ASU

ARIZONA STATE UNIVERSITY

School of Human Evolution & Social Change

Digital Antiquity

PO Box 872402, Tempe, AZ 85287-2402
(480) 965-6909   FAX: (480) 965-7671
http://tdar.org/confluence/display/DIGITAQ/Home
fpm@digitalantiquity.org

DIGITAL ANTIQUITY COMMENT TO THE OFFICE OF SCIENCE AND TECHNOLOGY POLICY EXECUTIVE OFFICE OF THE PRESIDENT

21 JANUARY 2010

RESPONDING TO THE REQUEST FOR COMMENTS ON PUBLIC ACCESS POLICIES FOR SCIENCE AND TECHNOLOGY FUNDING AGENCIES ACROSS THE FEDERAL GOVERNMENT-- IMPLEMENTATION, TECHNOLOGY, MANAGEMENT

We applaud the Office of Science and Technology Policy (OSTP) for establishing the Public Access Forum and seeking formal comments on how to improve access to the trove of scientific data produced by federally funded or required projects. Digital Antiquity, the organization that we represent participated in the forum portion with a posting on 20 December. Our concern expressed in the comment regarded the importance of including federally-generated archaeological data as one of the kinds of data for which improved access should be considered. We hope to ensure that the public access initiative extends beyond federal agencies that produce research as their primary product. The majority of archaeological research in the United States is performed pursuant to federal projects and undertakings that have other primary objectives, such as water management, natural resource extraction, improvements of the communications, energy and transportation infrastructures, or the conduct of military exercises. However, the archaeological data and research from these projects are essential for improving our understanding of American archaeology and the past human behaviors and cultures of the Americas that can be derived through the appropriate

analysis of these data. We emphasize this concern in our comment here, along with other considerations of how access to the research data can be improved. Our comments are organized according to the three general areas that OSTP used in soliciting comments: implementation, features and technology, and management.

Digital Antiquity (http://digitalantiquity.org) is a new organization dedicated to establishing an on-line digital repository of archaeological data and documents. Its primary goals are to expand dramatically access to the digital records of archaeological investigations and to ensure their long term preservation. Based at Arizona State University (where it is sponsored jointly by the School for Human Evolution and Social Change and the Arizona State University Libraries), Digital Antiquity is multi-institutional organization operating collaboratively with the University of Arkansas, Pennsylvania State University, the SRI Foundation, the University of York's Archaeology Data Service, and Washington State University.

Implementation: We wish to ensure that any federal policy and administrative actions developed out of this initiative to improve public access to scientific data include archaeological data that are produced by federal agencies for the management and protection of archaeological resources for which they are responsible or that are impacted by undertakings that involve federal agencies. We would welcome the opportunity to discuss this matter in more detail and at greater length with OSTP representatives.

Federal agencies annually produce, or require the production of, most of the archaeological research and associated data in the United States. The data from these individual research efforts can be substantial and have addressed important anthropological and historical issues, such as the development of agriculture; the actions ancient human societies took in the face of changing climate; and, interactions among different ethnic groups during ancient times and the historic period. However, the mass of archaeological data from this large overall research effort are not effectively shared, integrated, or utilized by other scientists and scholars.

United States government agencies reported producing or requiring the production of 86,000 archaeological overviews or record searches, 103,000 archaeological field studies, and 518 archaeological excavations during 2008 (http://www.nps.gov/archeology/SRC/index.htm, accessed 18 December 2009). In addition to the National Science Foundation and the National Endowment for the Humanities, nearly three dozen federal agencies conduct or require archaeological research. Agencies with the largest archaeological programs or that fund large amounts of archaeological research include: the Forest Service, the Bureau of Land Management, the National Park Service, the Corps of Engineers, the Bureau of Reclamation, the Federal Highway Administration, the Department of Energy, and the Department of Defense services (see *The Goals and Accomplishments of the Federal Archeology Program: The Secretary of the Interior's Report to Congress on the Federal Archeology Program, 1998-2003* for a description of the Federal Archeological Program; http://www.nps.gov/archeology/SRC/index.htm, accessed 18 December 2009).

Much of the archaeological research in the United States results from environmental or historic preservation reviews required by federal statutes, such as the National Historic Preservation Act, the Archaeological Resources Protection Act, or the National Environmental Policy Act. The research typically is organized in relatively small projects focused on specific areas where some kind of environmental impact is expected. Research involves checking these areas to see if archaeological resources exist there, and if they do, conducting historical and scientific research to determine the significance of the resource. If significant resources are identified and the project cannot be relocated to avoid further disturbance of them, additional research to recover the data that will be destroyed by the planned project is conducted.

Federal agencies already have the legal responsibility (e.g., under federal regulation 36 C.F.R. 79) to require curation of digital data in a form that will be accessible and survive in perpetuity. Yet, despite federal mandates requiring preservation and access to digital archaeological data and collections, the vast majority of data from federal research are difficult or impossible to access. Enforcement of the existing mandates would encourage widespread professional

participation. Of course, enforcement presumes repositories that are capable of meeting the existing data access and curation requirements.

Much of the archaeological research data produced by or for federal agencies over the past century exists in technical, sometimes lengthy, limited-distribution reports scattered in offices across the nation. Some of the data that underlie these reports are encoded in computer cards, magnetic tapes and floppy disks degrading in archives, book shelves, file cabinets, or desk drawers, while the technology to retrieve them and the human knowledge to make them meaningful rapidly disappears (Michener et al. 1997).

Rather than systematically archiving computerized information and making it available electronically so that it is useable, museums and other repositories typically treat the media on which the data are recorded as artifacts – storing them in boxes on shelves. Childs and Kagan (2008) report that only a few of the 180 archaeological repositories that responded to their recent survey charge a fee to upload digital data from the collections and records they curated to computers for preservation and access. This implies that the repositories recognized the seriousness of this activity and costs inherent in uploading and providing access, but that they are not able to provide digital access and preservation. Along with Childs and Kagan, we are concerned that the default preservation treatment for digital data used by almost all of the repositories that responded to their survey preserves the digital media, but leaves the data on the media actually inaccessible. Moreover, as computer software and hardware change and as the bits on the magnetic and optical media gradually, but inevitably "rot," the data will be completely unavailable for future research.

We believe that the agencies conducting or requiring archaeological research should ensure that the results of this research, publications, technical and popular reports, and data of various sorts, should be made more easily accessible. We understand, however, that simply requiring agencies to do something is not very helpful, if these agencies do not have readily accessible means of complying. To that end, we believe the creation of trusted repositories as well as

software designed to allow for the successful digital archiving of these materials is crucially important.

Features and Technology: Today, archaeologists in public agencies, private sector consulting firms, and academic settings spend a great deal of time searching for and acquiring relevant archaeological datasets and reports. Once found, more time is required to hunt for key data in volume after volume of hard copy reports that sometimes extend to more than a thousand pages.

The ability to reanalyze existing data can make present-day investigations more productive. Easy and complete access to existing data also reduces the likelihood of costly and unnecessary redundant projects. The ability to identify and integrate existing data that are comparable with new data sets being analyzed provides the opportunity for comparative investigations that have the potential for expanding and extending the scope of knowledge creation.

One example of how money could be saved if easier and wider access to existing archaeological data were available is found in a recent investigation in New Mexico. SRI, a private sector consulting firm, conducted archaeological investigations as part of a federal undertaking in the Loco Hills, a 460 square mile area in southeastern New Mexico. The firm carried out a field survey of 75,000 acres to identify and evaluate archaeological sites within the area and assess the impacts of proposed energy extraction activities to significant archaeological resources. In assessing the results of their field survey, it was learned that about 12,000 of these acres were areas that had been previously archaeologically surveyed. The reason for the re-survey was that the information on what had been surveyed previously was only available in files at the New Mexico State Historic Preservation Office. The state office is years behind in placing information about already investigated area on their statewide GIS. Oil and gas companies, such as the one that funded the Loco Hills investigation as part of their environmental review requirements, find it easier to resurvey plots than to send someone to Santa Fe and go through the paper records. If we estimate the average cost per acre for an archaeological survey at

$100, the re-survey of the already investigated portion of the Loco Hills project cost about $1.2M. If such unnecessarily redundant studies occur in 50 other situations, roughly $60M is wasted conducting archaeological field investigations that are not needed. By contrast, entire budget of NSF's archaeology program is only $7.5M annually. This example suggests that improving the availability and ease of access to archaeological data for environmental compliance activities alone would accrue savings that could fund the bulk of American academic archaeology for 8 years.

In recent years, the National Science Foundation has funded the development of a prototype digital repository for archaeological data, known as the Digital Archaeological Record (tDAR). The digital repository software is being refined and expanded as a part of the Digital Antiquity implementation. Digital Antiquity's repository will encompass digital documents and data derived from ongoing archaeological research, as well as legacy data and documents collected through more than a century of archaeological research in the Americas. The information resources preserved and made available by tDAR will be documented by detailed metadata submitted by the user before uploading the data and documents. Metadata may be associated generally with a project or specifically with an individual information resource (such as a database, document or spreadsheet). In addition to technical and other bookkeeping data, these metadata provide spatial, temporal, and other keyword information that will facilitate other users' discovery of relevant datasets and documents. They also include detailed information about authorship and other sorts of credit that must (as a requirement of the tDAR user agreement) accompany any use of information downloaded from the repository.

For databases and spreadsheets, the metadata include column-by-column descriptions documenting the observations being made including, "coding sheets" that will decode numerical values or string abbreviations associated with the appropriate labels of nominal categories.

tDAR now accommodates databases, spreadsheets, and documents in a limited number of formats. While the digital files are maintained as submitted, they are also—whenever

necessary—transformed into a format that can be sustained in the very long term (e.g. translation of Word files into a more sustainable PDF/A format). Planned development includes the expansion of the data and document formats accepted, as well as the inclusion of images, GIS, CAD, LiDAR and 3D scans, and other remote sensing data.

The inclusion of these more exotic forms of data awaits the completion of another component of the Digital Antiquity project, development of "best practices" guidelines for the creation and preparation of metadata descriptions and standards for different sorts of archaeological digital data. These guidelines build on the well-developed guideline series published by the Archaeology Data Services (ADS) in the United Kingdom http://ads.ahds.ac.uk/project/goodguides/g2gp.html. Julian Richards, Director of ADS, and Fred Limp of the University of Arkansas are leading the preparation of these guidelines.

Individual repository data sets and documents will be assigned persistent URIs that will provide permanent, citable web addresses. When content is revised, earlier content will be automatically versioned, so that the exact content as of a given date always can be retrieved. Sensitive information, such as site locations, can be restricted to qualified individuals. Investigators also can mark content (notably for ongoing projects) as "private" for a defined period, prior to a public release.

The development of tDAR, an easily accessible archive of digital archaeological data, offers the potential for more efficient and effective background research of past archaeological work, saving time and money for public archaeological management and preservation efforts, as well as for scholarly research. This online archive also will permit broad, comprehensive upgrading of digital data as new platforms for data storage and retrieval develop.

Management: To achieve this potential, we must transform archaeological practice so that the digital archiving of data and the description of metadata necessary to make it meaningful for general searching and access become a standard part of all archaeological project workflows.

Federal agencies can and should play an important role in facilitating this transformation. Agencies with land and resource managing responsibilities (such as, the Forest Service, Bureau of Land Management, Fish and Wildlife Service, National Park Service, Defense Department services, and Tennessee Valley Authority) and agencies with development or licensing responsibilities (for example, the Federal Highway Administration, the Environmental Protection Agency, the Federal Energy Regulatory Commission, and the Corps of Engineers) either fund or require tens of thousands of archaeological investigations annually (see the first section of this comment for references). By including among the requirements in scopes of work for these investigations the digital archiving of documents, data, images, and other products agencies can have a widespread, immediate, and lasting effect on American archaeological research.

Agencies like the Advisory Council on Historic Preservation, the Council on Environmental Quality, and state agencies responsible for archaeological and historic resources in each state (the State Historic Preservation offices established by the National Historic Preservation Act and partially funded by federal grants) also can influence archaeological practice by requiring that final reports of these public archaeology investigations demonstrate that the digital archiving of the results of the studies has been accomplished before approving any final report, which often is a project requirement.

As noted in the first section of this comment, Federal agencies already have the legal responsibility (e.g., under federal regulation 36 C.F.R. 79) to require curation of digital data in a form that will be accessible and survive in perpetuity. A new policy that promotes wider access to government data will underscore this responsibility. Emphasis will support efforts by archaeologists within the federal agencies to procure funding to support the digital archiving activity.

New policy development, led by OSTP, opening access to federal archaeological data presents an exciting opportunity for advancing knowledge through improved and wider-ranging comparative analysis of archaeological data and easier synthesis of these data. Already developing within the discipline of American archaeology, are mechanisms (such as Digital Antiquity and tDAR) for federal agencies and other public institutions to satisfy their legal mandates and professional responsibilities to provide access to the digital records of archaeological research and to effect long term curation using professional archival practices. These mechanisms will not only store data, but will provide the tools required by archaeologists to identify and access those data. It is anticipated that these mechanisms will enable private sector consulting archaeology firms, public agencies, and academic archaeologists to work much more effectively. It will enormously increase the accessibility – and impact – of the important work that the consulting firms and agencies do in managing, preserving, and protecting America's archaeological record.

Indeed, widespread digital access to archaeological data of the sort envisioned using tDAR has the potential to transform the practice of archaeology by enabling synthetic and comparative research on a scale heretofore impossible. The moment is right for this initiative. To succeed, however, cooperation and coordination throughout the discipline is needed. Those of us involved in Digital Antiquity look forward to working with OSTP and other organizations through mutually beneficial partnerships to achieve the potential that the is possible.

Francis P. McManamon, Ph.D.   Executive Director, Digital Antiquity

Sander van der Leeuw, Ph.D.   Chair, Board of Directors, Digital Antiquity
Director, School of Human Evolution and Social Change, Arizona State University

References

Childs, S. Terry and Seth Kagan 2008 A Decade of Study into Repository Fees for Archeological Collections. *Studies in Archeology and Ethnography* Number 6. Archeology Program, National Park Service, Washington, DC. http://www.nps.gov/archeology/PUBS/studies/STUDY06A.htm, accessed 28 November 2009

Departmental Consulting Archeologist 2009 *The Goals and Accomplishments of the Federal Archeology Program: The Secretary of the Interior's Report to Congress on the Federal Archeology Program, 1998-2003*. Archeology Program, National Park Service, Washington, DC. http://www.nps.gov/archeology/SRC/index.htm, accessed 18 December 2009).

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B. and Stafford, S.G. 1997. Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications* 7: 330-342.

Other Related Documents

Kintigh, Keith W. The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71(3):567-578. (2006). http://tdar.org/confluence/download/attachments/17399811/2006+AmAntiq-Kintigh-Challenge+and+Promise+Data+Integration.pdf?version=1&modificationDate=1264100062005

Snow, Dean R., Mark Gahegan, C. Lee Giles, Kenneth G. Hirth, George R. Milner, Prasenjit Mitra, and James Z. Wang. Cybertools and Archaeology. *Science* 311:958-959. (2006) http://tdar.org/confluence/download/attachments/17399811/2006+Science-DSnow+et+al+Cybertools+Archaeology.pdf?version=1&modificationDate=1264099734617