To: Archaeoinformatics Technical and Disciplinary Advisory Board Members
From: Keith Kintigh, Steering Committee Convener
Re: Archaeoinformatics.org Advisory Board Meeting February 15-18

First, I'd like to again thank you all for agreeing to participate with us in this effort to build an information infrastructure for archaeology. Following this cover memo is a preliminary agenda and a list of participants. A brief report on the activities archaeoinformatics.org since its formation about a year ago will follow. I have also attached a copy of our proposal, funded by the Andrew Mellon Foundation, which is underwriting this effort.

In many ways, archaeoinformatics.org is a follow-up to a 2004 NSF-sponsored workshop held in Santa Barbara and hosted by the National Center for Ecological Analysis and Synthesis. That workshop focused on problems of access, integration, and preservation of digital databases. Several of our participants in this Santa Fe meeting participated in that workshop. Archaeoinformatics.org is concerned with a broader scope that includes text, images, geospatial and more exotic forms of data (such as from LiDAR and 3-D artifact scans). However, as you will see from our proposal to the Mellon Foundation, the conclusions of the Santa Barbara workshop have substantially informed our effort.

We'd ask that, if at all possible, you read the proposal and the report of the Santa Barbara workshop, The Promise and Challenge of Archaeological Data Integration <http://archaeoinformatics.org/articles/Kintigh2006CyberinfrastructureAmAnt.pdf> published in *American Antiquity*. I also recommend a related article by Dean Snow published in *Science,* Cybertools and Archaeology <http://archaeoinformatics.org/articles/CybertoolsandArchaeology.pdf>. (Other relevant references are provided at http://archaeoinformatics.org.)

As outlined in our proposal, this summer the Archaeoinformatics.org Steering Committee plans to submit an implementation proposal to the Mellon Foundation. The proposal will not only need to outline the functional requirements and some core technical specifications, it will also need to present a persuasive plan for promoting acceptance and use of the infrastructure by the discipline and a financial model that can sustain it in the long term. Our goal for the Santa Fe meeting of the Disciplinary and Technical Advisory Boards is to report on our progress to date and to get key input that will help guide the development of a successful implementation of a broad-based digital information infrastructure for archaeology.

We recognize that the success of this effort depends simultaneously on functional capabilities and technical qualities, on acceptance and use by the discipline, and on fiscal sustainability. We also recognize that there will continue to be important roles for more specialized cyberinfrastructure components. Nonetheless, we hope to be

able to make substantial inroads into the compelling problems of access and preservation for both digital data and the gray literature.

There are several important international efforts in archaeology—some much farther along than ours—and quite a few additional national and international cyberinfrastructure initiatives, both generalized and residing in specific disciplinary domains. While we have investigated some of these, we will be seeking your advice on how take advantage of the accomplishments of and to plan and prioritize integration with these other efforts. If you have a chance it would be particularly worthwhile to investigate England's Archaeology Data Service web site (http://ads.ahds.ac.uk/).

Also on the agenda is a demonstration of ASU's related NSF-funded effort called tDAR (the Digital Archaeological Record; http://tdar.org). As a part of this package I have provided a brief memo that describes that related, but distinct, project.

We look forward to meeting those we do not know and to seeing all of you at the Hotel Santa Fe on the evening of February 14th (with our apologies to spouses and significant others for the timing). I'm sure that you'll enjoy Santa Fe. Please let us know of any ideas or concerns.

Keith Kintigh

JOINT MEETING OF THE ARCHAEOINFORMATICS.ORG STEERING COMMITTEE, DISCIPLINARY ADVISORY BOARD AND TECHNICAL ADVISORY BOARD
February 15-18, 2008
Hotel Santa Fe, Santa Fe, New Mexico

TENTATIVE AGENDA – FEBRUARY 13, 2008

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 (at Montezuma, 1 blk W of Guadalupe)

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 (Kiva B)
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

PARTICIPANT LIST

Brief bios and email addresses at http://archaeoinformatics.org

Steering Committee

Keith Kintigh, Arizona State University (convener)
Jeff Altschul, Statistical Research, Inc.
Tim Kohler, Washington State University
Fred Limp, University of Arkansas
Dean Snow, The Pennsylvania State University

Working with the Steering Committee

Debbie Harmon, University of Arkansas
John Howard, Arizona State University
Allen Lee, Arizona State University

Disciplinary Advisory Board

Brian Crane, Versar, Inc (Springfield, Virginia)
Katherine (Kitty) Emery, University of Florida
Eric Kansa, University of California at Berkeley, Alexandria Archive Institute
Francis (Frank) McManamon, National Park Service (Washington DC)
Fraser Neiman, Monticello, University of Virginia
Vincas (Vin) Steponaitis, University of North Carolina
Phillip (Phil) Walker, University of California at Santa Barbara
Willeke Wendrich, University of California at Los Angeles
Thomas Whitley, Brockington & Associates, (Norcross, GA)

Technical Advisory Board

Worthy Martin, University of Virginia
Julian Richards, Archaeology Data Service, University of York
Herbert Van de Sompel, Los Alamos National Laboratory

TOPICS FOR BOARD MEMBER PRESENTATIONS AND DISCUSSION

Phil Walker – Could you briefly describe 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 – Could you, very briefly, present the core elements of the Archaeology Data Service effort, including a comment on how you are linking the data archives to publications. Please include a brief discussion of the success of ADS's efforts to establish best practices relative to digital data.

Eric Kansa – You've done some useful thinking about and development of software solutions to some the core sociological problems that a successful infrastructure must face. Could you outline what you see as the most important of these? Please include a discussion of enhanced citation or other credit accruing to digital publication.

Fraser Neiman – Based on your experience with the Digital Archaeological Archive of Comparative Slavery, what are the key elements to achieving archaeological community buy-in to an information infrastructure? What have been the biggest challenges? What do you see as the most important elements in your successes?

Worthy Martin – Understanding that implementation of a cyberinfrastructure can be staged, do you see 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? Are there elements that are too conservative or too much at the bleeding edge?

Herbert Van de Sompel – How do you assess the opportunities and advantages vs. the costs and disadvantages of achieving interoperability with other science informatics and digital library efforts? (John Howard, especially, can also chime in on this discussion.)

Frank McManamon – Could you discuss the two or three most important values that an integrated text and data information infrastructure could bring the mission of the National Archaeology Program? Can you see ways in which those values can be "sold" to agencies? In light of previous governmental initiatives, what would be required for them to adopt a solution from the outside?

Vin Steponaitis - Could you provide a brief statement of what you see as the most important values a cyberinfrastructure could bring to the academic research community? How can the contributions to this infrastructure be rewarded in ways that "count" in an academic setting?

Brian Crane – 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? What aspects of an information infrastructure could simultaneously serve the interests of DOD and its contractors?

Tom Whitley – From the standpoint of the practice of CRM, what do you see as the two or three most important values of a cyberinfrastructure and what do you see as the two or three largest impediments to its adoption? How could this infrastructure improve efficiency or effectiveness of cultural resource management as it is practiced in the US? How might we convince people of these advantages?

Willeke Wendrich – Do you see large-scale, long-term, international multi-investigator research as posing particular challenges to or presenting particularly persuasive opportunities for an information infrastructure for archaeology?

Kitty Emery – Could you reflect on the values of a research infrastructure such as we propose here to the missions of anthropology or natural history museums? How might we best proceed to attain positive participation from such institutions?

BREAKOUT GROUP DISCUSSION TOPICS

On Saturday we will break up into four groups. Groups will reconvene briefly on Sunday. We will ask that each group consider one of these clusters of topics on Saturday and a different one of Sunday. We'll all hear reports of all the groups, so there will be opportunities for everyone to comment on all topics. A review of these topics, however would give you a good idea of the issues on which the Steering Committee thinks it needs your advice. Of course, you should also bring up other topics you consider important.

A. Organization and Logistics

A.1. Organization

A.1.1. Assuming a reasonable financial base, how should the ongoing operation of the infrastructure be organized?

A.1.2. Should there be a physical center?

A.1.2.1. If so, would it best be managed by some existing institution (e.g., a university)?

A.1.2.2. How should one go about choosing the location/institution of the center?

A.1.2.3. Are there good models that we should consider?

A.1.2.4. What should be the mission of the center.

Should it just be an infrastructure provider?

Should it actively seek data ?

Should it sponsor integrative research, as NCEAS (the national Center for Ecological Analysis and Synthesis) does?

A.1.3. Could the infrastructure we are contemplating run, in the long term, without a center?

A.1.3.1. Are there models for that?

A.1.4. Apart from whether or not there is a center, how should the infrastructure be governed (e.g., should the goivernance have a multiinstitutional component)? (The current plan is laid out in archaeoinformatics.org's proposal to the Mellon Foundation.)

A.1.4.1. How can an executive body be maintained that is seen to have legitimacy by the professional community, that the potential for some permanence but the flexibility to respond to changing conditions?

A.1.5. How can the advisory boards best be organized? (The current plan is laid out in archaeoinformatics.org's proposal to the Mellon Foundation.)

A.1.5.1. What are the proper roles for the boards?

A.1.5.2. What is the best size?

A.1.5.3. What sort of make-up should be stipulated?

A.1.5.4. What selection process should be used?

A.1.6. Give the considerable success of the ADS and English Heritage in addressing many of these same concerns, what organizational components should we consider adopting? How does the different situation in US (e.g. more diffuse organizational structure, lack of national system, etc.) mean that some lessons learned from ADS won't be applicable or would need to be modified?

A.1.7. What if any role should the key professional societies or RPA (Register of Professional Archaeologists) play in the organizational structure?

A.2. Private Data.

A.2.1. We've drawn a distinction between data sources that are designated private in the sense that they are shared with only a specific designed circle of user-colleagues) and types of data that are confidential (such as site locations) and to which access might always need to be restricted to a class of qualified archaeologists. Our idea is that allowing people to maintain data in the infrastructure in a state that is not publicly accessible would encourage the registration of data sources and metadata relatively early in the research process, increasing the chances both that the data are registered at all and the the metadata are of high quality. Nonetheless, archaeologists' professional ethics are clear that archaeological data may not be kept private indefinitely.

A.2.1.1. One approach is to allow users to specify a date (with an upper limit), so that at most after X years, a private resource becomes public. (Such a date might be specified in a grant or contract.)

A.2.1.2. A variant on that approach would be to allow the dataset owner to continue to move the expiration date for privacy forward, with the condition that at no time could it be more than Y years from the present.

A.2.1.3. Do either of these appear to be reasonable alternatives? Can you recommend another? If there is to be a limit, what would be a reasonable period?

A.3. Confidentiality

A.3.1. What are the archaeological concerns with respect to the confidentiality/sensitivity for an archaeological information infrastructure. Site location is one obvious one. Within-site locations of cemeteries can also be sensitive. What other sorts of data may be sensitive?

A.3.2. Some people suggest that the way to protect archaeological resources is to put signs on them urging preservation rather than trying to hide them; in the US the emphasis has generally been on protecting site location information. If site location information is to be restricted (as many agencies would doubtless require):

A.3.2.1. What is a reasonable process to decide who has legitimate access?

A.3.2.2. What sort of body might be generally accepted as likely to do a fair job of vetting requests. Would the decision-making have to be state/regional or could it be national? Is there a role for RPA or the professional societies here? Would they accept it?

A.3.2.3. Is there some role for community "policing" here.

A.3.3. For site location, on output, is reducing the precision of the data (for example, by adding noise or rounding right-hand digits of coordinates a reasonable way to to make cruder data public even if the precise data is restricted. How much has to be rounded? Are there other technical solutions? How might they work?

A.4. Who Controls Access Control.

It seems likely that there may be entities (e.g., agencies or tribes) who control data that it would be advantageous to include and who might be useful partners, yet who would insist on elaborate access control mechanisms and an access control policy that each would control. While this may be technically possible to achieve,

it would lead to a system that is probably more cumbersome and in which legitimate and trustworthy users pursing large scale or synthetic research would practically be unable to get at the data they need because of bureaucratic impediments. Further, implementing this would require substantially more human administration and additional technical complexity. We'd appreciate comments on this dilemma. We already see a couple of possible arguments here.

A.4.1. One is that the infrastructure develop a policy and procedure that does a reasonable job of protecting confidentiality and that is develop the overall functionality of its product (the infrastructure) in a way that is *so attractive* that an agency would decide that it can't afford not to take advantage of it (of course we ant to do this anyway).

A.4.2. Another is to accommodate these requests with the hope we can at least get the data into the system and hope that (with urging) the agencies will see the light and open up their data to a shared access policy.

A.4.3. Do you see this as an issue? Are there other approaches? Are there way's to finesse this potential problem?

B. Business Model

(Statistical Research, Inc.'s report provides a good basis for discussion)

B.1. Assuming that we can obtain development and startup funding from the Mellon Foundation, NSF, or similar sources--these are the general questions, better specified below:

B.1.1. How do we fund long term operations?

B.1.2. How do we fund long term development and improvement?

B.2. Demontrating Value.

B.2.1. We believe that an operational infrastructure where, one way or another, large fractions of information resources generated are registered, will save money and improve research. How can we most effectively demonstrate that?

B.3. Elements of a Business Model. What elements should we consider for a sustainable business model?

B.3.1. Curation fee premised on a curation responsibility for the data. (While we might most obviously think that the infrastructure would be *the* repository, it could also provide repository services at a cost for museums with responsibilities to curate digital data.)

B.3.1.1. How do we convince archaeological sponsors (i.e., grant institutions and government permitting/contracting agencies) that data are part of their curatorial responsibility?

B.3.1.2. How realistic is it to think we can get the requirements from granting or regulatory agencies that would be needed to effect curation fees?

B.3.1.3. What steps can be taken to get this accepted as an contract/grant obligation or ACHP recommendation or DOI standard?

B.3.1.4. What would be the elements of a reasonable fee structure.

B.3.1.5. Could there reasonably be exceptions to a fee, e.g., for unfunded or clearly underfunded projects (e.g. much thesis research).

B.3.2. Subscription or per-use fee premised on the value of data going out.

B.3.2.1. How would one price this to individuals or institutions or agencies?

B.3.2.2. How likely would different sorts of individual and instuitutions (e.g., firms or universities) be to actually pay these fees or these subscriptions?

B.3.2.3. Can this be established in such a way that the value will be recognized and the revenue stream is there before start-up funding ends?

B.3.2.4. If we go this route, would we still need requirements from agencies that data be entered?

B.3.3. Annual agency support premised on reduced agency costs and greater services or simply enhancing research value

B.3.3.1. How likely would agencies be to come up with the money for this?

B.3.3.2. What steps could be taken to drum up this support?

B.3.4. Mixed model. For example, costs on both ends with reduced costs for institutions that agree to register the information resources deriving from all of their funded projects..

B.3.4.1. Suggestions here?

B.3.5. Endowment.

B.3.5.1. Where might money come from?

B.3.5.2. What entity could hold an endowment?

B.3.6. Other Revenue Sources. There may be a revenue source in providing specific products from the infrastructure to, for example, agencies.

B.3.6.1. What sorts of data-based information services might have this value?

B.3.6.2. Is there potential revenue to be derived for somehow facilitating data entry or improving data quality at or near registration time (off-loading effort from the data holder to the infrastructure).

B.4. Digital Rights Management

B.4.1. What are the elements of digital rights management that we must accommodate?

B.4.1.1. Is there an opportunity, for example, to index JSTOR or an electronic for-profit journal so that relevant articles can be located, but with an attempt to download such an article taking you to a site where you have to pay for it? There seem to be an advantages to scholarship in general as well as to the journal and the infrastructure.

B.4.1.2. Should this be pursued. What should be the priority?

B.5. Enhancing Adoption

B.5.1. Where are the specific areas that application of the approach will have the most intellectual and financial value to the field? Are there some areas that should be considered first because of their greater "bang for the buck?"

B.5.1.1. What would be the characteristics of a range of compelling pilot studies that would best demonstrate the value of the infrastructure.

B.6. Credit

B.6.1. Incentives for registering data could come from increased citation of one's efforts, both published and data sources. Similarly agencies or granting organizations would gain some credibility for the expense of archaeological research if it can be shown that the existing and newly generated information is being used.

B.6.1.1. What actions can the infrastructure take to enhance the value of this credit to individuals or institutions?

B.6.1.2. What steps can other parties, such as employers or professional organizations, take to enhance the value of the credit?

C. Technical Issues and Data Standards

C.1. Data Standards

C.1.1. To the extent that the infrastructure has a mission to enhance comparative research, we would hope that there would be some practical way to keep out clearly bad information, particularly in databases.

C.1.1.1. Who should be allowed to contribute information resources?

C.1.1.2. Would some sort of community tagging or rating/reviewing of information resources be valuable?

C.1.2. To the extent that agencies would require use of the information infrastructure to make available (to the public and the professional community) reports and databases, they might wish to impose some standards on the quality and completeness of the metadata. They might for example review the metadata records as they now do the reports.

C.1.2.1. How could this be operationalized to insure that there is a level playing field for participants who are in some sense competing with one another?

C.1.3. Should less rigorous and more variable standards apply where there is no permit or contractual obligation of for entering orphan or legacy data?

C.1.3.1. What would be considered a bare minimum of metadata?

C.2. Technical Consideration

C.2.1. What technical issues/solutions/approaches do the board members suggest that the project consider - especially any not already under our review or consideration?.

C.2.2. Are there any special issues/problems identified in the survey (or other sources) that a specific technical strategy could address.

C.2.3. Archaeology has a substantial but complex and disputatious history in considering semantics and ontology. Much work in CS has focused on tools and approaches to both. As a result - how should we approach these issues today?

C.2.4. How do you evaluate the tradeoffs of a centralized vs a peer to peer model for this infrastructure. (tDAR like GEON anticipates a central repository data and metadata that accepts but obviously cannot guarantee the longevity of distributed data. It allows for the possibility of separately "branded" but fully inter-operable and interconnected "central" nodes each running the same core software stacks.)

C.2.5. What other current efforts (digital libraries, interoperability etc.) have particular relevance to this project?

C.2.5.1. Which would it be most useful for us to consider or contact?

C.2.5.2. While we want to design for operability with other informatics efforts, what is the payoff for implementing that sooner vs. later.

C.2.6. Do you have any reservations about sticking to open-source development?

C.2.7. What are the choke points in any reasonable system we might establish and how do we deal with them?

REPORT: UNIVERSITY OF ARKANSAS

Administrative coordination

The University of Arkansas (Uark) serves as the "prime contractor" for the current Mellon grant. Upon the award sub-contracts were executed with Arizona State University, the Pennsylvania State University, Statistical Research, Inc. and Washington State University. Additionally, Uark is the lead and fiscal agent for travel arrangements for the February meeting in Santa Fe.

Web development and hosting

The Uark has been responsible for the development, management and hosting of he consortium web site archaeoinformatics.org. The site currently provides background information on the initiative including the objectives, current status, participants and a substantial amount of relevant information including a number of key publications (and links) to relevant articles and books as well as an archive of the consortium's virtual lecture series (see below). As of Jan 31st the video files of the lectures had been downloaded in aggregate 268 times.

Virtual Lecture series

The Uark has been responsible for the organization and operation of the Consortium's Virtual Lecture series. The series began on March 26, 2007, prior to the grant award. The website has streaming video of the lecture as well as the original Power Point presentations for each lecture.

Archived Lectures that are available at archaeoinformatics.org are:

1. April 9, 2007
Eric C. Kansa
Executive Director of the Alexandria Archive Institute
"Open Context:Community Tools for Publishing Research Data on the Web"

Kansa, discussed the archaeological project OpenContext - an ArchaeoML based system for sharing diverse, nonstandardized data and media.

2. April 23, 2007
Chaitan Baru
Director of Science Research and Development at the San Diego Supercomputer Center *"GEON: Geosciences Network"*

Baru spoke on GEON, geology's successful analog to some of what we believe archaeology needs to accomplish.

3. September 19, 2007
Michael J. Halm (1), John Yoo

(1)Senior Strategist and Manager for the Special Project activities for the Teaching and Learning with Technology group, Penn State University,
"LionShare: Secure P2P File Sharing and Collaboration"

Halm 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.

4. October 17, 2007
Mark Gahegan (1), Chaitan Baru, Boyan Brodaric
(1) Professor of Geography and affiliate professor of Information Science and Technology at the Pennsylvania State University
"Sharing our resources, sharing our understanding: Cyberinfrastructure for Archaeology"

Gahegan is a GEON Co-PI and has worked on other cyberinfrastructure projects in the fields of plant pathology, e-education and human-environment interaction. This talk introduced 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 highlighted 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.

5. October 31, 2007
Fred Limp
Leica Chair and Director 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

6. 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 described 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.

7. November 28, 2007
Julian D Richards,
Professor of Archaeology, University of York and Director, Archaeology Data Service
"Current challenges for digital preservation and delivery"

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

8. 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), for the Electronic Cultural Atlas Initiative (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 - characterized more by the fascinating variety of its members than the focus of its mission.

the Digital Archaeological Record

Arizona State University's effort has included intensive work on the National Science Foundation-funded[1] implementation of a prototype digital information infrastructure, called tDAR for "the Digital Archaeological Record." This effort is led by Professors Keith Kintigh and K. Selçuk Candan with assistance of Professors Katherine Spielmann, Margaret Nelson, Hasan Savulcu, Subbarao Kambhampati, Dr. Huiping Cao, and graduate assistants Yan Qi, Mallorie Hatch, and Ben Schoville.

The tDAR proposal was submitted to NSF before the archaeoinformatics.org proposal to the Mellon Foundation, but was funded after the formation of archaeoinformatics.org. As a consequence, the NSF-funded tDAR prototype implementation grant has proceeded in parallel with the archaeoinformatics planning grant. The design of tDAR has been substantially influenced, but not dictated by, archaeoinformatics.org discussions. While the Mellon Foundation is not directly funding the development of tDAR, tDAR is very much directed to addressing the goals of archaeoinformatics.org and illustrates of one possible set of software design decisions.

While tDAR's focus is on systematically collected datasets (databases), as implemented, tDAR readily archives text and images. The particular challenge that it addresses is the semantic integration of archaeological data collected by different projects using inconsistent coding schemes. tDAR is intended as an open-source, Internet accessible foundation for a global archaeological information infrastructure. In its full implementation, tDAR aims to provide integrated, cross-project, sustainable, Internet access to dynamic archives of archaeological, physical anthropological, and environmental data that will advance our ability to conduct synthetic and comparative research and preserve the long-term utility and accessibility of archaeological data. tDAR incorporates some elements of geosciences' GEON platform. While tDAR is based at ASU, its servers reside at the San Diego Supercomputer Center. More information is available at the project web site, http://tdar.org.

ASU is now in the final phases of internally testing the soon-to-be-public beta release of Version 1 of tDAR, which will be demonstrated in Santa Fe. Version 1's web interface provides the ability to upload information resources in a variety of database, text, and image formats. Web registration of an information resource includes not just uploading the raw file, but also an interfacse that facilitates the entry of the detailed metadata needed to the resource. Part of the metadata acquisition is a facility to attach "coding rules" to databases, providing translations of numeric codes into searchable natural language labels. It also provides metadata and information resource content based search and, of course, download capabilities.

[1] National Science Foundation Grant No. 0624341. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Relation databases (and spreadsheets) are maintained in their original forms but also converted to postgreSQL so that they can be effectively maintained and readily searched. The next phase of tDAR's implementation focuses on semantically based information integration capabilities. The semantics will be actualized (with the assistance of a groups of experts) for subdomain of archaeology (archaeological fauna) with numerous, diverse datasets from two areas of the US.

A poster on tDAR was presented at the recent Southwest Symposium, attended by approximately 250 Southwestern Archaeologists.[2]

Mexico Case Study

Under the current, Mellon Foundation grant, archaeoinformatics.org will develop a proposal to fund the implementation of an archaeological information infrastructure. An important component of the planned proposal will be the development of a number of "case studies" intended to persuasively demonstrate the research value of the system through their role in advancing research on substantive archaeological questions. Archaeological research in central Mexico has been identified as an excellent candidate for such a case study. This case study would be driven by a compelling research question, would have both international and multi-institutional participation, and would have a large component of legacy data.

ASU is responsible for drafting a design and preparing a budget for such a case study to be included in that proposal. To this end, with funding from the Mellon Foundation grant, Dr. Michelle Elliott (a recent ASU PhD), is beginning to interact with Mexican, US, and other International scholars working in Central Mexico (construed broadly). This effort has several objectives:

- to introduce these scholars to the archaeoinformatics initiative;
- to identify potential partners for a implementation case study;
- to work toward development of a compelling substantive topic and identify potential leaders and participants;
- through discussion and demonstrations of TDAR or other existing tools, to elicit special concerns or opportunities presented by international collaborations;
- to attempt to identify existing databases and unpublished reports (including *informes*) that these scholars would be willing to make publicly available (a special efforty will go into scholars that are retired or are soon to retire);
- where possible, to assist these scholars in registering datasets in tDAR.

In addition, Prof. Ben Nelson has agreed to make publicly available his substantial database that documents his many years of work at the site of La Quemada in Zacatecas, Mexico. With Mellon Foundation funding, Vincent Schiavitti (in consultation with Nelson) will be registering the complex La Quemada database with

[2] Benjamin Schoville, Mallorie Hatch, and Keith Kintigh, 2008. tDAR: Introducing "the Digital Archaeological Record." Poster presented at the 10th biannual Southwest Symposium, January 17-19, 2008, Tempe, Arizona.

complete metadata in tDAR. This will not only make publicly available a major archaeological dataset, it can be used by Elliott in demonstrating the value of contributing data and documents to an information infrastructure.

SAA Forum

For the upcoming Annual Meeting of the Society for American Archaeology in Vancouver, BC (March 26-30), Kintigh has organized a SAA Board-Sponsored forum designed to heighten engagement of the broader discipline in the objectives of the archaeoinformatics.org initiative. Participants include the Archaeoinformatics.org steering committee members plus Eric Kansa (University of California at Berkeley), Tom McCulloch (Advisory Council for Historic preservation), Fraser Neiman (Monticello), Charles Niquette (Cultural Resource Analysts, Inc.), and Julian Richards (University of York).

Digital Antiquity: Planning an Information Infrastructure for Archaeology

Abstract. The utility of archaeological data is severely diminished by its inaccessibility. Archaeology needs to take transformative steps to develop an information infrastructure that can provide Internet access to archaeological research documents, databases, and images. This infrastructure will require software development, a management structure and financial model, and real commitment by practicing archaeologists. Archaeoinformatics.org has been funded to develop a plan for such an information infrastructure. This forum offers an opportunity for discussion of the desirable capabilities of the infrastructure, how it should be managed and funded, and how the discipline can best be engaged in its development and use.

REPORT: THE PENNSYLVANIA STATE UNIVERSITY

Archaeological and other historical sciences depend heavily upon legacy data. This is because the record is finite and new data are difficult to acquire. Field archaeology actually destroys the physical resource as it generates new data, so archaeologists necessarily depend upon not only the data they acquire through their own efforts, but also upon data acquired by other projects for different purposes. The use of legacy data entails several practical problems, including variable systems of weights and measures, incommensurate databases, and missing metadata. These data may take many forms, such as field notes, field forms, photographs, maps, remote sensing imagery, and the specialized data sets produces by technology such as magnetometers, resistivity surveys, and ground penetrating radar. Laboratory analysis produces additional data, including GIS files, database files, word processor files, digital images, and PDF files. All of these must be made accessible to future researchers if legacy data are to be useful to them.

Archaeological collections are not themselves legacy data, but rather the sources of data. Conservation of sites and curation of collections is important, but not the focus of our efforts in archaeoinformatics. The continuing generation of data from curated collections and unprocessed assemblages will be much improved with the development of standards and protocols for data and metadata recording. However standardization going forward will not improve accessibility to legacy data already recorded. The latter exists largely in the form of incommensurate databases and published reports, voluminous but nearly inaccessible "gray" literature, and archival sources. In many cases the sites and collections from which they were generated no longer exist.

The Penn State team is working on one solution to part of this problem. We have created a search engine called ArchSeer, which can crawl published and unpublished text and generate an excellent return at high precision given reasonable queries. When coupled with projects to convert thousands of gray literature reports into PDF files with optical character recognition, the tool will open up a vast resource for archaeological researchers, agencies, and private individuals. With the cooperation and assistance of JSTOR we are currently using 75 years of *American Antiquity* back issues to test and refine ArchSeer. When it is fully tested we will release ArchSeer as open source software.

A second part of the Penn State contribution to the larger effort is image recognition. This involves both automatic map recognition and registration and object recognition in photographs. The latter is built upon SIMPLIcity, image recognition software that allows us to find objects in published or unpublished photographs without the benefit of tagging or cataloging. The software facilitates automatic recognition of things such as arrow points, and with sufficient training the system should be able to recognize specific point types. One version of this software can already trace the outline of an object and correctly classify it to one or another of a set of defined types within that class of objects.

REPORT: WASHINGTON STATE UNIVERSITY

At Washington State University, work on the Mellon Grant has been completed primarily by Dr. Timothy Kohler and Matt Bailey. Bailey was replaced by Charles Reed as Kohler's primary assistant in early January, although Bailey continues to provide answers to questions and technical support as time constraints from his new job allow.

Matt Bailey was responsible for the conversion of the Dolores Archaeological Project (DAP) data into the PostgreSQL format used on the Galisteo server for the Archaeoinformatics project. Since the conversion he, and now Charles Reed, have worked largely on updating the data dictionary for the various databases to convert the data from numerical format into more user-readable definitions, gathered from the dictionaries supplied on the DAP data CD as well as "The Dolores Legacy: A User's Guide to the Dolores Archaeological Program Data", compiled by Richard Wilshusen (1999). There are a few fields within the data for which it will not be possible to provide dictionary definitions for various reasons. These range from values that are not explained/not defined, to cases where Wilshusen is unsure of how the coding was done in the laboratory. Currently, work is waiting for some additional information on the Bunny10 database being supplied by the Anasazi Heritage Center, as well for a small bug to be worked out that creating a situation where dictionary definitions from one table are used in similarly named tables in other databases erroneously. Another problem that needs to be addressed is that of execution time out errors that have occurred when attempting to open the larger databases via a web interface; this is not a problem when using the command line interface.

Preliminary work is being done to convert the database compiled by the Bandelier Archaeological Excavation project into PostgreSQL format. From there, similar protocols will be used as for the DAP data.

Table 1. Summary of components of the DAP database.

DAP Databases		
Database Name	**Number of Variables**	**Size of original DBF**
Bunny10	23	142 KB
Ceram10	28	48.6 MB
Corn10	49	687 KB
Date10	33	2.08 MB
Debtg10	34	12.57 MB
Fauna10	32	21.02 MB
FeatInk	34	4.62 MB
Flt10A	29	15.45 MB
Flt10B	29	8.13 MB
Haft10	25	327 KB
Macro10	27	12.27 MB
Misc10	43	459 KB
Nflt10	30	7.83 MB
Ornmt10	22	93 KB
Photo01	20	13.46 MB
PolIn10	13	1.88 MB
Prov	23	15 MB
Prov10	27	16.6 MB
Rarerox	24	395 KB
Sample10	12	2.93 MB
Seds10	21	217 KB
Tscer4	88	126.35 MB
Tsflt4	87	47.28 MB
Tsnfl4	88	19.54 MB
Tsprv4	69	43.52 MB
Wbone20	40	2.02 MB

REPORT: STATISTICAL RESEARCH, INC.
FRAMEWORK FOR DISCUSSION : BUSINESS AND SOCIOLOGICAL ISSUES

To be considered a success, in addition to its technical merits, an archaeological digital data infrastructure must 1) be financially self-sustaining and 2) be adopted and used as an "official" repository. Technology, finance, and adoption are not independent, and must be considered together in project planning and implementation. For example, operating funds (finance) must be sufficient to allow state-of-the-art technology and to drive research in archaeoinformatics (technology). Adoption by a large segment of archaeological users (adoption) plays directly to financial revenues (finance). Adoption by a large user base (adoption) is partially dependent upon the technological function, interface, and ease of use (technology). Because of this interrelatedness, the technical aspects of the project cannot be undertaken in isolation from the financial and sociological components. The informatics project must be undertaken within the framework of a business plan.

A number of topics from the Archaeoinformatics survey and lectures have been identified that must be achieved to make the project a success. These topics of design and implementation can be considered to be lessons learned from other informatics projects and market research from a good sample of potential future users. This list of topics is not comprehensive, could be expanded, and should be prioritized. Additionally, how to implement these key success items is a topic for discussion, debate, and consideration. These key items include:

- Reduce technological barriers for data contributors and users through a well designed and intuitive interface.
- Quickly make available a critical mass of data that has future research potential and high archaeological interest.
- Adoption of digital archiving provision by regulatory agencies at the federal and state levels.
- Development of a fee structure to incentivize, not penalize, data contribution.
- Creation of a "level playing field" for private-sector firms.
- Change the professional value and reward system to provide professional credit for data authors and contributors.

Components of a Business Model

Any business model presupposes financial sustainability; put simply, revenues must be sufficient to cover expenses. Building a business model for the digital infrastructure project requires sound financial data, which we are in the process of collecting. Thus, while presenting a business model is premature, we can outline the major components of such a model. We structure the following discussion first by outlining anticipated costs and then examining possible revenue sources to cover those costs.

Costs

We anticipate three major cost areas: (1)technical costs in terms of hardware and software; (2) facility costs related to maintaining the national center; and (3) staff

costs of personnel to oversee the inclusion of adequate amounts and types of data and the distribution of these data upon request can be forecasted over time. A further assumption of the business model for the digital infrastructure project will be that expenses will change over the life of the project, and we need to adjust revenues accordingly.

Initially, substantial investments in infrastructure will be required. Expenditures in all three cost areas outlined above will be required prior to releasing the infrastructure for use. Over time, however, we expect costs to moderate and then stabilize. We anticipate the demand for adding databases per year will become relatively constant. Moreover, on-going costs to maintain the system—hardware will need to be updated, data validated and migrated, and tools created—can be anticipated at least one to two years in advance. It is important to remember that there is flexibility in the timing of maintenance costs. Of course, we need to stay abreast of Hardware and software upgrades, but we control when we actually make these changes, and therefore, when monies are expended. Similarly, the amount of data added to the system or tools created to manipulate data are dependent on the number of staff positions. Ideally, we will have a very quick response time, but there is also nothing wrong with creating a backlog of work (say on the order of one fiscal quarter), which will allow us to adjust workforce levels in advance of payroll or independent contractor expenditures.

Revenues

We divide revenue sources into five categories: sponsored research grants, compliance contracts, curation fees, user fees, and endowments. Because the exact mix of revenue streams will greatly influence the trajectory of the digital infrastructure project, the pros and cons of each source are examined below.

Sponsored research grants include funds from traditional sources such as the National Science Foundation, the National Endowment for the Humanities, and even the Mellon Foundation. These grants are generally focused on advancing archaeological research. The digital infrastructure project would be of interest because it provides a means of utilizing already collected data in ways that go beyond the limited goals of the research projects that collected them. Because sponsored grants will provide us with the least restrictive funding source from which to demonstrate the importance of the digital infrastructure on the future of archaeological research, their use in the beginning of the project will be essential to create "stakeholder buy-in" from archaeologists creating new data (e.g., from new field projects) and those using these data sources. It is absolutely essential that we convince the archaeological community that the performance of archaeological research in the future depends on a digital infrastructure.

Compliance contracts relate to funding available from government agencies responsible for managing cultural resources. Although we anticipate interest in the digital infrastructure at all levels of government—federal, state, tribal, and municipal—we expect that the federal agencies will be the most inclined to fund the project. Government agencies generate huge amounts of data that could more effectively be utilized in decision making if they were integrated. For example, the

Department of Defense (DoD) administers 42,000,000 acres within the United States. Of this total, DoD believes, but cannot verify, that it has surveyed about 12,000,000 acres and has found 150,000 archaeological sites. It would be of great value, indeed it is a legal requirement, for DoD to know exactly where the agency stands vis-à-vis its responsibility to inventory and evaluate cultural resources under its jurisdiction. Furthermore, the ability to query across DoD would provide data both to assist the agency in indentifying resources critically important to the country's cultural heritage as well as opportunities to compare and interpret resources across military installations. The paleoIndian boat building project is a good example. Through serendipity military archaeologists identified similar tool kits of stone tools used to build boats about 10,000 years ago at installations across the country ranging from Fort Drum in upstate New York to San Clemente Naval Air Weapons Station off the coast of California. Although the project is held up as a best practice in DoD's cultural resource program, project personnel have repeatedly stated that similar advances in the future depend on the ability to link military databases (see http://www.cemml.colostate.edu/paleo/projectsummary.htm)

Government agencies focus on cultural resources as they relate to their overall mission and legal requirements. Consequently, to invest in digital infrastructure, an agency would need to see an immediate return on their investment. The digital infrastructure project would have to evaluate the extent of overlap between agency goals and project objectives. Using a "case" approach to build the infrastructure may be a sound financial move, but it will require project principals to ensure that the project not be unnecessarily sidetracked.

Curation is both an ethical and legal requirement for most archaeological projects. Most repositories currently curate digital material, although standards and requirements for accepting and maintaining digital information vary greatly. In almost no case are repositories positioned to maintain digital information into the future; few ensure that digital media remains viable and few migrate digital material with changes in software and hardware.

Curation represents an opportunity and a challenge for the digital infrastructure project. We are cognizant that repositories are financially strapped and will be reticent to forego any fees they currently receive. At the same time, repositories and sponsors of archaeological research recognize the compelling challenges of curating and maintaining digital material and archives as well as the fact that they are currently failing to meet these challenges. We believe that by partnering with curatorial repositories we can accept and maintain their digital material and share in the curation fees. Of course, curation fees will have to increase. However, because these fees are already part of current archaeological practice, we will not have to create a new fee requirement. Our challenge will be to demonstrate to sponsors of academic and compliance research that to maintain their digital collections they must use the digital infrastructure. We are confident if we can make that case, increased fees will follow.

In addition to curatorial fees associated with placing information into the system, we might consider charging users of the system a fee to obtain data or to use tools that

we create to manipulate data. A current **user fee** that serves as a good analogue are access fees levied by many states that allow archaeologists in cultural resource management to access state site files. In these cases, archaeologists that meet certain requirements, such as holders of state permits or registered professional archaeologists (RPAs), pay an annual or as-you-go fee (the latter usually based on time and material expenses). For digital data, fees might be assessed annually, fixed fee per set, or fee by download size.

User fees are a sensitive issue in archaeology. Although we might expect major universities, museums, and government agencies to shoulder the burden of most of the fees, we anticipate that individual archaeologists will show great resistance. Because our ultimate goal is to have data used, we remain committed to insuring that user fees, if used at all, will not become an impediment to accessing the digital infrastructure as a research tool. A sliding scale fee structure in which large university departments and government agencies pay larger fee then smaller institutions and ultimately only modest fees would be assessed on credentialed individuals is a likely solution.

Endowments are an excellent means of assuring that certain fixed costs are always met. Although endowments can be established for the general operation of an organization, often it is easier to raise money for specific objectives; for example, endowing a permanent research program into new and innovative ways of using digital data in archaeological research. Regardless, endowments are generally secured through fundraising. Soliciting donations is easier once an organization has demonstrated its "proof of concept".

Before soliciting donations, we need to show that digital infrastructure is the way of the future for archaeological research and the digital infrastructure project is the best mechanism for the archaeological community to achieve this goal. Accordingly, we do not anticipate fundraising until after the digital infrastructure has been in operation for at least a year.

Matching Revenues to Costs

Our business plan will take into account not only aggregate costs, but incremental ones as well. We anticipate that initially, we will require a large infusion of cash to establish and staff a national center. In addition to physical costs, we also will budget for adequate staff time to incorporate into the digital infrastructure sufficient data of various types so that once the infrastructure comes on line it will be immediately useful to archaeological research. We suspect that most of this upfront cost will have to be absorbed by sponsored research grants.

Over time, we expect that government agencies will decide to make their data available through the digital infrastructure. Many agencies have on-going database projects, but there remains a critical need to integrate these databases that no federal agency has been tasked with or is willing to undertake. We expect that these "case studies" will occur periodically, each with their own funding source. We will need to work out policies for accepting contracts to ensure that the key objective of the digital infrastructure, that of serving data to advance archaeological research is met.

While the incorporation of major data sources will have to be funded by individual sponsors, a steady and reliable funding source will be required to maintain the system. Because these costs consist of staff salaries and hardware/software upgrades, they are relatively easy to project. Anticipated costs will then be offset by adjusting curatorial fees and user fees. To achieve long term sustainability, the project will look toward creating one or more endowments through fundraising.

The types and mix of revenue sources affect the staffing structure of the project. We anticipate that some positions, such as the executive director and administrative support, will be considered largely overhead expenses. Other staff positions will be project dependent. Our objective is to maintain a relatively constant staff structure so it is important that we accurately assess the number of research grants, compliance contracts, and the flow of curatorial and user fees. At the same time, we need to have an adequate number of project positions to generate sufficient overhead to fund administrative salaries and facilities. In the beginning, it will be critical to keep the number of administrative positions low, thereby minimizing overhead costs.

Determining the feasibility of sustaining the financial model will require us to match anticipated costs with revenues. Initially costs will be relatively easy to define: facilities, hardware and software, and personnel. The much more difficult part of the business model will be to assess: (a) the likelihood of sponsored research grants; (b) the willingness of government agencies to integrate their databases as a step toward fulfilling their legal compliance requirements; (c) the response by repositories to partnering with the digital infrastructure as a means of solving their digital curation crisis; (d) the willingness of the archaeological community to incorporate user fees as a necessary expense to conducting research; and (e) the potential of raising targeted funds through fundraising campaigns.

Business plans must be flexible. Although a good plan should identify all cost and revenue components, it is important that the plan not be tied too closely with particular expectations. The failure of any one funding source should not critical to the success of the project. For example, if we are unable to secure sponsored grants at the outset, we can accept more “case” studies from government agencies. However, the method of compensating for the loss of revenues will affect the trajectory of the project, and if we are not careful, the final character and composition of the digital infrastructure. It is critical, therefore, that the business plan faithfully encapsulate the mission of the project and chart a path for making financial decisions that assures that the mission is met.

REPORT: STATISTICAL RESEARCH, INC.
THE ARCHAEOINFORMATICS.ORG SURVEY

Methods

The survey was an internet-based survey. Participants were solicited by e-mail invitations to the Society for American Archaeology's Digital Interest Group members, individuals in the ArchaeoInfomatic.org network, and by those finding an invitation on the ArchaeoInformatics.org web site. It is important to note that participants self-selected themselves and, given the invitation methods, were biased towards those with an interest in digital data; this was not a random sample of archaeologists.

Respondents

A total of 445 individuals visited the survey. These visits resulted in 232 completed and 38 partial surveys. The plurality (48%) held masters degrees as a terminal degree. Thirty-six percent had received doctoral degrees and 16% held a bachelors degree. A wide range of archaeological experience was represented. Respondents claimed 3-56 years in the discipline; the mean was 19. The mean age of respondents was 44 years. Males dominated (66%) those responding. Eighty percent of respondents lived in the United States of America; almost all other respondents were from European countries, with the United Kingdom dominating (8% of total respondents). All eight major employment sectors were represented with academic departments (31%) ranking first followed by private for-profit firms (24%) and academic museums or research centers (10%). The primary research setting of the vast majority (70%) was North America.

Use of Technology

Respondents were skewed, in a self ranking of "geekness", towards the geek end of the spectrum. Web search engines, digital imagery, and databases on local computers were used very frequently. Geographic information systems and web-based databases were used with moderate frequency. Statistical software and 3D imaging were used infrequently. None of these technologies were unused by the majority. There was little difference in those using technologies versus creating data sets via individual technologies. Thus, it appears that users are active, data creating, users.

Access to Data

Respondents indicated that learning of the existence of other relevant research is not a significant problem. However, gaining access to the research materials of others is difficult. Databases are the most difficult to access followed by primary paper documentation, artifacts, and technical reports. Even 44% of respondents ranked obtaining technical reports as difficult (compared with 29% who stated it was easy). When asked about their anticipated use of potential tools to assist in accessing data, nearly half (48%) would use a tool to search gray literature on a weekly basis. Ranked second (45%) for use on a weekly basis was a tool that could display maps and images from gray literature. Ranked third was a tool that could search for unpublished

excavation databases. Fourth was the ability to search multiple state site-files simultaneously. Overall, 94% of respondents stated that they would access electronic data or use databases more frequently if it were easier to work with such data.

Values

Seventy percent of respondents believe that documentation of the archaeological record is being irretrievably lost quickly (32%) or that it is a crisis (38%). It is believed that the responsibility of preventing this loss is shared. However, institutions employing archaeologists have the greatest primary responsibility followed by research sponsors, SHPO/THPOs, and individual archaeologists. Museum's and repositories ranked slightly lower. Archaeological professional organizations was the only category to which little primary responsibility was attributed. Respondents indicated that there is a clear need for some organization to archive digital archaeological data and that there should be officially sanctioned data standards to which archaeologists should adhere. There was a high level of agreement that there should be a web-based tool for searching for relevant data.

Archiving Data

Ninety percent of respondents thought that research sponsors should carry the responsibility and financial burden of curating archaeological digital data. Most (63%) thought that a repository at the national or regional scale dedicated to digital archiving was the appropriate location within which to curate these data, although some (21%) indicated that digital data should reside with the artifacts. Cultural consulting firms were thought to be the least suitable location for digital archiving. A slight majority (55%) believed that scholarly journals and presses have a responsibility to make available the data upon which a published work is based. The greatest concern about contributing data to a digital archive were issues of confidentiality and data security followed by the cost and amount of time necessary to prepare digital data sets for curation.

Costs

Most respondents (61%) thought that there should not be a fee to access to digital data sets. A slightly greater percent (68%) thought that there should not be a fee to submit data. If there were a fee to pay for submitting data, 56% stated that their project sponsors would pay this fee.