ARTICLE

DIGITAL ANTIQUITY

TRANSFORMING ARCHAEOLOGICAL DATA INTO KNOWLEDGE

Francis P. McManamon and Keith W. Kintigh

Frank McManamon is Research Professor and Keith Kintigh is Professor in the School of Human Evolution and Social Change at Arizona State University.

Digital Antiquity (http://digitalantiquity.org) is a new organization dedicated to establishing an online 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. Through a web interface users worldwide will be able to discover and download data and documents relevant to their research. Users also will upload their own data and documents along with the metadata (the data about the data) to the repository, known as tDAR (the Digital Archaeological Record), thereby making it possible for others to discover and effectively use the uploaded information. The access provided to documents and databases will permit scholars to create and communicate knowledge of the long-term human past more effectively and to enhance the management and preservation of archaeological resources.

The Need for Digital Archiving

Much of the information produced by archaeological research 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, museums, book shelves, file cabinets, or desk drawers, while the technology to retrieve them and the human knowledge to make them meaningful rapidly disappears (Eiteljorg 2004; Michener et al. 1997). Rather than systematically archiving computerized information so that it can remain 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) found that only a few of the 180 archaeological repositories that responded to their recent survey reported charging fees to upload digital data from the collections and records they curated to computers for preservation and access. By far, the most common preservation treatment for digital data used by the repositories that responded to the Childs and Kagan survey preserves the media on which the digital data files are stored, but leaves the data on the media actually inaccessible. This physical curation is an inadequate long-term preservation approach as computer software and hardware change and as the bits on the magnetic and optical media gradually, but inevitably, "rot."

Much of the archaeological work in the United States involves federal funds, lands, or permits and is subject to federal law. Federal agencies already have the legal responsibility (36 C.F.R. 79; Sullivan and Childs 2003:23–38) to require curation of archaeological collections and associated records, including digital data, in a form that is accessible and will survive in perpetuity. Yet, despite federal mandates requiring preservation and access to digital data, the vast majority is difficult or impossible to access and will not be preserved in the formats in which they currently reside. The existing mandates already are in place to justify widespread professional participation. However, compliance with the mandates requires the existence of repositories capable of meeting the data access and curation needs.

The intertwined problems of data access, preservation, and synthesis are not new to archaeology. In the late 1990s, a series of meetings and panels were sponsored by the Society for American Archaeology, the Society of Professional Archaeologists (now the Register of Professional Archaeologists), and the National Park Service on the general topic of "Renewing Our National Archaeological Program." Improving the management of archaeological information through greater data access and synthesis was one of the major topics covered in this effort (Lipe 1997; McManamon 2000). The challenges of data access and preservation are not unique to archaeology. The September 10, 2009 issue of *Nature* began with an editorial calling for broader sharing of data and its long-term preservation and related reports on data access and preservation challenges (*Nature* 2009a, 2009b; Nelson 2009; Schofield et al. 2009). The editorial cited particular successes:

> Pioneering archives such as GenBank have demonstrated just how powerful such legacy data sets can be for generating new discoveries—especially when data are com-

> bined from many laboratories and analysed in ways that the original researchers could not have anticipated [Nature 2009a:145].

However, the editorial emphasized that most scientific disciplines

> still lack the technical, institutional, and cultural frameworks required to support such open data access—leading to a scandalous shortfall in the sharing of data by researchers. This deficiency urgently needs to be addressed by funders, universities, and researchers themselves...[Furthermore] funding agencies need to recognize that preservation of and access to digital data are central to their mission, and need to be supported accordingly [Nature 2009a:145].

Also in 2009 the National Academies released a book-length report on efforts to ensure the integrity, accessibility, and stewardship of digital research data (National Academies 2009). At the same time we look back on legacy data, we must also look forward. A substantial amount of public archaeological work is carried out annually. Federal agencies report approximately 50,000 field projects involving archaeological resources conducted in the United States, mostly by cultural resource management firms or agency staff (Departmental Consulting Archeologist 2009). Given the volume of data and reports produced each year, even archaeologists working in the same area often are unaware of important results that others have already reported. Archaeological studies are generating loads of data, but the data cannot be used efficiently and effectively to advance knowledge of the past. The difficulty of sharing information about and from existing research is exacerbated by the demographic transition underway in the ranks of professional archaeologists. Large numbers of archaeologists entered the profession in the 1960s and 1970s. These individuals are retiring or passing away. Now is the time to capture for long-term preservation and access the digital data associated with the work carried out by this cohort of archeologists. Accessing the information by relying on the memories of individuals, no matter how prodigious these memories might be, will be impossible once these individuals are no longer available.

Today, a great deal of time is spent searching for and acquiring relevant 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. Yet, the ability to reanalyze existing data can make present-day investigations more productive and has the potential to recognize and reduce costly redundant projects.

The Digital Archaeological Record (tDAR)

In 2004, the National Science Foundation funded a workshop focused on the integration and preservation of structured digital data derived from archaeological investigations. The workshop included 31 distinguished participants from archaeology, physical anthropology, and computer science. The workshop report concluded

> for archaeology to achieve its potential to advance long-term, scientific understandings of human history, there is a pressing need for an archaeological information infrastructure that will allow us to archive, access, integrate, and mine disparate data sets [Kintigh 2006:567].

A subsequent $750,000 NSF grant funded the development of a prototype of tDAR, the digital repository software that will be refined and expanded as a part of the Digital Antiquity implementation. Development and testing of the tDAR prototype was led by Kintigh and involved a team that included Arizona State University archaeologists (Ben Nelson, Margaret Nelson, and Katherine Spielmann) and computer scientists (K. Selçuk Candan and Hasan Davulcu), as well as the Associate University Librarian (John Howard),

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 are 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 (e.g., 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. Finally, for databases and spreadsheets, they include column-by-column metadata that document 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 inclu-

sion 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 Mellon-funded project, development of "best practices" guidelines for the creation and preparation of metadata descriptions 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 soon all have persistent URLs that will provide permanent, citable web addresses. When content is revised, earlier content is 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.

To achieve this potential, we must transform archaeological practice so that the digital archiving of data and the metadata necessary to make it meaningful become a standard part of all archaeological project workflows. To help jumpstart this transition Digital Antiquity has allocated $225,000 to a grants program to encourage the deposit in tDAR of important archaeological documents and data that already exist in digital form. More information about the criteria for grants and their availability will be widely distributed as the program develops.

Digital Antiquity

Digital Antiquity, the organization that manages tDAR repository, is the direct product of a multi-institutional effort to plan a sustainable digital repository for archaeological documents and data that was funded by the Andrew W. Mellon Foundation The Mellon Foundation has now funded the implementation of Digital Antiquity and tDAR in response to the $1,290,000 proposal that grew out of the multi-institutional planning grant. The proposal was authored by Keith W. Kintigh (Arizona State University), Jeffrey Altschul (SRI Foundation), John Howard (University College, Dublin), Timothy Kohler (Washington State University), Frederick Limp (University of Arkansas), Julian Richards (University of York), and Dean Snow (The Pennsylvania State University).

Digital Antiquity confronts several challenges to succeed as a sustainable digital repository. Its business plan envisions either a transition from an entity incubated by the University into an independent not-for-profit or to a unit of an established nonprofit with compatible goals that can manage Digital Antiquity's services and data assets in the long term. Digital Antiquity's business plan is based on a model in which those who are responsible for archaeological investigations will pay a fee for the deposit of data and documents in the tDAR repository. In return, long-term preservation of the data will be assured and access to the data and documents will be freely available over the Internet, with controlled access to sensitive data.

The Mellon Foundation implementation grant has funded the establishment of Digital Antiquity as an independent organization that, for a four-to-five year startup period, is hosted by Arizona State University. In November 2009, Francis P. McManamon, formerly Chief Archeologist of the National Park Service and Departmental Consulting Archeologist for the Department of the Interior, began working as the full-time Executive Director. The staff will include two full time software engineers, a data curator, user support specialist, and clerical staff.

Digital Antiquity is governed by a 12-member Board of Directors who oversee the performance of the Executive Director and provide entrepreneurial and disciplinary guidance. The Board of Directors is chaired by archaeologist Sander van der Leeuw, Director of ASU's School of Human Evolution & Social Change (formerly, Department of Anthropology), and has as members the individuals from six institutions whose efforts succeeded in obtaining the Mellon grant, plus four directors from the private sector with expertise in business, law, finance, management, and commercial information technology. A 12-member Science Board, composed of archaeologists representing different sectors of the discipline, computer scientists, and informatics experts, has been established to advise Digital Antiquity on technical and disciplinary matters. The memberships of both boards are available on the Digital Antiquity home page: http://digitalantiquity.org.

Conclusion

Digital Antiquity represents an exciting opportunity for advancing knowledge through improved and wider-ranging comparative analysis of archaeological data and easier synthesis of these data. Through tDAR, Digital Antiquity provides a mechanism for public agencies and other 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. Digital Antiquity will not only store data, but will provide the tools required by archaeologists to identify and access those data. It is anticipated that once tDAR is fully established and data begin to populate it, consulting archaeology firms and public agencies, as well as academic archaeologists, will be able 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 provided by 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 through mutually beneficial partnerships with diverse organizations and individuals to achieve the potential that the initiative offers.

Acknowledgments. We appreciate and have used the comments and suggestions of our colleagues: Jeff Altschul, Terry Childs, Tim Kohler, Fred Limp, Peggy Nelson, Julian Richards, and Dean Snow, on an earlier draft of this article. The Digital Antiquity initiative and tDAR, the Digital Archaeological Record, have been funded by grants from the Andrew W. Mellon Foundation and by the National Science Foundation (0433959, 0624341). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Mellon Foundation.

References Cited

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. Electronic document, http://www.nps.gov/archeology/PUBS/studies/STUDY06A.htm, accessed November 28, 2009.

Departmental Consulting Archeologist
2009 *The Secretary of the Interior's Report to Congress on the Federal Archeological Program, 1998–2003*. Archeology Program, National Park Service, Washington, DC. Electronic document, http://www.nps.gov/archeology/SRC/src.htm, accessed November 28, 2009.

Eiteljorg, Harrison, II
2004 Archiving Digital Archaeological Records, In *Our Collective Responsibility: The Ethics and Practice of Archaeological Collections Stewardship*, edited by S. Terry Childs, pp. 67–73. Society for American Archaeology, Washington, D.C.

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

Lipe, William D.
1997 Report on the Second Conference on Renewing Our National Archaeological Program, February 9–11, 1997. Electronic document, http://www.saa.org/AbouttheSociety/GovernmentAffairs/NationalArchaeologicalProgram/tabid/240/Default.aspx, accessed 1 December 2009.

McManamon, Francis P.
2000 *Renewing the National Archaeological Program: Final Report of Accomplishments. A Report to the Board of the Society for American Archaeology from the Task Force Chair*. Society for American Archaeology, Washington, D.C. Electronic document, http://www.saa.org/AbouttheSociety/GovernmentAffairs/NationalArchaeologicalProgram/tabid/240/Default.aspx, accessed December 1, 2009.

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

National Academies
2009 *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. The National Academies Press, Washington, D.C.

Nature
2009a Editorial: Data's Shameful Neglect. *Nature* 461(7261):145.
2009b Opinion: Prepublication data sharing. *Nature* 461(7261):168–170.

Nelson, Bryn
2009 Data Sharing: Empty Archives. *Nature* 461(7261):160–163.

Schofield, Paul N., Tania Bubela, Thomas Weaver, Stephen D. Brown, John M. Hancock, David Einhorn, Glauco Tocchini-Valentini, Martin Hrabe de Angelis, and Nadia Rosenthal
2009 Opinion: Post-publication Sharing of Data and Tools. *Nature* 461(7261):171–173.

Sullivan, Lynne P., and S. Terry Childs
2003 Curating *Archaeological Collections: From the Field to the Repository*. AltaMira Press, Walnut Creek, California.