Sustaining the Digital Archaeological Record

Keith W. Kintigh and Jeffrey H. Altschul

Keith W. Kintigh, School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona 85282-2402 (kintigh@asu.edu)

Jeffrey H. Altschul, Statistical Research, Inc. and the SRI Foundation, P.O. Box 31865, Tucson, Arizona 85751 (jhaltschul@sricrm.com)

Economically sustainable digital preservation is a necessary condition for securing the long-term future of our scholarly and cultural record.

Brian Lavoie, research scientist at OCLC quoted in Zverina (2008)

In the United States, widespread recognition of the “curation crisis” is generally focused on the absence of repositories willing to accept new collections of archaeological materials. Most attention is drawn to the “stuff” of archaeology — decaying cardboard boxes of artifacts, faded photographs, yellowed maps and notes, and the like — with the face of this crisis generally being an old building, in desperate need of more space, repair, renovation, and staff. Even though the political will to fix this problem is more often than not lacking, the public and our elected officials have no trouble grasping the need. Artifacts that archaeologists recover belong in museums. Land developers who destroy archaeological sites have the responsibility to pay to excavate these sites, analyze the material, and curate the collection in perpetuity.

There is a second curation crisis that receives much less attention but whose importance is growing daily: a digital curation crisis (Eiteljorg 2004; Sullivan and Childs 2003: 37-38). In the information age, more and more archaeological work is being recorded, analyzed, and reported digitally. Increasingly, our records lack the traditional paper forms; photographs, databases, maps, survey and excavation forms, reports, etc. are “born” digital, with no paper copy. However, much of the digital record of a project is often not curated. Even when a repository accepts some or all of the digital record, it is mostly as files on a CD or DVD (Childs and Kagan 2008). But as we all know through painful experience, these media are short lived. Unless the files are uploaded to a computer, backed up and checked regularly to ensure that the files are not corrupted, and migrated as software and hardware change, their lifespans fall very far short of perpetuity (see Eiteljorg 2004; Michener et al. 1997). Those who have toiled so hard to create these archaeological archives of primary data are the losers as are current and future archaeologists who could have used these data to further investigate the past or manage the archaeological record.

The library sciences have developed a term for an institutional structure that provides for the proper preservation, discovery, and access to digital material: it is called a “trusted” repository (Research Libraries Group 2002). No trusted repository exists for American archaeology — and we need one desperately.
Digital Antiquity

Although many have decried the situation (Eiteljorg 2004; Kintigh 2006), the development of a digital archive for American archaeology has remained elusive. The startup costs are considerable and the long term viability of such a trusted repository is not guaranteed. Success of such an endeavor requires that American archaeologists agree on the need, commit to a solution, and find our own political will to convince the sponsors of archaeological research in the United States to pay for it. This is a long journey, the first stop of which must be to create a trusted repository.

Digital Antiquity (http://www.digitalantiquity.org/) is a new, multi-institutional organization established to address the crisis in the curation of digital documentation of archaeological investigations (McManamon and Kintigh 2010). Its primary goals are to provide access to the digital records of archaeological investigations and to ensure their long term preservation. Digital Antiquity is achieving its goals through tDAR, the Digital Archaeological Record, a new on-line digital repository of archaeological data and documents (http://tdar.org).

The Andrew W. Mellon Foundation has generously underwritten the startup costs of Digital Antiquity. In the long term, however, Digital Antiquity must be self sustaining; we must have a plan to remain current with the technology as well as with the evolving social and professional needs of the discipline. We also must become financially sustainable and we must achieve financial independence within four to five years.

In developing Digital Antiquity’s business plan, we have been substantially assisted by Julian Richards and informed by the experience of the United Kingdom’s (UK) Archaeology Data Service (ADS) that he directs. Although we explored a number of alternative funding models (Guthrie et al. 2008), we have pursued a funding model similar to that employed by ADS (Richards et al. this volume).

The Digital Model—Costs

We conceptualized the anticipated costs of running Digital Antiquity as either fixed or variable and as either present or future costs. Fixed costs, such as the salary of the Executive Director, are those that are essentially independent of the volume of data that come into the digital repository or the amount of use of the archive. The variable costs are those that are expected to increase with the volume of data, such as computer storage and staff salaries for providing user help. Present costs are those incurred at approximately the same time as the data are acquired, such as salaries for processing and review of the incoming data. Future costs are the long-term anticipated costs for such things as migration of data as digital formats and media change and the future costs of computer storage. A sustainable financial model, then, must provide for both the present and future costs.

As we modeled Digital Antiquity’s cost structure, it became clear that through the startup period our budget would be strongly dominated by fixed costs. The largest fixed costs are salary, benefits, and travel associated with the organization’s administration and software development and maintenance. Additional fixed costs cover marketing and the operation of Digital Antiquity’s Board.
of Directors and Science Board. Initially, Digital Antiquity’s annual fixed costs are estimated at approximately $600,000. It is important to note, however, that these costs are subsidized substantially by Arizona State University, which provides free of charge, office space, insurance, legal support, and auditing services. Digital Antiquity will have to sustain substantially higher fixed costs as the organization moves into independent non-profit status from its initial incubation stage.

Variable costs are also dominated by salaries and benefits, in this case associated with user support staff and the staff involved in the human curation components of ingesting—that is, properly incorporating—documents and data into the digital repository, including the processing and review of input materials. It appears that the costs of servers and storage are comparatively small.

Future fixed costs are modeled to rise primarily due to inflation. However, the future variable costs are an issue of concern, particularly with respect to the costs of periodically refreshing and migrating as well as storing and delivering the data. The experience of ADS (Richards et al. this volume) provides some assurance that these future costs can reasonably be projected to decline over time. Adding it all up, Digital Antiquity needs to have annual revenues on the order of $1 million.

The Financial Model—Revenue

How do we make a million dollars a year? As we designed Digital Antiquity, we considered a number of revenue models. The first, we label “contributor pays.” In this model, the institution, agency, or individual that is responsible for the digital data pays a fee at the time the data and documents are ingested by the repository. A contributor pays model is attractive primarily because it parallels the current method of paying for curation by both grant-supported research and compliance-sponsored cultural resource management (CRM). Both modes of financing archaeological work budget curation costs up front, with curation paid for by the sponsoring agency. Although focused primarily on physical remains, such as artifacts, documents, photographs, curation facilities also accept digital materials (see Lyons and Vokes this volume), which are typically included in their overall fees. Sponsors, therefore, already pay for the curation of media encoding digital data and it would be a small, but critical step to also cover the costs of curating digital data—which is to say the actual information—in a trusted digital repository. Assuredly, sponsors will be concerned with the overall cost of curation, but few are likely to object to the payment for digital data curation in principle.

The downside to the contributor pays model is that, in the US, curation in general is woefully underfunded. In the UK, ADS projects that proper digital curation alone costs between one and three percent of total project costs. In the US, total curation costs for both physical and digital materials on many projects do not reach three percent (but see Sonderman 2004; Majewski this volume). It is unrealistic to assume that sponsors in the US will pay significantly more for curation than they currently do. It is also unrealistic to assume that museums and other repositories which are struggling financially will readily agree to share their current fees with Digital Antiquity.
The key then is to have a price structure for Digital Antiquity which is low enough that sponsors do not balk at paying the going curation rate for museum curation and also paying a modest fee to Digital Antiquity to curate the digital data. How low would such a fee have to be? We do not yet have an answer to this question, but we built the Digital Antiquity business model considering fees in the low hundreds for small projects and as much as the low thousands of dollars for larger ones. To hit our financial target, we are relying on volume and on the preparation and documentation of the digital data sets by the submitting archaeologist or institution, assumptions to which we will later return.

Before moving on, we will briefly examine the other forms of revenue models. One popular method is to charge for access to data. While there have been both successful (iTunes) and unsuccessful efforts to charge for content delivered over the web (numerous newspapers), we were dissuaded from this approach by more fundamental considerations. Most importantly, such a charge would be a strong disincentive to the very use we want to encourage to improve our understandings of the past and our management of archaeological resources. Furthermore, a large fraction of these data are collected with public funding whose legal justification is premised on the preservation of the cultural resource information for the benefit of the public. Nonetheless, we might ask, what sort of charge would it take to cover the anticipated costs. Although we consider public access to be a significant asset, we expect that the primary market consists of professional archaeologists and the organizations that employ them. If we use the membership of Society for American Archaeology (SAA) as a proxy for this market, to cover the initial annual fixed costs of $600,000 would require on the order of $85 per SAA member per year. To state it mildly, we anticipate that a substantial fraction of that market would not be willing to pay this cost. To make things very much worse, this figure would not cover any of the quite significant variable costs. The latter are primarily dependent not on the amount of access (that is, what is being charged in this model) but on the volume of what is being put into the archive. Thus, even if one wanted to pursue this revenue model, it would be difficult to budget effectively. The result would likely be that the amount of data being entered in the repository would have to be limited, greatly diminishing the utility of the archives and making it much harder to convince archaeologists that the annual fee will be rewarded with access to data critical to their work.

The information infrastructure will be of substantial value for science and for resource management. As such, we anticipate that granting institutions and federal land managing agencies might support it with grants and fee agreements. Certainly, planning and development of our own and other initiatives have been supported with grant funding (in our case from the National Science Foundation, the Mellon Foundation, and by a joint grant from the National Endowment for the Humanities and the UK’s Joint Information Systems Committee). It is our experience and observation that funding agencies may fund development but are — to say the least — loathe to commit to permanent or even long-term funding.
As such, our business plan does not rely on grants to finance the long term operations of Digital Antiquity.

There are lots of marvelous resources on the web that are sustained by advertising (Google) or some combination of contributed effort and donations (Wikipedia). Can we use either of these models? While we did not do a detailed analysis, our market is likely too small for advertising to sustain any significant fraction of the effort. We hope that contributed effort will be an important factor for Digital Antiquity — particularly with respect to the daunting problem of incorporating legacy data.

We will establish an endowment that will allow Digital Antiquity to ride out the vicissitudes of the business cycles in cultural resource management (CRM). These revenue generating initiatives, however, are in addition to and not part of our core revenue generating methods.

In short, our current business plan relies primarily on a contributor pays model to sustain data acquisition, access, and ongoing operations. We anticipate that it will continue to be possible to get grant funding for new development, but not for ongoing operations. Finally, it may be possible to supplement revenue through other mechanisms. For example, if tDAR were able to index paid digital content provided by commercially owned scholarly journals, it might collect a fee associated with paid downloads initiated through its searches.

**Trade-Offs**

In the economic climate of American CRM and grant-sponsored archaeology, a digital archive can only succeed by keeping its cost low and its volume high. In this respect, we depart significantly from the UK. The model that is successfully employed by ADS requires relatively little of the user but charges a relatively large fee to cover its staff’s curation efforts. This model is successful both from the standpoint of the quality of the curation and from the standpoint of financial sustainability. However, a by-product of this high staff investment, high charge model is that the volume of projects processed is relatively low — on the order of a few hundred annually.

In contrast, Digital Antiquity’s design anticipates eventually being able to process a substantial fraction of the projects conducted each year in the US. Annually, federal agencies report approximately 50,000 field projects involving archaeological resources conducted in the United States, mostly by CRM firms or agency staff (Departmental Consulting Archaeologist 2009). Even though this is a far from complete count, it signals a difference in scale of a couple of orders of magnitude.

Because of the issue of scale, Digital Antiquity has taken a different approach to this problem from ADS. Digital Antiquity requires the typical user to execute much more of the curation process. The largest component of this process is the provision of structured metadata, which provide the detailed technical and semantic documentation that are required for the data to be meaningfully used and technically sustained into the future. This approach has the potential advantage of directly engaging those who created the data and know it best in providing the substantive documentation that will enable others to effectively find and use it. It has the liability of placing curatorial responsibilities for which they are not trained in the hands.
of archaeologists. To mitigate that liability, Digital Antiquity has had to invest heavily in the development of Web-oriented software tools that can effectively guide the users to provide appropriate metadata for their digital documents, data sets, and images before they are ingested into the digital repository.

This design has the effect of off-loading some of the data curation labor requirements to the data contributors and to shift what would otherwise be variable costs of processing by human curators to fixed costs in higher levels of software development. The two key results are that we keep the out-of-pocket digital curatorial costs low and the solution is scalable to very high project volumes because the incremental costs of providing the computer services are quite low. A side effect is that Digital Antiquity’s design essentially rewards good behavior with respect to systematic digital data collection. Both out of pocket costs and contributor effort can be kept low by adherence to good practices.

With on-line submission of data to Digital Antiquity’s repository, many problems, errors, and inadequacies with the data or metadata will be immediately detected and can readily be remedied by the submitters. Additional human audits of submitted data will, in some cases, identify additional problems. However, even with our very large investment in a user-friendly ingest interface, some submitters may be unwilling to invest the time to adequately document their data or may lack the expertise to provide the data in acceptable formats. In these cases, Digital Antiquity can assist submitters in ingesting their digital collections on a fee-for-service basis.

**Reality Check**

In the end, proper curation of digital material in a trusted repository, whether Digital Antiquity or some other, will increase the cost of curation. Why would federal agencies pay more—or require project sponsors to pay more—for digital data curation when they appear to be meeting industry and discipline standards with their current curation arrangements for the physical objects? Our answer is that agency leadership probably will resist any additional costs unless they are convinced that it is in their self-interest to make this investment.

Influences that will change prevailing attitudes about digital curation will come from two directions. First, agencies will need to be convinced that preservation and access to digital data is central to completion of their management responsibilities. Second, agencies will need to be convinced that their current curation policies fail to meet their regulatory mandates. Who will lead the charge to convince agencies of these facts? We believe it is the archaeological community; archaeologists are going to have to insist on the changes needed to ensure proper preservation of and access to digital data.

It is in the self-interest of public agencies with archaeological resource responsibilities to support services that can provide them with efficient access to the digital documents, data sets, and images. Most federal agencies cannot adequately access documents and data about the archaeological resources for which they are responsible. The size of this problem grows daily with increases in the volume of this inaccessible information. Technical developments over the last decade have enormously improved our ability to store,
search, and share digital data. To take advantage of these developments, the compliance practice and standard regarding curation of digital documents, data sets, and images must change. Such a change requires that American archaeologists agree that digital data curation is essential to the practice of archaeology and that this position be supported by the agencies and institutions that guide and regulate American CRM and sponsored research.

The flip side of this coin is that the failure to preserve digital data puts agencies in jeopardy of failing to meet regulatory mandates to which they are subject. To make this case, archaeologists need to demonstrate that they actually use these data and that their loss is consequential to our ability to learn about the past and to safeguard the public’s interest in the past.

Archaeologists who work within and outside public agencies have a lot at stake in ensuring that digital data are properly curated. As consumers, archaeologists genuinely desire effective, on-line access to the results of others’ work, allowing them to enhance their own work products (Kintigh 2006) at the same time saving time and effort. As producers of archaeological information, we want to see our efforts put to use by our colleagues. CRM archaeologists, in particular, have felt that their work is ignored, when in fact these data have languished in large part because they are inaccessible. Academic archaeologists have expressed the fear that a digital archive will lead to the premature release of their data. Yet, adequate technological protections have been built into tDAR to ensure against this eventuality. Even without these protections, the likelihood of being “scooped” by someone else using your data is probably much less than the probability of losing your data entirely, because they are inadequately maintained outside a trusted repository.

Beyond self interest, we share an ethical responsibility to make available the results of our work. These responsibilities have been repeatedly articulated by our professional societies, including SAA, the Society of Historical Archaeology, and the Register of Professional Archaeologists. This responsibility is articulated clearly in the Society for American Archaeology Principles of Archaeological Ethics, notably in Principles 1, 5, and 7 (SAA Ethics in Archaeology Committee 2000). For example, the principles state that,

...knowledge and documents created through the study of archaeological resources... should be treated in accord with the principles of stewardship rather than as a matter of personal possession.

and “Archaeologists should work actively for the preservation of, and long term access to, archaeological collections, records, and reports.” In relation to digital material, SAA explicitly states:

Field notes, photographs, maps, laboratory notes and data, and other records require the same levels of management, care, and preservation as artifacts and other recovered items. Records generated during collections research and treatment should be deposited with the collection. Data and records created or stored in electronic formats are fragile and require specialized long-term care and management [Childs 2004: vii].
The ethical notion that archaeologists should preserve and make accessible to other researchers the material they recover was codified in the federal regulation that provides legal standards for curatorial facilities holding federal collections (36 CFR 79). This regulation is both far reaching and comprehensive. At the time 36 CFR 79 was adopted, most archaeological data were not “born digital” and digital records were much less of a concern than they are now. Yet, the regulation explicitly covers digital records. It requires that agencies curate digital records in a repository “appropriate to the nature and content of the collections” and “using professional archival practices,” and it specifically provides for maintaining multiple copies of digital records. Repositories must “provide long-term curatorial services” including to “…catalog, store, maintain, inventory and conserve the particular collection on a long-term basis using professional museum and archival practices 79.9(a)” and to “Provide access to the collection….”

Agencies and their repositories rarely meet the legal standards set by 36 CFR 79 with regard to the curation of digital data. Agencies thus not only have a reason to be concerned about the preservation and access of digital records, they have an affirmative responsibility to meet the standards dictated by the regulation. But will they? Only if pressure is applied. And who will apply that pressure if not archaeologists, within as well as outside the agencies.

There is reason for optimism. Other disciplines, such as astronomy, ecology, and the environmental sciences, are grappling with this same issue of access to and long-term preservation of digital data related to their subject areas (National Academies 2009; Nature 2009; Nelson 2009; Schofield et al. 2009; Toronto International Data Release Workshop Authors 2009). The President’s Office of Science and Technology Policy is considering how to make federal agency research data more generally accessible. Rising expectations with respect to the treatment of digital data are evidenced by the National Science Foundation’s increasing emphasis on requiring grantees to rapidly make available the digital products of funded research. Many federal land managing agencies also are trying to bridge their own digital divide by setting standards in contracts, investing heavily in databases, and utilizing digital data in day-to-day resource management. The economics of digital curation could be easily managed if Digital Antiquity could preserve and provide access to these data, not simply for an agency but for American archaeology as a whole (or at least a significant proportion of it).

Cultural resource management in the US is on the order of a billion dollar industry annually (Altschul and Patterson 2010). We believe that the dedication of a tiny fraction of that amount to the preservation and access to digital records can fund the curation to professional archival standards that is required by regulation and demanded by our ethics. A digital trusted repository would secure the long-term preservation of irreplaceable archaeological records that would otherwise be lost, likely sooner than later. Finally, it could enormously enhance our ability to learn about the past, by permitting us to more effectively build on the research of our colleagues and to facilitate the comparative and synthetic research.
that would now be so burdensome as to be almost unthinkable.

To the extent that requirements for effective preservation and access to digital records were to become incorporated in archaeological contracts and grants, the administrative mandates and funding would exist to move the results of new investigations into a trusted repository. However, these requirements would not solve all our digital curation problems. We still have to grapple with the mountains of legacy data, digital and otherwise, that derive from investigations accomplished over the last century. But we need to start somewhere and we need to start now. If we can demonstrate the utility of digital curation for current projects, then we will be in a much better position to develop a strategy for handling old projects.

To accomplish the movement of a substantial fraction of the data from prospective projects into trusted repositories is a tall order and the need is pressing. It will require not just an effective organization and strong software tools; it will also involve a commitment from the archeological community and government agencies, a change in archaeological work flows, and eventually, a revision of our expectations regarding professional behavior. We need to reach the stage that archiving professional records with high quality metadata in a trusted repository is not the exception but the norm.

Acknowledgments
This paper is based on a paper presented in the invited symposium, Dollars and Sense in Recovering and Managing Archaeological Collections, organized by Terry Childs at the 74th Annual Meeting of the Society for American Archaeology in Atlanta, Georgia. We gratefully acknowledge the efforts of the Digital Antiquity Board of Directors in shaping the effort reported here: Sander van der Leeuw, Carol Ackerson, Kim Bullerdick, Jaime Casap, John Howard, Tim Kohler, Fred Limp, Harry Papp, Julian Richards, and Dean Snow. Digital Antiquity’s Executive Director, Francis McManamon, has also provided valuable assistance. This article was greatly improved by comments from Terry Childs and Frank McManamon. Finally, we acknowledge not only the financial support of the Andrew W. Mellon Foundation, but also the guidance and encouragement of Donald Waters, the director of the Foundation’s Scholarly Communications program. The Digital Antiquity initiative and tDAR, the Digital Archaeological Record, are funded by grants from the Andrew W. Mellon Foundation and the National Science Foundation (NSF; 0433959), and by a grant jointly awarded by the National Endowment for the Humanities (NEH; PX-50022-09) and the Higher Education Funding Council for England of the United Kingdom acting through the Joint Information Systems Committee (JISC). Planning grants were funded by NSF (0624341) and the Mellon Foundation. 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 Mellon Foundation, NSF, NEH, or JISC.
References Cited

Altschul, Jeffrey H., and Thomas C. Patterson

Childs, S. Terry (editor)

Childs, S. Terry, and Seth Kagan

Departmental Consulting Archaeologist

Eiteljorg, Harrison, II

Guthrie, Kevin, Rebecca Griffiths, and Nancy Maron

Kintigh, Keith

McManamon, Francis P., and Keith W. Kintigh

Michener, William K., James W. Brunt, John J. Helly, Thomas B. Kirchner, and Susan G. Stafford

National Academies

Nature

Nelson, Bryn

Research Libraries Group


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