, and Britain specifically, substantial progress has already been made in the development of key aspects of an archaeological cyberinfrastructure (or e-infrastructure), as represented primarily by the Archaeology Data Service (ADS) but also by efforts such as the International Committee for Museum Documentation (CIDOC), the Forum on Information Standards in Heritage (FISH), the Manual and Data Standards for Monument Inventories (MIDAS), and many others.

A key premise of the entire effort was that, while it is archaeologists who will be the first to benefit from an improved cyber infrastructure, it is critical to recognize that

cultural heritage is central to a people's identity. Nearly all countries mandate its protection, often through archaeological documentation and interpretation. The curation in perpetuity of irreplaceable archaeological artifacts, field notes, photographs, reports, and analytical data underlies the public's faith that their heritage is safeguarded for the future. Sadly, we are currently failing to fully meet this responsibility. In a digital age, we must not only care for physical objects, but also the digital data— reports, photographs, laser scans, satellite images, databases from laboratory analyses, and much more— that are a growing part of archaeological collections. Moreover, the data are only important if they are used to help us understand our common past. It is incumbent on the discipline to not simply store digital data, but to find ways of making these data accessible and useful to scholars and the public. (Archaeoinformatics 2007).

In the US, between \$650M and \$1B is spent annually on cultural resource management (CRM) (J. Altschul per. comm.) of which a large proportion is devoted to archaeology. Nearly all of this work is performed to comply with laws that require government agencies to take into account the effect of their actions on cultural and historical resources. On the order of 50,000 field projects a year are carried out by federal agencies under these mandates, with another 50,000 federal undertakings requiring record searches or other inquiries that do not result in fieldwork.

Nearly all of the 50,000 annual field projects and many of the additional 50,000 undertakings generate reports that together constitute the "grey literature" whose inaccessibility has long been an issue of major concern. Today these reports (ranging from a few pages to thousands of pages) are generated digitally but are generally not archived digitally; most survive only in their paper forms. A large fraction of projects will produce databases or spreadsheets of primary data and photographs and other sorts of images. Many will generate GIS or CAD files and some will have other kinds of digital data, including 3D scans and geophysical data. Project records (including digital records) and artifacts of federally mandated projects are generally subject to the federal curation regulation (36 CFR 79 "Curation of Federally-owned and Administered Archeological Collections") that requires preservation and access to the collections in a repository meeting federal standards.

With so much work being performed and so much data being generated, it is not surprising that archaeologists working in the same region do not know of each others' work. Decisions about whether to preserve particular sites, how many of specific types to excavate, and how much more work needs to be done are being made in an informational vacuum. The situation is so dire that the Bush administration's Preservation Summit targeted access to digital CRM information as one of the country's most important historic preservation goals for the next decade.

Currently, digital data are being curated at repositories (generally museums) that tend to focus on the preservation and access to the physical objects. Unfortunately, the digital data are usually treated not as information but simply as the physical objects (CDs, floppy disks, computer tapes, and computer cards)

on which the data are recorded and are simply stored on shelves. Few repositories are staffed or equipped to responsibly deal with digital data in the long term, and it is clear that digital data treated in this manner are being lost at an alarming rate.

Archaeologists have long realized that long-term preservation of digital data is not being adequately addressed, but to date little has been done. Officials of the agencies legally accountable for these collections have made clear their concerns for the sustainability of the digital collections. Yet without accepted standards, there has been no clear way forward.

It is within this context that the Archaeoinformatics.org consortium was created.

### **Archaeoinformatics.org**

Archaeoinformatics.org (AIO) was established in the Fall of 2006 as a collaborative organization to design, seek funding for and direct a set of cyberinfrastructure initiatives for archaeology. It seeks to coordinate with and develop interoperability of its own projects with other relevant data-sharing initiatives and offers to work with professional organizations and federal agencies to promote policies that will foster the development of cyberinfrastructure for archaeology.

Shortly after its formation, on March 20, 2007, the Consortium submitted a planning grant request to the Mellon Foundation.

In that document (pg 4) we proposed to:

... take critical, positive steps toward the development of an effective cyberinfrastructure for archaeology. Recognizing that it is an ambitious vision, we argue that the cyberinfrastructure should provide concept-oriented, integrated research access to the three types of data sources that we consider to be the most pressing with respect to the needs of archaeological research: documents, databases, and images.

We propose to develop a collaborative organization, *archaeoinformatics.org*, with a structure designed to foster and guide the development of this cyberinfrastructure. We further propose to accomplish a review of cyberinfrastructure initiatives within and outside archaeology. Following that review, we will develop an implementation plan that addresses functional and technical considerations as well as its financial viability and the sociological issues associated with engaging different constituencies within archaeology. We will also develop a financial model for the long-term viability of this cyberinfrastructure.

Success of the infrastructure will depend, in part, on maintaining active disciplinary interest. We need to avoid widely publicized promises whose rewards are only seen many years in the future. To this end, we will design the project so as to incrementally release tools with immediate research value to a substantial body of archaeological practitioners. We will also propose substantial investment in the development of substantive case studies. They will be designed to provide a range of constituencies with a compelling demonstration of the research value of the infrastructure.

## Organizational Plan

The initial structure of Archaeoinformatics included a **Steering Committee** advised by a Disciplinary Advisory Board and a Technical Advisory Board. However, we recognized that the structure of this organization should evolve in response to changing funding sources and policy environments designed to foster and guide the development of this cyberinfrastructure. In response to recommendations at the Joint Meeting in February 2008 in Santa Fe (see below) the Disciplinary and Technical Advisory Boards were reorganized into a single **Board of Directors**.

### Steering Committee

**Purpose.** As the decision-making body of archaeoinformatics.org, the Steering Committee manages the design, seeks funding for, and directs the development of cyberinfrastructure initiatives in archaeology. It administers central functions necessary for the development and operation of the cyberinfrastructure. It seeks to work with professional organizations to influence relevant policies and to engage the larger community of archaeologists in these efforts.

**Membership and Governance.** The Steering Committee is initially comprised of the five institutional submitters of the planning proposal, each represented by a single individual. An institution may be added or removed from the Steering Committee with the majority agreement of the members. Members may also be added based on the recommendations of the Board of Directors. The chair will rotate annually through all Committee members unless a chair is otherwise unanimously agreed upon by the members. The steering committee included:

**Keith Kintigh**, Arizona State University (Chair)

**Jeffrey Altschul**, SRI Foundation

**Tim Kohler**, Washington State University

**Fred Limp**, University of Arkansas

**Julian Richards**, University of York (joined in February 2008)

**Dean Snow**, The Pennsylvania State University

### Board of Directors

**Purpose.** The Board of Directors provides recommendations to the Steering Committee based on its evaluation of the plans for and review of the substantive results of the initiatives pursued by archaeoinformatics.org to insure that they are effectively addressing the needs of the profession. The Board reviews and evaluates the technical design and technical results of the cyberinfrastructure effort to insure effective development and appropriate interoperability with other science infrastructure efforts. It provides advice on the information engineering considerations of ontology development. It reviews technical staffing requests and provides recommendations to the Steering Committee. At a minimum, the Board meets annually and produces a brief report after each meeting.

## Board of Directors 2007-2008 Membership

**Brian Crane** Versar, Inc.  
**Katherine (Kitty) Emery** Florida Museum of Natural History of Florida, Gainesville  
**Sebastian Heath** Archaeological Institute of America/American Numismatics Society  
**Eric Kansa** University of California, Berkeley/Alexandria Archive Institute, Berkeley  
**Francis (Frank) McManamon** National Park Service  
**Fraser Neiman** Thomas Jefferson Foundation  
**Vincas (Vin) Steponaitis** University of North Carolina  
**Willeke Wendrich** UCLA  
**Phillip (Phil) Walker** University of California, Santa Barbara  
**Thomas (Tom) Whitley** Brockington & Associates, Inc.  
**Worthy Martin** University of Virginia  
**Herbert Van de Sompel** Los Alamos National Laboratory

More details on the membership are provided in Appendix I.

## Information Development and Dissemination

A central aspect of the planning efforts was to acquire information on the state of the field and on best-practices in the cyberinfrastructure community. As described in the proposal (pg. 14):

The implementation plan for the infrastructure will be informed by an investigation and evaluation of other cyberinfrastructure efforts in archaeology and allied sciences, both in the US and abroad. Through this investigation we expect to be able to identify existing codebases that could be adopted or adapted, to better understand the capabilities of other systems with which we would hope to collaborate or to provide interoperability, and to benefit more generally from their development experiences.

Three partially overlapping components will be investigated: software frameworks; science (including archaeology) and heritage management informatics efforts; and metadata standards initiatives.

A series of presentations were commissioned from leaders in both archaeology and the broader IT community. These lectures were offered via the NSF-sponsored Access GRID video conferencing and cyber-collaboration tools. Ten (10) lectures were presented. They were:

1. Eric Kansa, "Open Context: Community Tools for Publishing Research Data on the Web." Executive Director of the [Alexandria Archive Institute](#) and University of California Berkley.
2. Chaitan Baru "GEON: Geosciences Network. " Director of Science Research and Development at the San Diego Supercomputer Center.
3. Michael J. Halm "LionShare: Secure P2P File Sharing and Collaboration." Senior Strategist and Manager for the Special Project activities for the Teaching and Learning with Technology group, Penn State University.
4. Mark Gahegan "Sharing our resources, sharing our understanding: Cyberinfrastructure for Archaeology." Professor of geography and affiliate professor of information science and technology at the Pennsylvania State University.

5. Fred Limp "Interoperability and net-centric architectures: lessons for archaeoinformatics from the Open Geospatial Consortium." Open Geospatial Interoperability Institute and University of Arkansas.
6. Mark Schildhauer "Ecological informatics: challenges and approaches, and potential relevance for archaeology." National Center for Ecological Analysis and Synthesis, Santa Barbara.
7. Julian D Richards "Current challenges for digital preservation and delivery." Professor of Archaeology, University of York and Director, Archaeology Data Service.
8. Ian Johnson "Electronic Cultural Atlas Initiative: The snowball still survives." Archaeological Computing Laboratory, University of Sydney.
9. Katherine Skinner "Collaborative Adventures in Distributed Digital Preservation: The MetaArchive Cooperative and the Educopia Institute. " Digital Projects Librarian at the Emory University Libraries.
10. John Howard "Fedora Digital Repository for Long-Lived Data Collections." Arizona State University.

Additional details on the virtual lecture series is provided in Appendix II. The lectures themselves (as PDF and streaming video and pod-casts) are available on line at <http://archaeoinformatics.org>

## Web site

An extensive web site was created to host the background materials, presentations and to serve as a portal to the initiative. The site is [http://archaeoinformatics.org/lecture\\_series.html](http://archaeoinformatics.org/lecture_series.html) and is hosted by the University of Arkansas.

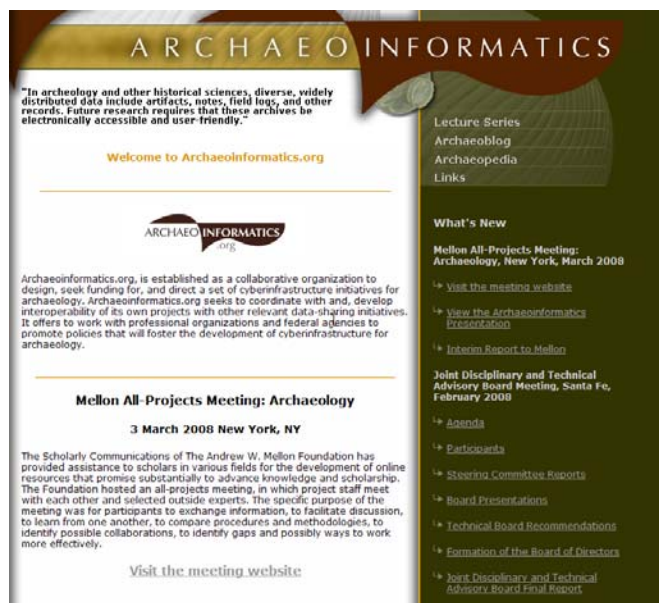


Figure 1. Archaeoinformatics.org home web page



The web site contains information on all aspects of the initiative, including the lectures, background materials, various reports, the results of all meetings and links to other related sites.

## Survey of profession

A web-based survey was conducted to assess the current circumstances in the field, the levels of interest and key concerns. There were 270 responses, primarily from members of the SAA's Digital Data Interest Group. Key findings were:

- 94% responded that documentation of the archaeological record is being lost.
- 94% responded that they would use electronic data more if it were accessible.
- 90% responded that it is the responsibility of a project sponsor to fund and ensure curation of databases.
- More than 60% responded that users should not be charged access fees.

### Archaeoinformatics Survey

The screenshot displays a web-based survey titled "Archaeoinformatics Survey". The survey is presented on a green background with a circuit-like pattern. The first question, "Please tell us a little about yourself.", is followed by six numbered questions:

- 1 What is your highest degree?  
☐ Bachelors  
☐ Masters  
☐ Doctorate
- 2 What is your age?
- 3 How many years have you spent in archaeology?
- 4 What is your sex?  
☐ Female  
☐ Male
- 5 In which country do you live?
- 6 What is your primary employment setting?  
☐ Academic Department  
☐ Academic Museum or Research Center  
☐ Government, State or Local  
☐ Government, Federal or Tribal  
☐ Museum or Research Center (not academic)  
☐ Private For-Profit

Figure 2. Portion of the first page of the on-line survey.

More results of the survey are presented in Appendix III.

## **Santa Fe Meetings**

A three day meeting was held in Santa Fe NM (Feb 15-18, 2008), bringing together the Steering Committee, the Board of Directors and outside experts. The meeting started with a series of presentations made covering the background to the project, current initiatives and other related projects. The presentation's PowerPoint slides are available on the Archaeoinformatics.org web site at [http://archaeoinformatics.org/Board\\_Present.html](http://archaeoinformatics.org/Board_Present.html)

The Steering Committee, led by **Keith Kintigh**, reviewed the progress of the consortium to date, and the objectives for the meeting. This was followed by a series of presentations by Board members. The presenters and their topics follow. More details on the presentations are provided in Appendix IV.

**Phil Walker** –in-field recording and repatriation

**Julian Richards** –the core elements of the Archaeology Data Service effort

**Eric Kansa** –development of software solutions to some the core sociological problems

**Fraser Neiman** –key elements to achieving archaeological community buy-in to an information infrastructure

**Worthy Martin** – staging the implementation of a cyberinfrastructure and an evaluation of the AIO vision

**Herbert Van de Sompel** – the opportunities and advantages vs. the costs and disadvantages of achieving interoperability with other science informatics and digital library efforts.

**Frank McManamon** –the two or three most important values of an integrated text and data information infrastructure and its relationship to governmental initiatives

**Vin Steponaitis** – the most important values a cyberinfrastructure could bring to the academic research community

**Brian Crane** – how would their integration and interoperability with major datasets not from DOD lands benefit the larger DOD mission

**Tom Whitley** –the two or three most important values of a cyberinfrastructure from the standpoint of the practice of CRM

**Willeke Wendrich** –how large-scale, long-term, international multi-investigator research poses particular challenges and opportunities

**Kitty Emery** –the values of such an infrastructure to the missions of anthropology or natural history museums.

The participants then broke into a series of smaller working groups, followed by a meeting as the committee-of-the-whole. The Technical and Disciplinary groups prepared a joint report which was initially distributed at the meeting and then underwent revisions culminating in a final consensus document on March 1, 2008. The report is presented here in full.

### **Santa Fe Joint Disciplinary and Technical Advisory Board Report**

#### **Initial version - 17 Feb 2008**

The initial meeting of the Archaeoinformatics.org Disciplinary and Technical Advisory Boards with the Steering Committee was held from February 15-17, 2008, at the Hotel Santa Fe, in Santa Fe, New Mexico. (One Board member, Sebastian Heath, was unable to attend due to a scheduling conflict.)

The advisory boards had been supplied with brief written reports on archaeoinformatics.org activities. The meeting began with presentations by Steering Committee and Advisory Board members, accompanied by considerable discussion. That full-group session was followed by discussion on targeted topics pursued by three breakout groups, each with Advisory Board and Steering Committee members. The breakout groups then reported back to the full group.

Following these presentations, the Disciplinary and Technical Advisory Boards met jointly, without the steering committee present, in order to evaluate the project's work to date. It seemed appropriate for us to prepare a joint, rather than separate board reports. We found the Steering Committee receptive to our suggestions and we expect that these suggestions will substantially inform the final plan.

We provide this report as a statement by the Boards of important issues confronting the discipline and an evaluation of the planning efforts to date pursued under the Digital Antiquity grant from the Andrew W. Mellon Foundation. Our comments are generally not at variance with the goals and activities and plans of the Steering Committee. Rather, they are intended to highlight points that we consider most important.

#### 1. Understanding of Disciplinary Needs

- General disciplinary needs have been well defined, as have the payoffs of a potential information infrastructure. The boards concur that major disciplinary problems are: the inaccessibility of data and gray literature; the potential and ongoing loss of data; and the inability to perform important new and innovative research.
- It would be useful to more specifically articulate how the more general needs and benefits would apply to particular classes of stakeholders. Stakeholder groups that should be separately considered include academics, CRM companies, public agencies, museums, and the general public.
- As development is planned, the needs should be prioritized so that initial demonstrations will be seen to be responsive to widely perceived needs and stakeholder interests.

#### 2. Scope of Archaeoinformatics.org Mission and Objectives

- The scope is properly focused on archaeology (including prehistoric and historic archaeology and bioarchaeology) with appropriate connections to key constituencies including cultural resource management (CRM) and academic practitioners. It would be unwise at this time to extend the scope to include, for example, standing historic structures. However, the possible articulation with other databases of information about cultural resources, for example, historic documents, historical analytical databases, historical architectural structures, museum collections, traditional cultural properties, and cultural landscapes should be recognized.
- The scope needs to be more clearly specified and integrated with an implementation timeline.
- The objectives should be addressed to specific stakeholder audiences taking into consideration what each would consider as its priorities. These might include CRM archaeologists, academics working in US and international contexts, public agency archaeologists and State Historic Preservation Officers (SHPOs), and educators.
- A key component of the plan must be community building. Developing the social components will be just as critical as developing the technical capabilities.

- Development of a widely accepted body of shared principles and practices can lead to much-needed transformations of agency practice and disciplinary culture in ways that support data accessibility, interoperability, and preservation.
- Sustainability—financial, technological and social—of an archaeological information infrastructure must be emphasized.

### 3. Evaluation of Effort to Date

- Administration: The current structure has worked well to date.
- Advisory Boards: Having reached this point, we recommend some reorganization as the initiative moves forward:
  - The two advisory boards and steering committee should be merged into a single board of directors.
  - The board should be led by an executive subcommittee, which initially should consist of the current steering committee members.
  - Some members should continue to be nominated by professional societies; others with special expertise should be nominated by board members.
  - There is a need for representation of SHPOs and current data repository managers familiar with site-file information management and access.
  - The diverse stakeholder constituencies of the infrastructure should continue to be considered when reviewing and making any changes to disciplinary advisory board composition.
- Financial Model: The possible sources have been well identified. Of course, more concrete plans for tapping available revenue sources will need to be developed.
- Lecture Series:
  - The virtual lecture series is a good idea and has been executed with strong content. Although the lectures are archived and accessible, the series needs better publicity. It might be good to offer downloads as podcasts.
  - As the initiative moves forward and tools become available, this lecture series could move more into a widely promoted training mode, including “slide share” documents.
- Survey: The survey was useful as far as it went.
  - In the future, more formal marketing research is needed, including expected yield-for-service (identification and quantification of potential income from stake holders who are willing to pay for specific services).
  - More insight is needed into community perspectives and needs.
- Technology:
  - The Steering Committee has identified a number of important attributes of a desirable archaeological cyberinfrastructure.

- In the course of the meeting, a consensus has developed on appropriate Level 1 and Level 2 (initial and subsequent stage) functionality of cyberinfrastructure services.
  - Some key technological “building blocks” have been identified and the Steering Committee recognizes that more specific technological specifications need to be developed.
  - Some outside tools have been identified. While they need to be tested, appropriate adoption of open source software applications supported by broad user communities will improve sustainability.
  - Some experimentation with the software tools will be needed to identify optimal configurations.
- Case-Study Development:
    - The current examples presented are useful, but are closely associated with Steering Committee members (leading to potential sociological liabilities that may need to be mitigated).
    - Other case studies are needed that will illustrate the broad scope and large scale of the initiative and its relevance to diverse audiences in the US and internationally.
    - For each case study, the purpose needs to be clearly justified in terms of what is to be learned and how it can be used in the overall promotion of the infrastructure.
  - Prototype Software:
    - Both tDAR and ArchSeer are useful in illustrating and providing specific functionality. Additional models (including ADS and OpenContext) will be useful in both community building and in demonstrating semantic interoperability enabled by the Archaeoinformatics cyberinfrastructure.

4. Overall Progress: EXCELLENT.

5. Additional Recommendations:

- The initiative will require active outreach and training programs. This relates to the earlier point regarding the need for community building as well as the need for specific training in the use of data access tools, identification of metadata, and other technical skills.
- An informal survey to ascertain the current state of affairs and needs of existing data repositories and collections would be informative.
- Archaeoinformatics.org should establish liaisons with other relevant efforts (e.g., US FGDC and international standards for documenting heritage properties).
- There should be support for the notion of a spectrum or levels of interoperability. For example, one can designate certain fields to be required, and others to be optional with varying degrees of importance. This approach will maximize the utility of datasets without unduly compromising interoperability.

The document (as presented above) was finalized 1 March 2008 as based on board comments on a circulated draft.

## **Subcontracts**

Subcontracts were initially issued to Arizona State University, Penn State University, Washington State University, and Statistical Research, Inc.

In addition to their participation in all the general activities, these sub-contracts involved specific tasks. For ASU and WSU this involved conducting pilot case studies that would provide critically needed practical information on issues of cyberinfrastructure implementation and execution. SRI was to focus on fiscal and organizational efforts and PSU was funded to extend earlier work on a development of full-text, map and text search capabilities that would increase the accessibility of documents and reports. As the project moved along it became clear that the participation of University of York was essential. York hosts the Archaeology Data Service and ADS is an exemplar of many objectives of the AIO initiative. In addition ADS has close links to the Europe efforts in archaeological e-science and would serve as a bridge to these. ASU indicated that they were able to release \$3,000 of their initial award and a subcontract was signed with York. These funds were to be used to cover expenses associated with York's AIO collaboration.

The selected case studies planned for ASU and WSU were described in the proposal (page 19) as follows:

The implementation proposal will make recommendations with respect to substantive case studies to be developed. At the implementation stage, we envision a small number of case studies focused on important substantive topics that require the synthesis of information from diverse data sources. Each case study would be developed by a group of scholars with an interest in and expertise on the topic. The group would help define the questions to be addressed, identify the crucial documents and datasets, and facilitate their registration in the system. The group would then use the tools and data sources provided by the proposed infrastructure to conduct innovative research on the target topic. The case studies would result in:

- a demonstration for archaeologists of the research value of the infrastructure;
- new scholarly research that would have been impractical without the infrastructure;
- feedback on the capabilities of the software tools leading to their refinement; and
- on-line availability, through the infrastructure, of substantial bodies of data and documents relevant to significant topics and geographical areas that will have continuing research value to a range of scholars.

Criteria for the selection of case studies based on our discussions to date are:

- 1) they must be driven by a set of compelling research questions;
- 2) multi-institutional cooperatives are encouraged;
- 3) at least one should be international;
- 4) at least one should contain a large component of legacy data; and
- 5) at least one should have a large component of recent data with ongoing contributions by cultural resource management firms working under federal mandate.

The current [planning] proposal will take initial steps toward the development of two such case studies. ASU's experience in its NSF grants shows that acquiring the relevant datasets and the associated metadata can be quite time consuming. In this proposal we plan to jump-start two of these case studies with the goal that, by the time that the implementation project (the subsequent grant) is ready for presentation to the archaeological community, two powerful cases will be available to showcase the infrastructure's potential. The Basin of Mexico case study was identified as highly desirable both at the NSF-sponsored Santa Barbara cyberinfrastructure workshop (Kintigh 2006) and the meeting of the planning grant principal investigators with the Mellon Foundation staff. It satisfies criteria 1-4.

The second case study will be focused on the Four Corner's region of the American Southwest, an area of with some of the highest precision archaeological data in the world. This case study satisfies criteria 1, 2, 4, and 5. Initial contacts with curators of candidate datasets from southwestern Colorado and southeastern Utah have already been made. The two cases provide a useful contrast in that the first focuses on complex societies (Teotihuacan and Aztec) the second on villagers (Ancestral Pueblos).

### **Arizona State University Case Study Results**

As described in their sub-contract report (provided in full in Appendix VI), the ASU sub-contracted efforts in this case-study

... focused on laying the foundation for a digital archive strength in the archaeology of central Mexico (in and around the Basin of Mexico, including Classic Period Teotihuacan and Aztec Tenochtitlan). This involved introducing tDAR to US, Mexican, and other scholars working in this area. To the extent possible, once Prototype tDAR was operational, we assisted these scholars in acquiring and uploading datasets. Funded efforts were pursued by young independent scholars Michelle Elliott and Vince Schiavetti both of whom received PhDs under Professor Nelson. These individuals worked on the grant under contract and were supervised by Nelson with the assistance of Kintigh.

Because Michelle Elliott is based at Universidad Nacional Autónoma de México (UNAM) in Mexico City, she was in an excellent position to promote the proposed archive. She visited Tempe and was introduced to the project and worked closely with Ben Nelson and Keith Kintigh via videoconference and email. Elliott was able to present the Digital Antiquity project to numerous scholars (Mexican, US, and French) who lived in or passed through Mexico City during the grant period. As a part of this effort, she created a useful PowerPoint tutorial to introduce the key concepts ([http://tdar.org/confluence/download/attachments/131075/tDAR\\_Instructions.pdf?version=1](http://tdar.org/confluence/download/attachments/131075/tDAR_Instructions.pdf?version=1)).

Elliott was able to discuss the Digital Antiquity initiative with a number of important scholars. Through Elliott's efforts we were able to incorporate into Prototype tDAR Jeffrey Parsons' (University of Michigan, emeritus) and his colleagues' database for their systematic survey of the Valley of Mexico Survey. Because of its focus on an area of enormous archeological significance and because of the author's willingness to share these data, the database has been widely reused. Its availability in Prototype tDAR (and tDAR) will further enhance its accessibility and usability for current and future scholars.

Elliott was also able to assist Mexican senior scholar Emily McClung de Tapia (UNAM) in registering information from a number of her completed projects. Additional scholars agreeing to contribute data include Tom Charlton of the University of Iowa, George Cowgill of Arizona State University, and Ian Robertson of Stanford University.

As anticipated, there were delays associated with contacting scholars, with obtaining their agreement to contribute data, and in preparing their projects for ingest. Additional delays were introduced by the fact that Prototype tDAR was not fully operational as early as anticipated, and because of Mexican scholars' need to consult with INAH on the appropriateness of participation. As a result, fewer datasets were registered in tDAR than we had hoped. However the degree of general support was somewhat greater than we anticipated.

We are quite encouraged by the results these efforts, especially in light of the fact that Mexican scholars have a somewhat different culture of sharing data than US scholars. We believe that we have met the major subcontract goal of laying the groundwork for international participation in tDAR, once the production version of tDAR is launched under the implementation grant.

During the grant period, the opportunity arose to incorporate in tDAR a major systematic database and associated documents for the site of La Quemada, Zacatecas. Because it seemed that this would be a very important resource for the archive, because it represented an excellent example of a senior US scholar (Ben Nelson) sharing a very large corpus of data from his long-term research, and because there appeared to be sufficient funds, we added working toward the incorporation of this database to the original scope of work.

Vince Schiavetti worked to incorporate Nelson's La Quemada database and associated documents into Prototype tDAR. While we have not yet completed this effort, we have already made important documents and data available and work on this project continues. Challenges presented by this large (300MB+) and complex dataset have provided valuable user testing and expert feedback and has substantially informed our planning for the tDAR archive.

Finally, we engaged in considerable discussion concerning the development of a series of international conferences directed to synthetic work on an important topic. In this model, selected scholars would pursue synthetic research using the corpus of participants' contributed data shared through tDAR. One actively discussed possibility for such a conference would have to do with the regional reorganization of central Mexico that occurred after the collapse of Teotihuacan. While firm plans must await further development of the production archive and better data integration capabilities (in progress), we have made considerable progress in the development of what we believe could be a very persuasive case study in the research value of the archive for which we can seek independent funding.

## **Washington State University Case Study Results**

As described by Kohler in his sub-contract report (provided in full as Appendix VII):

Our accomplishments with this funding were closely in line with our proposal, although we have completed most of the work on not one dataset, as proposed, but on two. Specifically, we have put on line, locally stored in a PostgreSQL database (an open-source database), and available for eventual incorporation into the next generation of the tDAR software, 25 datasets of the Dolores Archaeological Project (DAP; accessible at <http://galisteo.anth.wsu.edu/dap/>), and 11 datasets of the Bandelier Archaeological Excavation Project (BAEP; accessible at <http://galisteo.anth.wsu.edu/BAEP/>). With minor exceptions set forth in appendices 2 and 3, all



the labeling (data dictionaries) associated with these databases is complete and also accessible on-line. We chose the BAEP dataset instead of one of the other candidate datasets listed in our proposal because

- (1) the Cedar Mesa dataset was (and still is at this writing) incomplete, and
- (2) the BAEP datasets are much more at risk of becoming inaccessible than are the datasets of the “Village” project, for example.

One area in which we did not complete the goals of the proposal is that the metadata for these datasets, and the datasets themselves, are not yet entered into tDAR or its successor, although we believe we have collected the necessary metadata for this to take place in the future, and the databases themselves are ready to go. The necessary facilities for ingesting these data and metadata into tDAR were only becoming available, in trial form, as our funding was exhausted and the grant period was ending. It remains our hope that these two databases will be joined by others in the next phase of development of the digital archive, and that together these resources will serve as an attraction to other regional datasets.

### **Penn State University ArchSeer Results**

In addition to the participation of Dean Snow in all the various AIO activities, Penn State also participated in a pilot project. As described in the proposal (page 20):

PSU will pursue a pilot project involving intelligent access to documents and images. This is expected to include an application to the full text of *American Antiquity* and selected other journals (with access arranged through JSTOR).

The PSU team met weekly to supervise the four students that worked on the project. The results are an integrated search engine for archaeology that can search text, citations, maps, tables, locations and time. For the prototype data was drawn from 8,000 documents from JSTOR archaeology journals. The effort leverages other open source projects, specifically

- the Lucene indexer
- JSTOR metadata used for metadata extraction and indexing
- ChemXSeer, chemistry, table extraction and indexing

The prototype uses aspects of CiteSeerX ingestion, indexing and crawling. It supports table search and data extraction, extracting data from tables in an XML OAI format that can then be used in other experiments or data aggregation. The effort will also provide open source extraction tools for other initiatives. The results to date are accessible at <http://cxs02.ist.psu.edu:8080/archseer>.

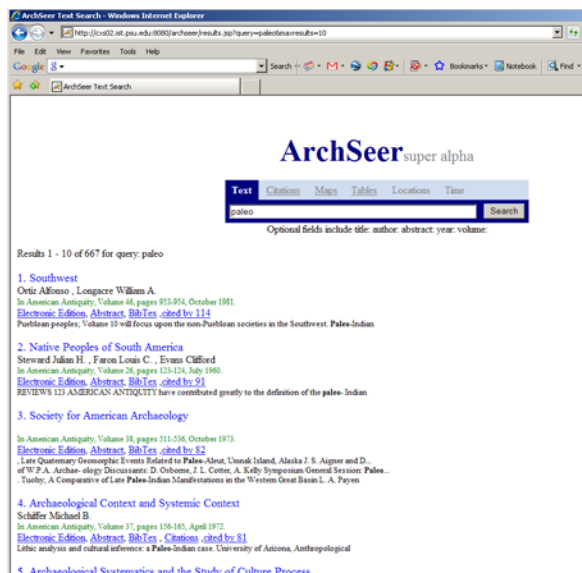


Figure 3. ArchSeer search text results

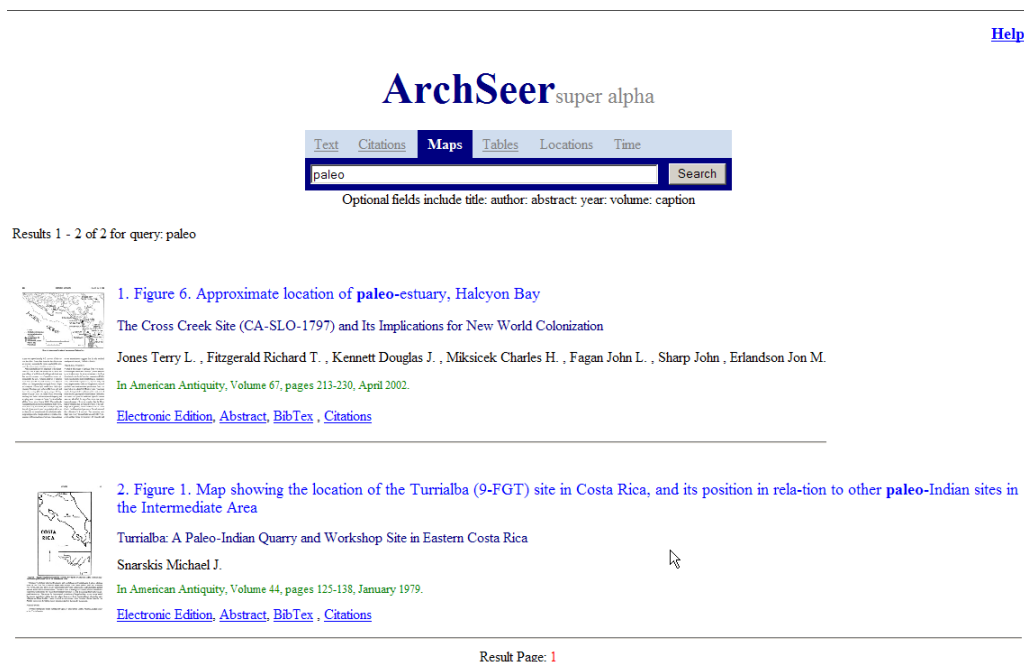


Figure 4. ArchSeer Map search results

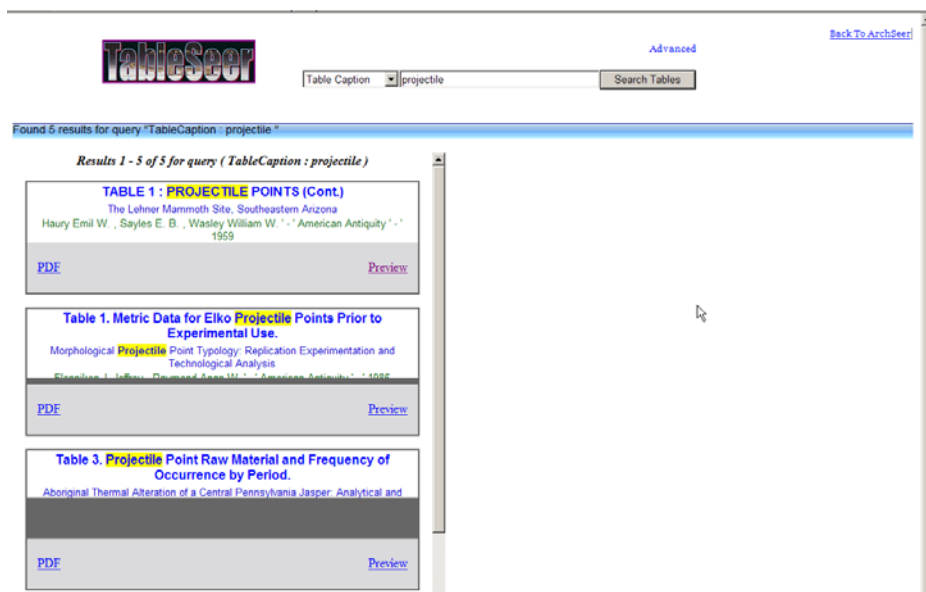


Figure 5. TableSeer search results

## Statistical Research Inc. Results

As described in the proposal (page 20):

SRI will develop a formal specification of the functional requirements necessary for the effective adoption of the cyberinfrastructure by the applied archaeological community. It will spearhead development of a *pro forma* financial model for the long-term sustainability of the cyberinfrastructure. It will perform a financial assessment of the potential efficiencies and cost savings that could accrue to contractors, regulators, and artifact repositories. SRI's investigation will also articulate the ways in which the infrastructure can lead to improved program performance with respect to legal mandates.

SRI completed all these tasks with the results integrated in the business plan elements of the final proposal. During the project time staff changes at SRI resulted in Jeff Altschul, SRI founder and Chairman, joining the AIO steering committee. Whereas previously, SRI's contribution were broad based (i.e., survey of the discipline and the development of a business plan based largely on economic assumptions), Dr. Altschul inclusion allowed the initiative to focus on specific CRM sectors. Convinced that the only viable means of success for the initiative was the implementation of a mandate to place digital material gathered in compliance with historic preservation law in a trusted repository, Dr. Altschul together with Dr. Kintigh and Dr. Frank McManamon presented the problem of digital curation and the possible solution via archaeoinformatics to the Chief Archaeologists and Federal Preservation Officers of Federal agencies. Drs. Altschul and Kintigh also engaged the CRM industry in the discussion by making a similar presentation to the board of directors of the American Cultural Resources Association.

## Fiscal status

Individual and summary fiscal reports are provided in the appendices. As documented in the appendices the funds allocated were expended consistent with the proposed tasks. The total project budget was composed of the \$152,000 award and the \$3,861.09 interest for a total of \$155,861.09. For the University of Arkansas distribution a positive balance of \$398.09 remains after all expenditures were paid. Within the University of Arkansas sub-budget revenue from interest income and savings from the projected consultants travel costs were reallocated to staff salaries when a graduate student who was managing the web site left for another position and had to be replaced with (a fraction of) a full-time salaried staff employee. Minor (positive) variances also exist between budgeted and allocated subcontracts for WSU (\$992.73), PSU (\$173.51) and Arizona State (\$679.66). One substantial deviation between the proposed and expended funds is in the SRI budget. Because of changes in SRI staff during the project, Dr. Clay Mathers was replaced by Dr. Jeff Altschul. Dr. Altschul is the founder and CEO of SRI. Due to the fiscal structure of SRI, Dr. Altschul's salary cannot be charged and as a result only \$8,289.41 of the total \$21,801.00 requested by SRI could be expended, leaving an unexpended balance of \$13,511.59.

The total aggregated unexpended amount (\$15,755.58) is being held by the University of Arkansas to be returned to the Mellon Foundation or otherwise distributed as directed.

## Conclusion

The efforts of the Archaeoinformatics Consortium involved an extensive suite of actions that included the formulation of a national organization with a well respected Board of Directors, development of an extensive web site, hosting of an extensive on-line lecture series and completion of case studies and business models and technical studies. These all served as key foundational elements to the effective creation of a cyberinfrastructure for archaeology in the US. Most significantly, the effort has lead to the preparation of a proposal ***Digital Antiquity : enabling and enhancing preservation of an access to archaeological information.***

## References Cited

Kintigh, Keith W. (ed.)

2006 The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71:567-578.

Snow, Dean R., M. Gahegan, C. Lee Giles, Kenneth G. Hirth, George R. Milner, P. Mitra, and J. Z. Wang

2006 Cybertools and Archaeology. *Science* 311:958-959.

Waters, Don

2007 Doing much more than we have so far attempted. *EDUCAUSE* September/October:8-9.

# Appendix I

## Board Structure and Membership

### Membership

**Brian Crane** Senior Archaeologist. Versar, Inc., Springfield, VA

Brian Crane (Ph.D.) is Senior Archaeologist with Versar Inc. Versar is a global project management company delivering construction management, environmental sciences and engineering infrastructure solutions to public and private customers. Versar provides extensive cultural resources support to many units of the Department of Defense.

**Katherine (Kitty) Emery** Assistant Curator, Environmental Archaeology, Florida Museum of Natural History UF Assistant Professor, Anthropology, University of Florida, Gainesville

Kitty F. Emery, Ph.D., is Assistant Curator of Environmental Archaeology at the Florida Museum of Natural History. She uses ancient animal and plant remains recovered from archaeological deposits in Central America to understand how the ancient Maya used (and sometimes abused) their natural world. Recent research has included compiling regional evidence of ancient Maya hunting and forest management as reflected in archaeological animal remains, and modern evidence for animal bone medicinal and ritual curation, and organic material discard, both with implications for animal remain recovery in the Maya region.

**Sebastian Heath** Archaeological Institute of America/American Numismatics Society, New York

**Eric Kansa** University of California, Berkeley/Alexandria Archive Institute, Berkeley/San Francisco

**Francis (Frank) McManamon** National Park Service, Washington DC

**Fraser Neiman** Director of Archaeology, Monticello, Virginia and Lecturer, Departments of Anthropology and Architectural History, University of Virginia.

Fraser Neiman directs ongoing archaeological research at Monticello ([www.monticello.org](http://www.monticello.org) <<http://www.monticello.org/>> ) into the ecological and social dynamics of the early-modern Chesapeake and the larger Atlantic world of which it was a part. His lab is home to the Digital Archaeological Archive of Comparative Slavery ([www.daacs.org](http://www.daacs.org) <<http://www.daacs.org/>> ). DAACS is an experiment in the use of IT and the internet to share detailed archaeological data, encourage comparative analysis, leverage collaboration, and accelerate progress in understanding the evolution of slave societies of the Chesapeake, Carolinas, and the Caribbean. He teaches courses in quantitative methods, historical archaeology, and archaeological theory at the University of Virginia ([www.people.virginia.edu/~fn9r](http://www.people.virginia.edu/~fn9r)).

**Vincas (Vin) Steponaitis** Director, Research Laboratories of Archaeology and Professor, Department of Anthropology , University of North Carolina

In collaboration with Steve Davis and others, Steponaitis worked on a variety of projects exploring the frontiers of digital publication in archaeology. The most elaborate and visible of these projects is a CD-ROM entitled *Excavating Occaneechi Town*, published by UNC Press in 1998. They have also developed a new edition of this work that will be formally published by UNC Press on the World Wide Web. A "beta" version of this new edition can be found at [www.ibiblio.org/dig](http://www.ibiblio.org/dig).

The history and political economy of Moundville, a large Mississippian town in Alabama that was occupied from the 11th to the 17th centuries AD, have long been subjects of Steponaitis' research. His current research attempts to reconstruct patterns of craft production and trade by attempting to identify the geological sources of the raw materials used to make "prestige goods" at Moundville.

**Willeke Wendrich** Associate Professor Egyptian Archaeology, Dept. of NELC, and Cotsen Institute of Archaeology at UCLA

Willeke Wendrich leads a large team of archaeologists and archaeological specialist in the study of land and water use, and the development of agriculture in the Fayum (Egypt). The research program focuses on the prehistoric and Greco-Roman remains in the region, and the work includes rescue excavations and cultural heritage management of the ancient landscape. As the Editor-in-chief of the online UCLA Encyclopedia of Egyptology and Faculty Director of the UCLA Digital Humanities Incubator Group, Wendrich is closely involved in digital publication and archaeological data management.

**Phillip (Phil) Walker** University of California, Santa Barbara

**Thomas (Tom) Whitley** Vice President of Brockington & Associates, Inc. , Norcross, GA

Dr. Whitley has specialized in GIS applications to Archaeology since the late 1980s. This has included small, medium, and large scale spatial analyses and predictive models in 17 different states. He is Vice President of Brockington and Associates, Inc., one of the largest CRM firms in the Southeast, where he manages the Atlanta Office. His current research focus is on complex GIS and 3D modeling of cognitive and other interpretive prehistoric/historic landscapes.

**Worthy Martin** Associate Professor and Associate Chair of Computer Science, University of Virginia  
Associate Director of the Institute for Advanced Technology in the Humanities, University of Virginia

Worthy Martin joined the Department of Computer Science at the University of Virginia in 1982 after completing his Ph.D. degree in Computer Science at the University of Texas, Austin, specializing in computer vision. He has published papers in dynamic scene analysis, visual acquisition of volumetric models, combinatorial optimization with evolutionary algorithms and visual control in robotic systems. His interests in image analysis and complex data structures lead to a research appointment with the Institute for Advanced Technology in the Humanities (IATH) in 2000.

Through IATH he has participated in numerous digital humanities projects, including The Monastery Plan of St. Gall project, The Chaco Digital Initiative, The Sustaining Digital Scholarship project, The Salem Witch Trial Archive project, and The Uncle Tom's Cabin & American Culture project (see <http://www.iath.virginia.edu/>). The common thread to these projects is the design and implementation of thematic repositories with rich interconnections among many levels of information components within the repositories.

**Herbert Van de Sompel** Los Alamos National Laboratory, Los Alamos, NM Team Leader, Digital Library Research & Prototyping Team Los Alamos National Laboratory, Research Library

Herbert Van de Sompel graduated in Mathematics and Computer Science at Ghent University, and in 2000, obtained a Ph.D. there. For many years, he was Head of Library Automation at Ghent University. After having left Ghent in 2000, he has been Visiting Professor in Computer Science at Cornell University, and Director of e-Strategy and Programmes at the British Library. Currently, he is the team leader of the Digital Library Research and Prototyping Team at the Research Library of the Los Alamos National Laboratory. The Team does research regarding various aspects of scholarly communication in the digital age, including information infrastructure, interoperability, digital preservation and indicators for the assessment of the quality of units of scholarly communication. Herbert has played a major role in creating the Open Archives Protocol for Metadata Harvesting, the OpenURL Framework for Context-Sensitive Services, the SFX linking server, and the info URI. With Carl Lagoze, he is currently leading the Open Archives Initiative Object Re-Use and Exchange effort.



## Appendix II

### Lectures and lecture archive

The Power Points from all the lectures as well as streaming video and podcast versions of the actual lecture are available at the [archaeoinformatics.org](http://archaeoinformatics.org) web site.

#### **March 26, 2007**

Executive Council members hosted a Town Hall Meeting with an introduction to the activities of Archaeoinformatics.org.

---

#### **April 9, 2007**

Eric C. Kansa

##### **"Open Context: Community Tools for Publishing Research Data on the Web"**

Eric Kansa, Executive Director of the Alexandria Archive Institute discussed the archaeological project OpenContext - an ArchaeoML based system for sharing diverse, nonstandardized data and media.

---

#### **April 23, 2007**

Chaitan Baru

##### **"GEON: Geosciences Network"**

Chaitan Baru, Director of Science Research and Development at the San Diego Supercomputer Center, spoke on GEON , geology's successful analog to some of what we believe archaeology needs to accomplish.

---

#### **September 19, 2007**

Michael J. Halm, John Yoo

##### **"LionShare: Secure P2P File Sharing and Collaboration"**

Michael J. Halm, a Senior Strategist and Manager for the Special Project activities for the Teaching and

Learning with Technology group, Penn State University, spoke about the LionShare project and its dedication to harnessing the promise of peer-to-peer (P2P) file-sharing and the integration of P2P with organizational services to create a collaborative environment for use in academic communities.

The LionShare project, which Mike will discuss in this lecture session, is dedicated to harnessing the promise of peer-to-peer (P2P) file-sharing and the integration of P2P with organizational services to create a collaborative environment for use in academic communities. The LionShare Peer application is built around the themes of collaboration, security, personal responsibility, and access control of shared resources, along with access to large digital repositories.

---

### **October 17, 2007**

Mark Gahegan, Chaitan Baru, Boyan Brodaric

"Sharing our resources, sharing our understanding: Cyberinfrastructure for Archaeology"

Mark Gahegan is professor of geography and affiliate professor of information science and technology at the Pennsylvania State University, USA. He is a GEON Co-PI and has worked on other cyberinfrastructure projects in the fields of plant pathology, e-education and human-environment interaction. His research interests are in knowledge computing, GIS, information visualization and e-science.

Science communities are beginning to construct elaborate cyber-infrastructures to try to overcome some of the fundamental inefficiencies in the science process. This talk introduces the idea of a layered cyber-infrastructure to support e-science activities, concentrating on the problem of sharing understanding via one layer in a cyber-infrastructure— *the knowledge layer* —whose purpose is to capture, preserve and communicate *meaning* associated with sharable science resources. The talk highlights one such e-science initiatives: the *Geosciences Network* (GEON: <http://www.geongrid.org>) and shows how knowledge-level computational tools can help communicate and mediate understanding between collaborating scientists. It also asks the question: 'is the solution designed for the geosciences suitable for archaeology?'

---

### **October 31, 2007**

Fred Limp  
Center for Advanced Spatial Technologies, University of Arkansas

"Interoperability and net-centric architectures: lessons for archaeoinformatics from the Open Geospatial Consortium"

The Open Geospatial Consortium, Inc. (OGC) is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location based services. Since its founding in 1994 it has developed a model process for the effective development of

consensus interoperability standards that have been adopted by the global community. While many of the standards will be of specific interest to the archaeological community, perhaps the most value is in the larger lessons on how to build an effective standards development community. In addition to serving on the Steering Board of Arcaheoinformatic.org Fred Limp was also one of the founders of the Open Geospatial Consortium and served on its Board of Directors from 1994 to 2002. Since 2002 he has served on the Open Geospatial Consortium's Interoperability Institute (OGCI). OGCI is a university research focused group working to move the standards process into the university research community.

---

### **November 14, 2007**

Mark Schildhauer  
National Center for Ecological Analysis and Synthesis, Santa Barbara

"Ecological informatics: challenges and approaches, and potential relevance for archaeology "

This presentation will describe the goals and progress in Ecological Informatics as undertaken by the SEEK (Science Environment for Ecological Knowledge) and KNB (Knowledge Network for Biocomplexity) research projects-- two multi-year, multi-institutional efforts in technology development that were funded by the National Science Foundation. Both projects involved partnerships among ecologists, technologists, and computer scientists, working together to develop usable, powerful tools and cyberinfrastructure to facilitate synthetic, integrative research in ecology and the environmental sciences.

---

### **November 28, 2007**

Julian D Richards, Professor of Archaeology, University of York  
Director, Archaeology Data Service

"Current challenges for digital preservation and delivery"

The Archaeology Data Service recently celebrated its 10th birthday. This wide ranging paper will look forward to some of the challenges of the next ten years, as seen for a UK perspective. It will also outline a range of current research and development initiatives that are seeking to address these issues.

Firstly, the growing complexity and size of digital data sets presents new problems for digital repositories in the cultural heritage sector. The ADS has recently completed the Big Data project for English Heritage, and published its recommendations. We are now working on the EU-funded VENUS project, looking at the archiving issues associated specifically with underwater data sets collected via remote operated vehicles and other deep water technologies.

Secondly, user expectations continue to increase from a generation no longer content with simple downloads, but wanting instead to explore data sets online. The LEAP project has been investigating the issues associated with the linking of electronic publications and archives, creating a series of four

exemplars with our sister e-journal Internet Archaeology. These also raise specific issues of long term sustainability.

A third, linked concern is that the Google generation expect to be able to find what they want on the first page of results by typing a single keyword into a basic search box. We are working with the Natural Language Processing Group in the Dept of Computer Science at the University of Sheffield on Archaeotools. This project provides a powerful way for users to browse through result sets according to an agreed ontology, and we hope to link this faceted classification interface with the results of data mining of the semantics from archaeological grey literature and antiquarian journals.

Fourthly, we need to find more efficient ways of aggregating data provided by a number of physically distributed services. We are collaborating with English Heritage on their Heritage Gateway project to use web services to provide interoperability and cross-searching across distributed data sets.

Finally, I shall describe how our thinking is currently evolving about business models for digital archives. The UK Arts and Humanities Research Council (AHRC) has recently withdrawn funding for our parent body, the Arts and Humanities Data Service. Funding for Archaeology has been ring-fenced for five years but we need to move towards a sustainable business model which relies more upon responsive mode funding rather than core grant. These issues are of concern to all those involved with the establishment of cyberinfrastructure services.

---

## **December 12, 2007**

Ian Johnson  
Archaeological Computing Laboratory, University of Sydney

"ECAI: The snowball still survives"

Starting in 1998, the Archaeological Computing Laboratory at the University of Sydney, under Johnson's direction, developed a novel metadata directory and distributed mapping system based on TimeMap ([www.timemap.net](http://www.timemap.net)), for the Electronic Cultural Atlas Initiative ([www.ecai.org](http://www.ecai.org)). The idea was collaborative online publishing of cultural datasets in map form. The definition of 'cultural' was as wide as the membership of ECAI - characterised more by the fascinating variety of its members than the focus of its mission.

The idea was somewhat under-resourced and technically and socially premature. On the technical front, they made it work and it is still up and running although perhaps more challenging to use than today's generation of instant web apps. On the social front, it long preceded social computing and web mashups, and little attention was given to the reward structure necessary to encourage the sharing of academic data. Consequently the snowball failed to gather substantial snow, except from its creators (a problem Johnson predicted at the 2000 ECAI conference in a paper entitled 'A Snowball's Chance in Hell').

In this presentation he will describe the history of the ECAI Clearinghouse and the lessons to be learned about situating technology in its social context. He will look at some examples of TimeMap use in

Archaeology and History and our more recent work in using some of the features of the ECAI Clearinghouse to build a collaborative social computing and web publication framework, Heurist ([HeuristScholar.org](http://HeuristScholar.org)), which brings together database, clearinghouse, bibliographic and mapping capabilities.

---

## **February 27, 2008**

Katherine Skinner  
Digital Projects Librarian at the Emory University Libraries

"Collaborative Adventures in Distributed Digital Preservation: The MetaArchive Cooperative and the Educopia Institute "

The challenges presented by the concept of digital preservation require and have inspired a number of institutions to work cooperatively in order to accomplish meaningful programmatic advances. Among these collaborative ventures, the MetaArchive Cooperative, established in 2004, has developed an organizational model and technical infrastructure (building on the LOCKSS software developed at Stanford University) for preserving the digital assets of archives, museums, data centers, and libraries in a geographically distributed framework. This lecture takes as its focus some of the strategies that the MetaArchive Cooperative has employed in order to support, sustain, and grow its cross-institutional collaboration. During the session, Katherine will explore some of the logistical and organizational issues that have arisen for the Cooperative over the last four years and will talk more generally about the strengths of different organizational structures for accomplishing particular goals.

The MetaArchive Cooperative (<http://metaarchive.org>) began in 2004 as a collaborative venture of Emory University, Georgia Institute of Technology, University of Louisville, Virginia Polytechnic Institute and State University, Auburn University, Florida State University, and the Library of Congress. The MetaArchive Cooperative has operated a distributed preservation network infrastructure for several years that is based on the LOCKSS software, and has now transformed into an independent, international membership association hosted by the Educopia Institute and based in Atlanta, Georgia.

---

## **April 23, 2008**

John Howard, Arizona State University

**"Fedora Digital Repository for Long-Lived Data Collections"**




No description provided

## Appendix III





























### Selected Survey Results










(from Zoomerang, web host for the survey)

#### 1. What is your highest degree?


Bachelors		36	16%
Masters		110	48%
Doctorate		83	36%
Total		229	100%

### 3. How many years have you spent in archaeology?

1		1	0%
2		1	0%
3		4	2%
4		6	3%
5		11	5%
6		8	3%
7		7	3%
8		8	3%
9		3	1%
10		14	6%
11		4	2%
12		16	7%
13		7	3%
14		5	2%
15		10	4%
16		6	3%
17		8	3%
18		1	0%
19		4	2%
20		16	7%
21		3	1%
22		4	2%
23		2	1%
24		3	1%
25		8	3%
26		2	1%
27		3	1%
28		8	3%
29		2	1%
30		11	5%
31		3	1%










32		6	3%
33		5	2%
34		2	1%
35		7	3%
36		2	1%
37		6	3%
38		1	0%
39		1	0%
40		3	1%
41		1	0%
42		0	0%
43		1	0%
44		2	1%
45		0	0%
46		0	0%
47		0	0%
48		0	0%
49		0	0%
50		2	1%

#### 4. What is your sex?










Female		76	34%
Male		150	66%
Total		226	100%



**6.** What is your primary employment setting?

Academic Department		70	31%
Academic Museum or Research Center		22	10%
Government, State or Local		21	9%
Government, Federal or Tribal		19	8%
Museum or Research Center (not academic)		7	3%
Private For-Profit		54	24%
Private Non-Profit		11	5%
Unemployed or Retired		11	5%
Other, please specify <a href="#">View Responses</a>		14	6%
Total		229	100%

## 7. What is your primary research area(s)?

North America		161	70%
South America		9	4%
Europe		35	15%
Near East		20	9%
Africa		9	4%
Asia		8	3%
Australia		2	1%
Pacific		2	1%
Other		13	6%

**10.** How often have you, as part of your professional role in archaeology, directly or indirectly **made use of** the following?

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Several times per week	Several times per month	Several times per year	Less often	Never
Use web search engines (Google®, etc.)	200 87%	20 9%	2 1%	4 2%	4 2%
Use databases on local computers	145 63%	53 23%	18 8%	10 4%	4 2%
Use databases via web interfaces	86 37%	62 27%	45 20%	27 12%	10 4%
Use statistical software	36 16%	53 23%	51 22%	65 29%	23 10%
Use geographic information systems (GIS)	89 39%	49 21%	48 21%	29 13%	15 7%
Use 3D visualization	25 11%	30 13%	50 22%	62 27%	60 26%
Use digital images or video	144 63%	52 23%	13 6%	15 7%	5 2%

11.

How often do you, as part of your professional role in archaeology, directly or indirectly **create** the following:

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Several times per week	Several times per month	Several times per year	Less often	Never
Create web pages	33 15%	28 12%	41 18%	43 19%	81 36%
Create databases on a local computer	65 28%	65 28%	67 29%	28 12%	5 2%

Create global positioning system (GPS) data sets	29 13%	41 18%	60 26%	41 18%	57 25%
Create geographic information system (GIS) data sets	49 22%	45 20%	44 19%	50 22%	38 17%
Create 3D visualization products	11 5%	22 10%	36 16%	56 25%	101 45%
Create digital images or videos	94 41%	71 31%	37 16%	13 6%	14 6%

**12.** How difficult is it **to learn of** the existence of others' research that may be relevant to your own?

Easy		10	4%
		46	20%
Neutral		71	31%
		72	32%
Difficult		29	13%
Total		228	100%



**13.** How difficult is it to **get access to** others' research materials relevant to your own research?

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Easy	2	Neutral	4	Difficult
Technical reports providing basic documentation of archaeological research projects.	16 7%	51 22%	61 27%	72 31%	30 13%
Artifacts and related collections.	8 3%	22 10%	61 27%	89 39%	49 21%
Primary paper documentation of research.	12 5%	22 10%	51 22%	86 38%	56 25%
Databases systemizing primary research efforts.	5 2%	20 9%	52 23%	83 37%	67 30%





**14.** Please consider the frequency with which you might use the following, if they were readily available:

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Weekly	Monthly	Yearly	Never
A tool that could search currently difficult-to-access gray literature	111 48%	100 43%	15 7%	4 2%
A tool that could query currently unpublished excavation databases	84 37%	102 45%	39 17%	4 2%
A tool that could simultaneously access multiple state site-file databases	87 38%	85 37%	44 19%	11 5%
A tool that could locate and display images and maps from published and grey literature sources according to their temporal and spatial proveniences and content	104 45%	95 41%	27 12%	4 2%
A tool that you could use to “publish” and archive your own datasets,	66 29%	87 38%	69 30%	5 2%
A tool that you could use to “publish” and archive your own reports	56 24%	93 41%	74 32%	6 3%
A tool that you could use to “publish” and archive your own images	70 31%	90 39%	58 25%	10 4%
A tool that you could use to “publish” and archive your own maps	60 26%	93 41%	66 29%	10 4%

- 15.** I would access electronic data or use databases more if it were easier to work with such data.

Yes		216	94%
No		13	6%
Total		229	100%

- 16.** Do you think documentation of the archaeological record is being irretrievably lost?

Yes, and it is a crisis		89	38%
Yes, quickly		74	32%
Yes, but slowly		54	23%
No		15	6%
Total		232	100%

**17.** Whose responsibility do you think that it is to do something about the loss of data?








Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Primary responsibility	Some responsibility	No responsibility
Agencies, companies, or individuals funding the work	117 55%	89 42%	6 3%
SHPOs/THPOs	108 55%	86 43%	4 2%
SAA, SHA, or other professional organizations	52 25%	130 62%	26 12%
Individual archaeologists	115 55%	90 43%	6 3%
Museums and/or curation facilities	109 51%	102 48%	2 1%
Institutions employing archaeologists	126 59%	80 38%	6 3%





**18.** Please indicate your level of agreement with the following statements:

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Disagree	2	Neutral	4	Agree	Not sure / it depends
A web based tool like Google <sup>®</sup> that would allow you to search for specific archaeologically relevant concepts (not just terms) in the published and unpublished literature would be very valuable to my professional role.	8 3%	8 3%	22 10%	53 23%	127 55%	12 5%
It should be an archaeological priority to develop and implement an interoperable archaeological information system for the discipline as a whole.	9 4%	12 5%	32 14%	66 29%	97 43%	12 5%
The professional community should adhere to a minimal set of officially sanctioned data standards.	9 4%	5 2%	27 12%	56 24%	120 52%	13 6%
The US archaeological community should set and apply minimal data standards.	12 5%	10 4%	26 11%	56 24%	97 42%	28 12%
There is a need for one or more organizations to archive and maintain digital resources.	16 7%	4 2%	15 7%	50 22%	140 61%	5 2%
The same organization that sets data standards also should be responsible for archiving digital resources.	23 10%	26 11%	68 30%	31 14%	46 20%	35 15%
All databases submitted to an archiving/access facility should be subject to peer review.	18 8%	29 13%	65 28%	37 16%	53 23%	28 12%
Peer review should be available, but not required, for submitted databases.	22 10%	16 7%	45 20%	65 29%	68 30%	12 5%

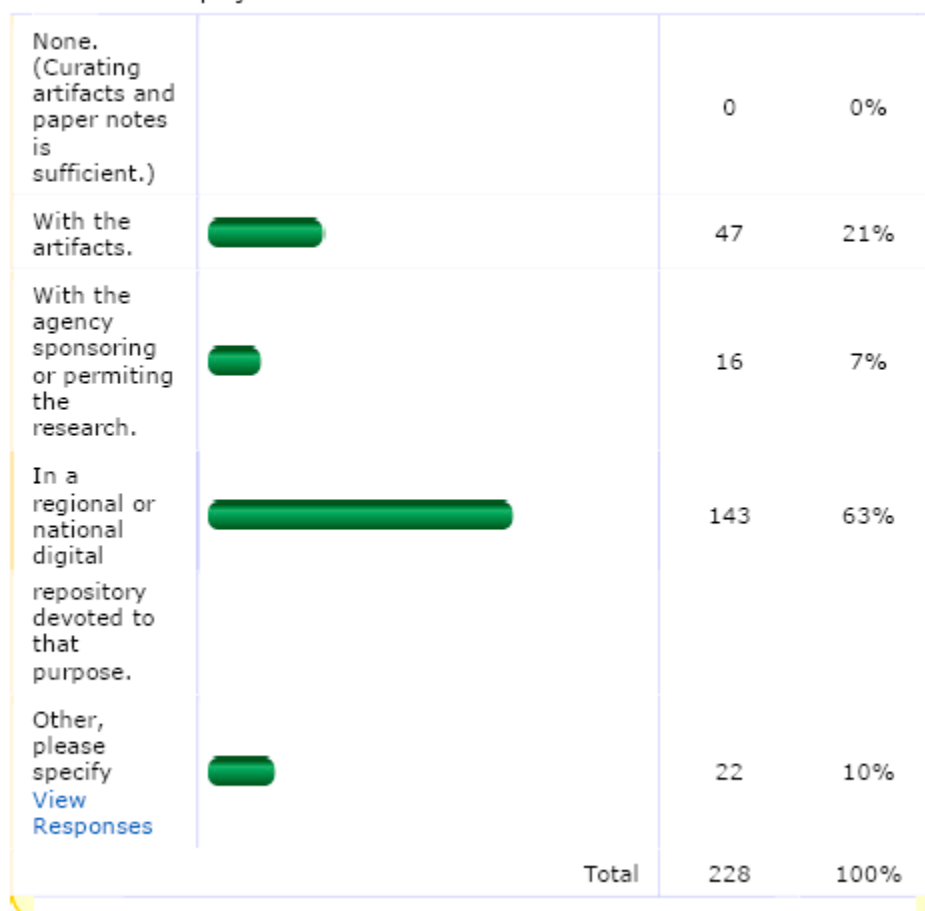
- 19.** For grant- or contract-sponsored research, how long after the termination of the grant or contract is it reasonable for an archaeologist to keep primary data private?

Primary data should never be private		66	29%
6 months		10	4%
1 year		55	24%
5 years		48	21%
10 years		6	3%
At the termination of the grant or contract, regardless of absolute length.		35	15%
Forever		7	3%
Total		227	100%

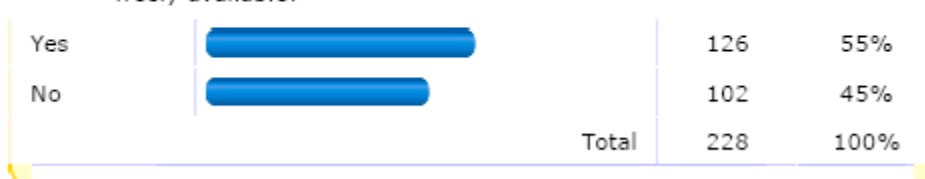
- 20.** Should research sponsors have the responsibility to fund and ensure curation of databases that represent basic documentation and analyses associated with the research?

Yes		206	90%
No		22	10%
Total		228	100%

**21.** Where is the most appropriate place to curate digital data resulting from a project?



**22.** Should scholarly journals or presses have a responsibility to make sure the data on which a published article or book is based be made freely available?



**23.** How suited do you believe the various institution types are to be able to maintain data usability and access in the long run?

Top number is the count of respondents selecting the option.  
Bottom % is percent of the total respondents selecting the option.

	Not suited	2	Neutral	4	Well suited
Cultural consulting firms.	102 44%	64 28%	39 17%	20 9%	5 2%
Local or state government agencies.	21 9%	29 13%	48 21%	86 38%	45 20%
National government agencies.	19 8%	24 11%	38 17%	87 38%	60 26%
Museum curation facilities.	11 5%	28 12%	35 15%	91 40%	65 28%
Academic departments.	39 17%	46 20%	56 24%	62 27%	27 12%
Digital data repositories.	5 2%	8 4%	36 16%	68 31%	105 47%





**24.** Should there be a cost to **access** archaeological data sets in order to fund their wider availability?

Yes		90	39%
No		138	61%

**25.** Should there be a cost to **submit** archaeological data sets in order to fund their wider availability?

Yes		75	33%
No		153	67%
Total		228	100%

**26.** If there were a requirement to pay for digital archiving or retrieval, who would pay?

Your project sponsor		127	56%
Your institution		43	19%
You		23	10%
Other, please specify <a href="#">View Responses</a>		34	15%
Total		227	100%

**27.** What concerns would you have about routinely submitting digital reports or datasets to a digital archive?

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Unconcerned	2	3	4	Very Concerned
Cost (including cost of preparation)	17 7%	27 12%	50 22%	70 31%	63 28%
Amount of time required	20 9%	30 13%	51 22%	74 32%	53 23%
Access to technology	44 19%	35 15%	52 23%	63 28%	33 15%
Sufficient knowledge of technology	51 23%	43 19%	46 20%	52 23%	33 15%
Issues of confidentiality and data sensitivity	37 16%	32 14%	38 17%	51 22%	69 30%
Issues of credit/citation	38 17%	37 16%	62 27%	50 22%	40 18%
Fear of criticism of your work	121 54%	59 26%	33 15%	10 4%	3 1%
Competition's access to your data	109 48%	43 19%	43 19%	17 7%	15 7%

## Appendix IV

### Santa Fe Meeting participants and Schedule

#### Participants

1. Keith Kintigh, Arizona State University (convener)
2. Jeff Altschul, Statistical Research, Inc.
3. Tim Kohler, Washington State University
4. Fred Limp, University of Arkansas
5. Dean Snow, The Pennsylvania State University
6. Julian Richards, Archaeology Data Service, University of York
7. Debbie Harmon, University of Arkansas
8. John Howard, Arizona State University
9. Allen Lee, Arizona State University
10. Brian Crane, Versar, Inc (Springfield, Virginia)
11. Katherine (Kitty) Emery, University of Florida
12. Eric Kansa, University of California at Berkeley, Alexandria Archive Institute
13. Francis (Frank) McManamon, National Park Service (Washington DC)
14. Fraser Neiman, Monticello, University of Virginia
15. Vincas (Vin) Steponaitis, University of North Carolina
16. Phillip (Phil) Walker, University of California at Santa Barbara
17. Willeke Wendrich, University of California at Los Angeles
18. Thomas Whitley, Brockington & Associates, (Norcross, GA)
19. Worthy Martin, University of Virginia
20. Herbert Van de Sompel, Los Alamos National Laboratory

#### Meeting Agenda

Thursday, February 14

Check In, Hotel Santa Fe - 1501 Paseo de Peralta Santa Fe, NM 87501 800-825-9876

Dinner on your own

Informal Reception (details to be announced)

#### Friday, February 15

Breakfast On your own (Coffee and light refreshments in Kiva A from 7:30)

8:30-9:45 Introduction to Archaeoinformatics.org & Participants (Kiva A)

Welcome & Participant Introductions

Summary of Archaeoinformatics.org Scope, Vision & Accomplishments

Fred Limp – University of Arkansas

Dean Snow – Pennsylvania State University

Tim Kohler - Washington State University

Keith Kintigh – Arizona State University

Jeff Altschul – Statistical Research, Inc.

9:45-10:00 Break

10:00-10:45 Demonstration of tDAR: the Digital Archaeological Record

- 10:45-11:30 When Our Data Sources Disappear & Our Responsibilities and Options  
Implications of in-field recording and repatriation (Phil Walker)  
Beyond Databases, Text, and Images (Fred Limp and Jeff Altschul)
- 11:30-1:00 Lunch on your own (see Restaurant Suggestions)
- 1:00-2:45 Advisory Board Member Presentations & Discussion (15-20 min each)  
Lessons from the Archaeology Data Service (Julian Richards)  
Problems of Adoption and Advantages of Due Credit (Eric Kansa)  
Community Buy-in and the Experience of the Digital Archaeological Archive  
of Comparative Slavery (Fraser Neiman)  
Situating an Informatics Vision in Computer Science Reality (Worthy Martin)  
Infrastructure Interoperability (Herbert Van de Sompel, tentative)
- 2:45-3:00 Break
- 3:00-5:00 Advisory Board Member Presentations & Discussion (15-20 min each)  
Cyberinfrastructure & the National Archaeological Program (Frank  
McManamon)  
Values and Rewards for the Academic Community (Vin Steponaitis)  
Values for Federal Agencies and their Contractors (Brian Crane)  
Values and Impediments for CRM Acceptance (Tom Whitley)  
Challenges of Long-Term, Multi-Investigator Research (Willeke Wendrich)  
Values for Natural History Museums (Kitty Emery)
- 6:00 Dinner Together -meet in Lobby at 6:00, or 6:15 at the Restaurant  
Café San Estevan 428 Agua Fria St
- 9:00PM Steering Committee Only - Form Small Groups and Assign Questions

### **Saturday, February 16**

- Breakfast On your own (Coffee and light refreshments in Kiva B from 7:30)
- 8:30-9:00 Combined Session (Kiva B) – Small Group Assignments
- 9:00-11:30 Breakout Session (Kiva B, Library, and Canyon rooms)  
Four groups: the Technical Advisory Board and the Disciplinary Advisory  
Board divided in thirds, each with representatives of the steering  
committee, will consider a set of questions, develop discussion points,  
and formulate recommendations.
- 11:30-1:00 Lunch on your own
- 1:00-3:00 Combined Session – Small Group Reports & Discussion  
This discussion will lead to specific charges to the Boards.
- 3:00- Explore Santa Fe & Dinner on your own

### **Sunday, February 17**

- Breakfast On your own (Coffee and light refreshments in Kiva B from 7:30)
- 8:30-9:45 Small Groups Reconvene (same rooms)
- 9:45-10:00 Break
- 10:00-11:30 Separate Advisory Board and Steering Committee Meetings  
Disciplinary Advisory Board (Kiva B)  
Technical Advisory Board (Canyon)  
Steering Committee (Library)
- 11:30-1:00 Lunch on Your Own
- 1:00-2:15 Combined Session (Kiva B)



Brief Board Reports; Discussion toward Recommendations  
2:15-2:30 Break  
2:30-4:00 Formulation of Final Recommendations

**Monday, February 18 (Steering Committee only)**

8:30-4:30 Assess recommendations  
Mellon Foundation presentation (March 6)  
Outline proposal/report & make assignments  
Refine plan for SAA Forum

**Santa Fe Meeting Presentations**

A series of presentations were made covering the background to the project, current initiatives and other related projects. The presentations power points are available on the archaeoinformatics.org web site at [http://archaeoinformatics.org/Board\\_Present.html](http://archaeoinformatics.org/Board_Present.html)

**Fred Limp** and **Jeff Altschul** - discussed the curation of digital data beyond databases, text and images- the archiving and discovery of other data including geospatial, remote sensing, high density LIDAR and laser scans, geophysical and lab instruments, and virtual reality models.

**Phil Walker** – described how today’s frequent insistence on in-field recording and repatriation (not just of human remains) puts extra demands on cyberinfrastructure and introduce how we can use an information infrastructure with technology (such as 3-D scanning) to help mitigate data losses.

**Julian Richards** – presented the core elements of the Archaeology Data Service effort, including a comment on how you are linking the data archives to publications. He includes a brief discussion of the success of ADS’s efforts to establish best practices relative to digital data.

**Eric Kansa** – has done some useful thinking about and development of software solutions to some the core sociological problems that a successful infrastructure must face. He outlined what he sees as the most important of these. He includes a discussion of enhanced citation or other credit accruing to digital publication.

**Fraser Neiman** – based on his experience with the Digital Archaeological Archive of Comparative Slavery, he identified the key elements to achieving archaeological community buy-in to an information infrastructure and the biggest challenges.

**Worthy Martin** – discussed the understanding that implementation of a cyberinfrastructure can be staged, and evaluates the vision presented in our Mellon Foundation proposal and the report of the 2004 workshop report to be reasonably positioned with respect to the state of the art in computer science.

**Herbert Van de Sompel** – assessed the opportunities and advantages vs. the costs and disadvantages of achieving interoperability with other science informatics and digital library efforts.

**Frank McManamon** – discussed the two or three most important values that an integrated text and data information infrastructure could bring the mission of the National Archaeology Program. He assessed ways in which those values can be “sold” to agencies? In light of previous governmental initiatives, he evaluates what would be required for them to adopt a solution from the outside.

**Vin Steponaitis** -provided a statement of what he sees as the most important values a cyberinfrastructure could bring to the academic research community. He discussed how the contributions to this infrastructure be rewarded in ways that “count” in an academic setting.

**Brian Crane** – discussed, in light of the kinds of large datasets that the Department of Defense generates and maintains, how would their integration and interoperability with major datasets not from DOD lands (by way of the proposed infrastructure) benefit the larger DOD mission. He identified aspects of an information infrastructure that could simultaneously serve the interests of DOD and its contractors.

**Tom Whitley** – discussed, from the standpoint of the practice of CRM, what he sees as the two or three most important values of a cyberinfrastructure and what do he sees as the two or three largest impediments to its adoption. He evaluated how this infrastructure might improve efficiency or effectiveness of cultural resource management as it is practiced in the US and how might we convince people of these advantages.

**Willeke Wendrich** – discussed how large-scale, long-term, international multi-investigator research poses a particular challenges to or presenting particularly persuasive opportunities for an information infrastructure for archaeology.

**Kitty Emery** – reflected on the values of a research infrastructure we proposd here to the missions of anthropology or natural history museums.

## Appendix V

### Publications and presentations

Altschul, Jeff and Keith Kintigh,

- 2008 Archaeoinformatics.org. A presentation to the Federal Chief Archaeologists, Washington, D.C.

Kintigh, Keith W.

- 2008 tDAR: The Digital Archaeological Record. Invited Presentation to Plenary Session: Mega Databases in Southwestern Archaeology: Current Status and Future Prospects. 2008 Pecos Conference. 9 August 2008. Flagstaff, AZ
- 2008 Digital Antiquity: Planning an Information Infrastructure for Archaeology, Keith Kintigh, organizer). SAA Board Sponsored Forum. 73<sup>rd</sup> Annual Meeting of the Society for American Archaeology, Vancouver, B.C., March 27, 2008. (Organizer, Moderator, and Presenter)
- 2008 A Look into the Research Life of Faculty. Invited Lecture, ASU University Libraries, March 25, 2008, Tempe.

Kintigh, Keith W., Fred Limp, and Dean Snow

- 2008 Digital Antiquity: Planning a Digital Information Infrastructure for Archaeology. Invited presentation at Archaeology All Projects Meeting, The Andrew W. Mellon Foundation, New York, March 5-7, 2008, New York.

Kintigh, Keith W., K. Selçuk Candan, John B. Howard

- 2008 The Digital Archaeological Record: Requirements for Database Preservation, Access, and Integration. Presentation at the 2008 Conference of Commission 4 of the Union Internationale des Sciences Préhistoriques et Protohistoriques, organized by François Djindjian & Hélène Noizet, June 3-4, 2008, Paris, France.

Kintigh, Keith W.

- 2007 The Challenge of Archaeological Data Integration. Invited Lecture, Statistical research, Inc., Tucson, AZ. January 29, 2007.
- 2006 The Challenge of Archaeological Data Integration. Paper presented in the invited session, Technology and Methodology for Archaeological Practice: Practical Applications for the Reconstruction of the Past, organized by Alexandra Velho and Hans Kamermans at the Congress XV of the Union Internationale des Sciences Préhistoriques et Protohistoriques, 4-9 September 2006, Lisbon, Portugal.
- 2006 Discussant in the invited forum, Toward a Cyber-infrastructure for Archaeology: Tools and Incentives, sponsored by The William and Flora Hewlett Foundation and organized by Eric Kansa and Michael Ashley at the 71st Annual Meeting of the Society for American Archaeology, April 26-30, 2006, San Juan, Puerto Rico

Kohler, Timothy A.

- 2008 Overview of Modeling Activities of the Village Ecodynamics Project and Introduction to the Digital Archiving Initiative of Archaeoinformatics.org. IGERT Program in Evolutionary Modeling Seminar, 25 September 2008, Washington State University.

Schoville, Benjamin Mallorie Hatch, and Keith Kintigh

- 2008 tDAR: Introducing “the Digital Archaeological Record.” Poster presented at the 10<sup>th</sup> biannual Southwest Symposium, January 17-19, 2008, Tempe, Arizona.

## **Appendix VI**

### **Arizona State University Report Digital Antiquity: Planning a Digital Information Infrastructure for Archaeology**

#### **Arizona State University Subcontract to the University of Arkansas Final Report to the Andrew W. Mellon Foundation**

**Keith W. Kintigh, Professor and Associate Director  
School of Human Evolution & Social Change**

The overall objective of the subcontract activity was to add an international component to the effort, as proposed in the Digital Antiquity planning grant proposal and as recommended by the 2006 NSF-funded workshop.<sup>1</sup> The ASU effort was focused on Mexico with three goals: 1) to raise awareness in Mexico of archaeological data preservation and access issues in general and of the Digital Antiquity initiative in particular; 2) to explore with scholars working in Mexico the possibility of uploading their datasets and documents into Prototype tDAR (the Digital Archaeological Record), including informers to the Instituto Nacional de Antropología e Historia [INAH]), and to assist them in doing so to the extent that they are willing; and 3) to identify likely topics for a substantive Mesoamerican case study that could be a component of a future proposal and begin acquiring datasets relevant to that case study.

During our work on the subcontract, advice from the Mellon Foundation staff made it clear that there would not be an international component to the initial Digital Antiquity implementation proposal; from that point on our emphasis shifted to the first two goals (so as not to raise unrealistic expectations regarding any future synthetic conferences).

Activities under the planning grant can be divided into two groups: funded efforts and cost-shared or contributed efforts. Funded activities focused on Mesoamerican scholars and their documents and data. These efforts also benefited from substantial contributed effort by ASU professors Ben Nelson and George Cowgill. Nelson very effectively directed the Mesoamerican component of the subcontract. Cowgill, in August 2007, presented the initiative to an international meeting of scholars working on issues surrounding the archaeology of Teotihuacan. Cowgill's presentation effectively launched our efforts in Mexico. A briefing paper prepared for that discussion is attached.

A large contributed effort was devoted to the advancing the Digital Antiquity initiative generally and to developing the initial Digital Antiquity implementation proposal. Much of this effort was contributed by

---

<sup>1</sup> Kintigh, Keith W. (ed). 2006. The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71: 567-578.

Kintigh, with substantial contributions by faculty members K. Selçuk Candan, Allen Lee, Katherine Spielmann, Margaret Nelson, and Ben Nelson.

## **Funded Activities**

The funded efforts focused on laying the foundation for a digital archive strength in the archaeology of central Mexico (in and around the Basin of Mexico, including Classic Period Teotihuacan and Aztec Tenochtitlan). This involved introducing tDAR to US, Mexican, and other scholars working in this area. To the extent possible, once Prototype tDAR was operational, we assisted these scholars in acquiring and uploading datasets. Funded efforts were pursued by young independent scholars Michelle Elliott and Vince Schiavetti both of whom received PhDs under Professor Nelson. These individuals worked on the grant under contract and were supervised by Nelson with the assistance of Kintigh.

Because Michelle Elliott is based at Universidad Nacional Autónoma de México (UNAM) in Mexico City, she was in an excellent position to promote the proposed archive. She visited Tempe and was introduced to the project and worked closely with Ben Nelson and Keith Kintigh via videoconference and email. Elliott was able to present the Digital Antiquity project to numerous scholars (Mexican, US, and French) who lived in or passed through Mexico City during the grant period. As a part of this effort, she created a useful PowerPoint tutorial to introduce the key concepts ([http://tdar.org/confluence/download/attachments/131075/tDAR\\_Instructions.pdf?version=1](http://tdar.org/confluence/download/attachments/131075/tDAR_Instructions.pdf?version=1)).

Elliott was able to discuss the Digital Antiquity initiative with number of important scholars. Through Elliott's efforts we were able to incorporate into Prototype tDAR Jeffrey Parsons' (University of Michigan, emeritus) and his colleagues' database for their systematic survey of the Valley of Mexico Survey. Because of its focus on an area of enormous archeological significance and because of the author's willingness to share these data, the database has been widely reused. Its availability in Prototype tDAR (and tDAR) will further enhance its accessibility and usability for current and future scholars.

Elliott was also able to assist Mexican senior scholar Emily McClung de Tapia (UNAM) in registering information from a number of her completed projects. Additional scholars agreeing to contribute data include Tom Charlton of the University of Iowa, George Cowgill of Arizona State University, and Ian Robertson of Stanford University.

As anticipated, there were delays associated with contacting scholars, with obtaining their agreement to contribute data, and in preparing their projects for ingest. Additional delays were introduced by the fact that Prototype tDAR was not fully operational as early as anticipated, and because of Mexican scholars' need to consult with INAH on the appropriateness of participation. As a result, fewer datasets were registered in tDAR than we had hoped. However the degree of general support was somewhat greater than we anticipated.

We are quite encouraged by the results these efforts, especially in light of the fact that Mexican scholars have a somewhat different culture of sharing data than US scholars. We believe that we have met the

major subcontract goal of laying the groundwork for international participation in tDAR, once the production version of tDAR is launched under the implementation grant.

During the grant period, the opportunity arose to incorporate in tDAR a major systematic database and associated documents for the site of La Quemada, Zacatecas. Because it seemed that this would be a very important resource for the archive, because it represented an excellent example of a senior US scholar (Ben Nelson) sharing a very large corpus of data from his long-term research, and because there appeared to be sufficient funds, we added working toward the incorporation of this database to the original scope of work.

Vince Schiavetti worked to incorporate Nelson's La Quemada database and associated documents into Prototype tDAR. While we have not yet completed this effort, we have already made important documents and data available and work on this project continues. Challenges presented by this large (300MB+) and complex dataset have provided valuable user testing and expert feedback and has substantially informed our planning for the tDAR archive.

Finally, we engaged in considerable discussion concerning the development of a series of international conferences directed to synthetic work on an important topic. In this model, selected scholars would pursue synthetic research using the corpus of participants' contributed data shared through tDAR. One actively discussed possibility for such a conference would have to do with the regional reorganization of central Mexico that occurred after the collapse of Teotihuacan. While firm plans must await further development of the production archive and better data integration capabilities (in progress), we have made considerable progress in the development of what we believe could be a very persuasive case study in the research value of the archive for which we can seek independent funding.

## **Contributed Effort**

Contributed effort was applied toward three objectives: implementation planning; promoting the grant activities to the archaeological community; and development of a Mesoamerican component to the project.

Kintigh was the primary drafter and organizer of the implementation proposal and led archaeoinformatics.org through the grant period and up to the present. Implementation planning included a large number of physical and virtual meetings with the archaeoinformatics.org group (always attended by Kintigh, usually by Lee and Howard, and on several occasions by Candan). It also included numerous internal ASU meetings (attended by Kintigh, Candan, Lee, Howard, Spielmann, Margaret Nelson, and Ben Nelson) devoted more closely to defining the substantive objectives of the effort and to designing an appropriate functional specifications and architecture for the proposed archive.

Kintigh was actively engaged in promoting the Digital Antiquity initiative to the archaeological community. This involved presentations to a number of professional audiences that are detailed elsewhere in the overall report.

Meetings devoted to developing the Mexican case studies included Ben Nelson and Kintigh, sometimes attended by Elliott, Cowgill, and Schiavetti.

## **Expenditures**

Funds were generally expended as budgeted, with somewhat less than budgeted on salary and wages and materials and supplies and somewhat more on travel. The funded efforts, rather than being pursued by graduate students as budgeted, were pursued by the independent scholars indicated above. We believe that this was more cost effective (more effective people with somewhat higher salaries but drastically smaller benefits). No funds were needed for media translation. Travel monies were expended both for the Mexican work and for Kintigh to attend conferences and meetings to advance the goals of the Digital Antiquity initiative. Unexpended funds late in the grant period were transferred to the University of Arkansas for their use in implementation planning.



## Digital Antiquity: Designing an Information Infrastructure for Archaeology

### Proposed Central Mexican Case Study

Keith Kintigh – [kintigh@asu.edu](mailto:kintigh@asu.edu) – Arizona State University - 2007-08-27

A multi-institutional collaboration known as [archaeoinformatics.org](http://archaeoinformatics.org)<sup>2</sup> has been awarded a planning grant titled Digital Antiquity: Designing an Information Infrastructure for Archaeology by the Andrew W. Mellon Foundation. The grant funds planning for the implementation of a digital information infrastructure (cyberinfrastructure) for archaeology. As described below, we are seeking your advice as we plan a proposed central Mexico case study that would use the infrastructure and form a part of the implementation plan (and proposal) we are developing. More information on our current Mellon-funded initiative is available at <http://archaeoinformatics.org>.

Briefly, we anticipate developing a system that will at once facilitate synthetic research in archaeology and provide for the long-term preservation of documents, images, and especially systematically collected archaeological data stored in digital databases. The system would provide a sustainable repository for contributed datasets, documents, and images. It will not only store the datasets and permit locating and retrieving them, it will provide innovative capabilities to integrate information from multiple databases developed by different people at different times using different recording protocols. It would access data sources uploaded to a central server as well as datasets on distributed servers maintained by other institutions. The overall objectives of this initiative were formulated at broad-based, NSF-funded workshop in Santa Barbara that was reported in a recent American Antiquity article (Kintigh 2006). The workshop conclusions were endorsed by SAA, SHA and AAPA. George Cowgill, Linda Manzanilla, and Ian Robertson participated in that workshop.<sup>3</sup>

Under our current, year-long Mellon Foundation planning grant, [archaeoinformatics.org](http://archaeoinformatics.org) will develop a proposal to fund the implementation of an archaeological information infrastructure. While software development will, of course, be one component of the implementation, we are equally concerned with engaging the scholarly community of archaeologists in this enterprise. Without their buy-in, this effort can have little impact.

To this end, our planned implementation proposal to the Mellon Foundation will include the development of a small number of “case studies” that are intended to persuasively demonstrate the research value of the system through their role in advancing research on substantive archaeological questions. It appears that archaeological research in central Mexico provides an excellent candidate for such a case study. This central Mexico focus was identified as a high priority by the Santa Barbara workshop. Subsequent discussions of the [archaeoinformatics.org](http://archaeoinformatics.org) steering committee (and the Mellon Foundation staff) have confirmed this assessment.

---

<sup>2</sup> The Steering Committee is composed of Keith Kintigh at Arizona State University, Dean Snow at the Pennsylvania State University, Fred Limp at the University of Arkansas, Tim Kohler at Washington State University, and Clay Mathers at Statistical Research, Inc

<sup>3</sup> In addition to the Mellon Foundation planning grant, NSF has funded a team of ASU archaeologists and computer scientists to implement a pilot cyberinfrastructure for systematically collected faunal data. As a result of this NSF project, we expect to have basic data repository functions not restricted to faunal data (but not the sophisticated data integration) running within a few months (see <http://cadi.asu.edu>).

Over the next several months, ASU is responsible for designing and preparing a budget for a case study that will be included in the implementation proposal to be submitted in the spring of 2008. It is in this effort that we seek your advice. We have a great deal of latitude in how we do this, but permit to offer one possible vision. At the core of this case study would be a compelling archaeological research topic (or cluster of related topics) and a group of scholars committed to research on that subject. In this effort we envision a novel sort of collaboration leading to a new kind of synthesis.<sup>4</sup> In the traditional model, one might convene a conference of scholars, each of whom brings to the table a draft paper to be discussed, with the conference result being an edited volume of revised papers.

In contrast, we envision a series of two or perhaps three meetings spread over some time, in which the scholars instead bring databases of primary observations that they are willing to share, augmented by the other databases judged central to progress on the topic. The data integration tools provided by the infrastructure (with some human support as well) would allow the group to explore issues at a range of analytical levels. The result of this process would, of course, include traditional publications. It would also result in a powerful collection of datasets that could be made available to the scholarly community through the infrastructure, of course with appropriate control and credit to the contributors.

To be most effective, we would need to identify a research topic for which progress (or resolution) will be viewed as important by a substantial range of scholars. The topic should be one that the relevant experts believe would benefit from analysis or re-analysis of data from a number of projects (including both recent and older research) that would be impractical or impossible without advanced tools to facilitate the integration of the diverse datasets. Ideally, most of the databases would be in electronic form and in a state that would not require too much work to make them ready to share with other scholars. Similarly, we need to identify a group of scholars with a research commitment to the topic who are willing to share their own data and to accept the challenge of thinking through the issues in ways that may require reevaluations of their own and others' work. To the extent that the group is able to identify other crucial datasets for this research, and the scholars or institutions that control those datasets are willing, grant funds could be devoted to getting those data in digital form. It does not seem possible to anticipate new fieldwork under the grant. However, it seems quite plausible that the results of this effort might yield a compelling argument for new fieldwork.

Of course we cannot guarantee the implementation funding and we do not yet have a clear idea of the level of overall funding that we can anticipate. However as we see it, the implementation grant could fund research meetings of the relevant scholars, assistance in getting necessary datasets up and running, assistance in obtaining or making ready legacy datasets, and technical assistance in pursuing the synthetic research.

As indicated above, we now have a small planning grant and modest funds to moving this case study forward. Unfortunately, we cannot fund large meetings of senior scholars to help formulate a plan. However we do have graduate student and travel funding that can be used to get a head start on pulling together some of the datasets and associated metadata that are likely to be needed for the implementation grant. It may make sense to start with available datasets from central Mexico that are complete from the standpoint of the responsible scholars as well as important central Mexico datasets from retired or about to retire scholars (insofar as they are willing to make them available). To the extent that they are willing, this could include datasets from some of the Ometusco Seminar

---

<sup>4</sup> This sort of collaboration has been pioneered by the National Center for Ecological Analysis and Synthesis.

participants. For example, we have discussed with George Cowgill the possibility of assisting in making the Teotihuacan Mapping Project Data (or some substantial portion thereof) available on-line.

We would very much appreciate your guidance in developing this project. The primary questions on which we would appreciate your advice, either individually or collectively, are 1) what topics might be considered; 2) what scholars might be interested in participating; 3) what existing datasets that would be of value for this research in the implementation phase; and 4) what central Mexico datasets might be available now to be put on-line at modest expense from our existing grant.

## Appendix VII

Memorandum

August 26, 2008

To: Fred Limp, Lead PI, Archaeoinformatics Planning Grant  
From: Tim Kohler, PI for Washington State University  
Re: Final Report from WSU on Planning Grant Outcomes

This brief letter report has three appendices: (1) a final financial report from WSU on the uses to which the \$22,388.00 apportioned to WSU on this grant were put; (2) a report on remaining minor data issues with the Dolores Archaeological Program database; and (3) a report on remaining minor data issues with the Bandelier Archaeological Project database.

In addition to collaborating with the other co-PIs on furthering the general goals of the Archaeoinformatics initiative through participating in planning, bi-weekly Access Grid sessions, SAA sessions, and the like, WSU proposed to “pursue a case study in the Four Corners region of the American Southwest. Candidate datasets under consideration include the Dolores Archaeological Project datasets and the Cedar Mesa Project datasets (both requiring coordination with the Bureau of Land Management [BLM]) and various datasets developed by the “Village” project (<http://www.wsu.edu/~village/>), some of which will require coordination with BLM, the National Park Service and/or Crow Canyon Archaeological Research Center, Cortez, Colorado. These efforts will include coordination with data curators, coding of metadata, and “proofing” of software tools against these data and metadata. These efforts will be assisted by the resources of the WSU Museum of Anthropology.”

Our total expenditures (first appendix) were closely in line with our proposed expenditures, though we overspent 15 percent on salaries, and underspent 95 percent on benefits and 88 percent on goods & services, altogether leaving \$992.73 (4 percent) of the \$22,388.00 apportioned to WSU unspent.

Our accomplishments with this funding were closely in line with our proposal, although we have completed most of the work on not one dataset, as proposed, but on two. Specifically, we have put on line, locally stored in a PostgreSQL database (an open-source database), and available for eventual incorporation into the next generation of the tDAR software, 25 datasets of the Dolores Archaeological Project (DAP; accessible at <http://galisteo.anth.wsu.edu/dap/>), and 11 datasets of the Bandelier Archaeological Excavation Project (BAEP; accessible at <http://galisteo.anth.wsu.edu/BAEP/>). With minor exceptions set forth in appendices 2 and 3, all the labeling (data dictionaries) associated with these databases is complete and also accessible on-line. We chose the BAEP dataset instead of one of the other candidate datasets listed in our proposal because (1) the Cedar Mesa dataset was (and still is at this writing) incomplete, and (2) the BAEP

datasets are much more at risk of becoming inaccessible than are the datasets of the “Village” project, for example.

One area in which we did not complete the goals of the proposal is that the metadata for these datasets, and the datasets themselves, are not yet entered into tDAR or its successor, although we believe we have collected the necessary metadata for this to take place in the future, and the databases themselves are ready to go. The necessary facilities for ingesting these data and metadata into tDAR were only becoming available, in trial form, as our funding was exhausted and the grant period was ending.

It remains our hope that these two databases will be joined by others in the next phase of development of the digital archive, and that together these resources will serve as an attraction to other regional datasets, providing a rich regional resource for researchers and those in CRM alike.

Attachments: Appendix 1, Signed Financial Report  
Appendix 2, Remaining Data Problems, DAP  
Appendix 3, Remaining Data Problems, BAEP

**WASHINGTON STATE UNIVERSITY  
FINAL FINANCIAL REPORT**

**FUNDING AGENCY:**

UNIVERSITY OF ARKANSAS  
RESEARCH SUPPORT & SPONSORED PROGRAMS  
ATTN SUBAWARD ADMINISTRATOR  
120 OZARK HALL  
FAYETTEVILLE AR 72701

WASHINGTON STATE UNIVERSITY  
BUSINESS SERVICES/CONTROLLER  
SPONSORED PROGRAMS SERVICES  
240 FRENCH ADMINISTRATION BLDG  
PULLMAN WA 99164-1026  
(509) 335-2058 FAX (509) 335-2071

Project Title:

DIGITAL INFO INFRASTRUCTURE ARCHAEOLOGY

Grant/Contract No.

SA0809242

WSU NO.

13N-2482-0211

<b>COST CATEGORIES</b>	<b>BUDGET FOR PERIOD 7/1/2007</b>	<b>TO</b>	<b>EXPENDITURES FOR PERIOD 6/30/2008</b>	<b>BALANCE</b>
SALARIES	15,103.00		17,364.88	(2,261.88)
WAGES	0.00		0.00	0.00
Personal Services	0.00		0.00	0.00
G & S	1,000.00		115.80	884.20
TRAVEL	3,900.00		3,846.50	53.50
EQUIPMENT	0.00		0.00	0.00
BENEFITS	2,385.00		68.09	2,316.91
GRANTS	0.00			0.00
RESTRICTED	0.00		0.00	0.00
FACILITIES & ADMINISTRATIVE	0.00		0.00	0.00
<b>TOTAL</b>	<b>\$22,388.00</b>		<b>\$21,395.27</b>	<b>\$992.73</b>

CASH RECEIVED TO DATE:	\$ 21,395.27
EXPENDITURES TO DATE:	21,395.27
BALANCE ON HAND:	\$ -

BY:   
Kim Small, Accounting Manager

8/26/2008

DATE

## Appendix 2: Status of DAP database as of 5/6/08, including remaining problems

- Overarching issues/notes:
  - Fix database timeout (download speed)
  - SUNUM column: This column varies between either denoting a study unit number or the grid number based on the study unit type. Because of this a normal dictionary update cannot be completed. (100300 could be either 100300 or 100E 300S). As such, can do one of a few things:
    - Option 1) Create a code loop in the php code that reads "if column sutype >= 89 then sunum = \*\*\*E \*\*\*S. Else leave sunum as it is. This would modify the whole branch of the code, and might be difficult to implement. Further I did some glancing for a function that would dismantle sunum so that it would know to take the first three digits, add an "E" and a space then take the second three digits and add an "S", as of end April 2008 I was unable to find something that would work, (I am sure there is something out there)
    - Option 2) Look through the code and determine how many incidences of sutype >89 there are and then create two new columns, one a SUNUM' and the other a "SUNUM grid" and then populate these two rows with the attributes. Might not be too difficult to do in excel, although would have to do for each database with SUNUM independently. Also would look to see how difficult it is to copy to columns in postgresql (have copied to entire database tables, but not individual columns)
    - Several columns have a few values without labels while being mostly defined. Should we spend the time to track these down (if it means having to contact AHC etc)
    - There are several columns in which the definitions for the numeric code give the definition in terms of a X length numeral, but in the columns there are often several values that are shorter than the X length. An example is the color/streak columns in the rarerox database. The definitions are coded for a 6 digit numeral, but we have several 3,4, and 5 digit numbers. On some of them I believe it is safe to assume that it is a case of the database program cutting out zeros before the number (003 becomes 3). So on those (generally the 5 and 4 digits in the rarerox streak column) I completed the definition assuming the zeros had been cut. However for the three digit numerals I was unable to do this because the definition relies on there being at least one number in the first two (thus if I have 300, assume it is 00300 then I have a value of "00" for the first two numbers which does not translate to anything). Other columns have similar problems. Generally I have left these without definitions as I am unsure on what to do with them.
    - Table below describes all that (hopefully) is left to do on all of the databases in the DAP project.


Dataset	Database	Level of Completeness	Missing:
DAP	BUNNY10	Believe it is complete	colloc or disc column definitions? No info on either of these
	CERAM10	complete	
	CORN10	Believe it is complete	should cultigen number have definition values? Page we got from AHC does not have values, nor does Wilshusens
	DATE10	complete	
	DEBTG10	complete	
	FAUNA10	complete	
	FEATLNK	95%	SUNUM
	FLT10A	complete	
	FLT10B	complete	
	HAFT10	complete	
	MACRO10	Believe it is complete	Left class, family, genus, species columns without definitions per conversation with Dr. Kohler. Would have to work backwards from the Taxon column as we do not have definitions for those values. Are the whole and frag columns percents/numbers or are they missing definitions? DAP guide doesn't say much
	Mano10	40%	Matid- same as all nfl material class? All values in this column are "4", and the DAP guide says "refers to material ID. However, beyond this brief definition, no other explanation for this attribute was located. Presumably it keys for something as simple as nonflaked lithic artifact. All values for this attribute in the mano10 file are coded as 4. It should not be mistaken for CLASS" (W. 173)
			Exassoc and inassoc columns
			surface column
			spafunc column
			surface column
			LNGCRV columns (long curve)
			texture columns
			striation and grips columns
			shtrv column (short curve)
	Maps10	80%	missing some values for taxon (40s and 80s, low 90s)



		missing some values for area
		sunum
		exxassoc
		innasoc
		spafunc column
Metate10	40%	Matid- same as all nfl material class? All values in this column are "4", and the DAP guide says "refers to material ID. However, beyond this brief definition, no other explanation for this attribute was located. Presumably it keys for something as sim
		morhpo and morhpoB
		surface, latcav, straiton texture, end columns
MISC10	95%	missing def. for value of "84" for sstype
NFLT10	95%	missing def. for value of "84" for sstype
ORNMT10	98%	form1 column: two forms of ornament, not in data from AHC or in DAP GUIDE
PHOTO01	80%	matsurt? Ssfetype needs definitions, Otype needs defs
POLLN10	believe it is complete	
PROV	90%	area column, SUNUM,
PROV10	85%	area and subarea, SUNUM
RAREROX	97%	3 digit streak and color values (supposed to be 6 digit, so lots of interpretation open to how they could be analyzed)
SAMPLE10	Believe it is complete	
SEDS10	complete	
TSCER4	90%	SUNUM, ST2 column (AHC data doesn't have- says this is only info they have), STC1 values outside range?
TSFLT4	90%	same as TSCER4
TSNFL4	90%	same as TSCER4
TSPRV4	90%	same as TSCER4
WBONE20	90%	subclass. Not in AHC data

Dataset	Database	Missing/Comments
BAEP	BAEPDD	Not much known about this database, haven't seen anything in the coding, need to check with Dr. Kohler. Missing definitions for "dd" column and "species" column needs to be expanded (ie. Jun becomes juniper). Replace "." with n/a?
	BONETOOLS	"plan view" and "crossscn" and "tiptype" columns need expanded (shortform in columns presently). Columns ending with \$ coded as "a" or "p". Need to check to see if there is an expanded definition for these.
	CER92	SUNUM Sugrid need northing/eastings? (values such as 98). Suhoriz?
	CER92_RIMS	SUNUM, suhoriz?. Sugrid northing/eastings? "sitea" column looks like it could be expanded. Need to check
	FAUNA96	SUNUM, suhoriz?. Sugrid northing/eastings?
	FLD92	SUNUM, suhoriz?. Sugrid northing/eastings? "sitea" column looks like it could be expanded. Need to check
	FLT92	"material" "condition", "primary_func", "gpfc" columns need expanded SUNUM, suhoriz?. Sugrid northing/eastings? "sitea" column looks like it could be expanded. Need to check a couple columns
	MACROBOT96	Sunum, sugrid, bs columns. Taxon and condition columns need expanded. A couple column titles need explained in the codebook to be more understandable
	NFLT92	"morphou", "sutype" needs expanded definitions. Sunum and sugrid
	CODEBOOK	n/a

## Appendix VIII Penn State Report

  
**PENNS STATE**  
**INVOICE**

RECEIVED  
 JUL 07 2008  
 RSSP

**UNIV OF ARKANSAS**  
**RES SUPPORT & SPON PROG**  
**ATTN: SUBAWARD ADMIN**  
**120 OZARK HALL**  
**FAYETTEVILLE, AR 72701**

Invoice No.	41NC0-5
Date	06/30/2008
Contact	Marty Gillespie
Telephone	814-865-2680
Fund Name	UNIVARKANSAS SNOW
Award Amt	\$22,024.00
Account	04-021-12 UP 41NC0
Exception	CERT

Agreement: SA0809241		Document No.
Comment:		
Category	Expenditures for Month Ending May 31, 2008	To Date
GRAD ASST.	1,588.50	15,885.00
<b>TOTAL SALARIES, WAGES &amp; GRADS</b>	<b>1,588.50</b>	<b>15,885.00</b>
DOMESTIC TRAVEL	0.00	2,459.88
FOREIGN TRAVEL	0.00	1,011.70
GRAD ASST FRINGE BENEFITS	249.39	2,493.91
<b>TOTAL DIRECT COSTS</b>	<b>1,837.89</b>	<b>21,850.49</b>
<b>TOTAL COSTS</b>	<b>1,837.89</b>	<b>21,850.49</b>
<b>NET COSTS</b>	<b>1,837.89</b>	<b>21,850.49</b>

**Please Pay This Amount: \$1,837.89**

CERTIFICATION: I certify that all payments requested are for appropriate purposes and in accordance with the agreements set forth in the application and award documents.

Marty Gillespie

*M. Gillespie*

cug4

Send copy of invoice along with your check made payable:  
**The Pennsylvania State University**  
 Research Accounting, 227 W. Beaver Ave Ste 401 State College, Pennsylvania 16801-4819

08/14/2008  
11:08:49

Report: UnivRstr

Account: 04-021-12 UP 41NCO  
Account Name: UNIVARKANSAS SNOW  
Sponsor: UNIVARK  
Grant: SA0809241  
Fringe Rates -- Acad: 28.60 Non Acad: 28.60 Grad: 13.40 Wage: 8.30 Student: 0.60 F&A Rate: 0.00 Cost Share: 0.00

The Pennsylvania State University  
Financial Information System

Rpt Desc: University Report/Restricted

Fiscal Yr: 2007/2008 (Ctd)

Page 1

Line Description	Total Budget	Total Encm	Actual	Balance
INCOME				
0040 Other Income	22,024.00	0.00	15,539.57	6,484.43
TOTAL INCOME	22,024.00	0.00	15,539.57	6,484.43
EXPENSES				
SALARIES				
0164 Grad Asst-Academic	15,884.00	0.00	15,885.00	-1.00
TOTAL SALARIES	15,884.00	0.00	15,885.00	-1.00
WAGES				
TOTAL SALARIES & WAG	15,884.00	0.00	15,885.00	-1.00
DEPARTMENTAL ALLOTMENT				
033B Meals & Travel	3,900.00	0.00	0.00	3,900.00
0337 Travel CONUS	0.00	0.00	2,459.88	-2,459.88
0338 Travel O-CONUS	0.00	0.00	1,011.70	-1,011.70
0433 Grad Fringe Benefits	2,240.00	0.00	2,493.91	-253.91
TOTAL DEPT ALLOT	6,140.00	0.00	5,965.49	174.51
CAPITAL EQUIPMENT				
INDIRECT COSTS				
TOTAL EXPENSE	22,024.00	0.00	21,850.49	173.51
NET INCOME (INCOME-EXPENSE)	0.00	0.00	-6,310.92	6,310.92

Amount Due  
Penn State  
\$6,310.92

# Appendix IX University of York Report

Page 1 of 1

**Finance Department**  
University of York, York, YO10 5NH  
Telephone: 01904 432120/434262  
Fax: 01904 434125  
Email: fees-office@york.ac.uk

## INVOICE

VAT Reg No: GB 647 2055 41

FRED LIMP  
UNIVERSITY OF ARKANSAS  
CENTER FOR ADVANCED SPATIAL  
TECHNOLOGIES  
JBHT 304  
FAYETTEVILLE AR72701  
United States

Invoice Number	34141
Account no	9132
Inv/Tax Date	11-Jul-2008
Order No	4030462
Your Ref	

DESCRIPTION	VAT%	AMOUNT IN USD
SOFTWARE SYSTEM DESIGN AND METADATA STANDARDS - CONSULTING SERVICES 3/19/08 AND 6/30/08	0.00	3,000.00
<b>PAYMENT DUE BY 08-Aug-2008</b>		
Sub Total		3,000.00
VAT		0.00
Cheques: Payable to "The University of York" [in Esterling on UK bank account]		
BACS: HSBC PLC 40-47-31 20898201 Bank Transfer: SWIFT MIDLGB22 HSBC PLC 40-47-31 20898201 IBAN GB37MIDL40473120898201 BIC MIDLGB2109Y	TOTAL DUE in US Dollar	3,000.00

If this invoice has been paid in advance please accept this as a VAT receipt

Queries: please contact a member of the Fees Office on 01904 432120/434262  
email fees-office@york.ac.uk

When making a payment always quote ACCOUNT NO and INVOICE NO

## Appendix X SRI Report



ARCHAEOLOGY • ANTHROPOLOGY • HISTORY • HISTORIC ARCHITECTURE

August 27, 2008

W. Fredrick Limp  
Director Center for Advanced Spatial Technologies  
JBHT 304  
University of Arkansas  
Fayetteville, AR 72701

Dear Dr. Limp:

Statistical Research, Inc. (SRI) was responsible for three major tasks: (1) the creation of a spatially referenced digital document archive; (2) analysis of sociological issues in the CRM community; and (3) development of a business model. The first, in consultation with the AOI steering committee, was abandoned after Dr. Clay Mathers left SRI and when it became apparent that completion of item 3 required more effort than originally anticipated. The second was completed by Dr. Christopher Dore. Dr. Jeffrey Altschul spent considerable effort on the business model, largely completing the task. It should be noted, however, that the business model continues to be modified, long after completion of the grant.

The SRI budget for the Digital Antiquity Project was \$21,801.00, \$17,901.00 in labor and fringe benefits, and \$3,900.00 for travel. The total billable amount expended by SRI was \$8,289.41.

The SRI labor amount anticipated that the work would be performed by Clay Mathers and Christopher Dore. Dr. Jeffrey Altschul took over the project after both Mathers and Dore left the company. Although Dr. Altschul spent substantially more time than budgeted for Dore and Mathers combined, as Chairman of the Board, Dr. Altschul does not have an auditable billing rate. Therefore, SRI did not bill the project for his time; however, his expenses were billed. The total billable labor with fringe benefits expended by SRI for the project is \$2,878.51.

The additional \$5,410.90 expended was for travel to the following locations:

Denver, CO (Mathers) August 24-26, 2007  
Sante Fe, NM (Altschul) February 14-19, 2008  
New York, NY (Altschul) June 9-12, 2008  
Washington, DC (Altschul) June 17-19, 2008

Sincerely,

Peter G. Fox  
Chief Financial Officer

CALIFORNIA  
Redlands  
21 W. Stuart Ave.  
P.O. Box 390  
Redlands, CA  
92373-0123  
(909) 335-1896  
(909) 335-0808 (fax)

San Diego  
3990 Old Town Ave.  
Suite B-102  
P.O. Box 82404  
San Diego, CA  
92138  
(619) 299-9766  
(619) 299-9774 (fax)

Woodland  
211 Court St.  
Woodland, CA  
95695  
(530) 661-1400  
(530) 662-5500 (fax)

ARIZONA  
Phoenix  
P.O. Box 27748  
Tempe, AZ  
85285-2681  
(480) 774-1920  
(voice & fax)  
(480) 600-8692 (cell)

Tucson  
6099 E. Speedway Blvd.  
P.O. Box 31865  
Tucson, AZ  
85751-1865  
(520) 721-4309  
(520) 298-7044 (fax)

NEW MEXICO  
Albuquerque  
4425 Juan Tabo Blvd. NE  
Suite 112  
Albuquerque, NM  
87111-2681  
(505) 323-8300  
(505) 323-8314 (fax)  
(505) 331-2491 (cell)

TEXAS  
El Paso  
8201 Lockheed Dr.  
El Paso, TX  
79925  
(915) 781-2200  
(877) 781-2205  
(915) 781-2201 (fax)

PACIFIC NORTHWEST  
Burnaby, British Columbia  
5331 Meadevale Dr.  
Burnaby, British Columbia  
Canada V5B 2E6  
(604) 298-2701  
(voice & fax)

[www.sricrm.com](http://www.sricrm.com)



## Appendix X

### University of Arkansas Aggregate Fiscal Report

10/15/2008 20:31 4795757494

RESEARCH ACCOUNTING

PAGE 01/01

UNIVERSITY OF ARKANSAS  
FAYETTEVILLE, ARKANSAS

Cost Center: 0402-71051-21-0000  
Award Number: RSSP 07-000196  
"ANDREW MELLON FDN/DIGITAL ANTIQUITY"

Dr. Fred Limp

FOR THE PERIOD: 07/01/07 - 09/30/08

<b>REVENUE</b>			
GRANT	152,000		
INTEREST REVENUE	3861.09		
<b>TOTAL REVENUE</b>	<b>\$155,861.09</b>		
<b>COST ELEMENTS</b>	<b>EXPENDITURES TO DATE</b>	<b>Budget LTD</b>	<b>Budget Remaining</b>
SALARIES & WAGES	28,832.49	17,658.00	(11,174.49)
FRINGES	7,496.44	4,591.00	(2,905.44)
TRAVEL	25,538.64	19,140.00	3,601.36
COST DIRECT & SUPPLIES	6,087.43	16,964.09	10,876.66
Washington State	21,395.27	12,388.00	992.73
Penn State	21,850.49	12,024.00	173.51
Arizona State	17,615.34	18,295.00	679.66
University of York	3,000.00	3,000.00	0.00
Statistical Research	8,289.41	11,801.00	13,511.59
<b>TOTAL DIRECT COSTS</b>	<b>\$140,105.51</b>	<b>\$115,861.09</b>	<b>15,755.58</b>

I certify to the best of my knowledge that all expenses are  
for appropriate purposes and in accordance with the agreements set forth  
in the application and award documents. The documentation required by  
current Federal regulations for these costs is on file and remain  
available for audit by the University of Arkansas and the appropriate  
Federal auditors for the period required by Federal regulations.

Signature Stephen Guman Date: 10/15/08