

PRESERVATION OF ONLINE PUBLICATIONS

JOSEPH L. CHARTKOFF
DEPARTMENT OF ANTHROPOLOGY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MI 48824
JOECHARTKOFF@COMCAST.NET

The growth of online publications in California archaeology, such as with the Proceedings of the Society for California Archaeology, has many virtues in making new research more widely accessible. It also raises some serious concerns about the long-term preservation and accessibility of reports. These issues need to be explored, and responses need to be adopted, so that the priceless contributions of research on irreplaceable data are not lost forever.

In the September 2008 issue of the *Society for California Archaeology Newsletter*, SCA President Mark Allen reports that the annual collection of papers from the Society's conference, the *Proceedings of the Society for California Archaeology*, is now being published online rather than being printed in hard copies as was done for its first 20 volumes (Allen 2008:4). The posting of the conference anthology online provides some important benefits to both the Society and the archaeological community. At the same time, however, it raises some significant questions about the preservation of these scholarly contributions in comparison with the traditional hard copies. This paper is intended to look at both the benefits and the potential problems, and to suggest some possible strategies by which the benefits can be retained but the potential problems can be essentially overcome.

BENEFITS

The SCA is hardly unique in moving to the posting of important documents online. Nearly all of the important journals in archaeology, and all other scholarly fields, are now available online. Not only is the practice increasingly widespread throughout the archaeological community nationally and globally, but it is a growing trend in all areas of scholarship, as well as many other domains of life. A number of benefits have been gained from doing so, and President Allen has noted some in his article. For example, the posting of collections of papers online is far less expensive to do than to have them printed in order to distribute them to the Society's membership, so the financial resources of the publishing agency (in our case, the Society for California Archaeology) can be put to other, even more valued, purposes, such as the launching of the Society's first refereed scholarly journal, which President Allen (2008) also discusses.

In addition, the availability of these papers online makes them accessible to a vastly wider audience than does the publication of them in a printed anthology series whose distribution is largely limited to the Society's membership. Online publications can be accessible to the entire archaeological community nationally, and even globally, and for that matter, to far wider audiences than just the archaeological community—literally to millions of potential readers. The potential productivity in creative thought for such an expansion in readership is extremely important and valuable. The economic savings to the readership also can be extremely significant.

On a separate but relevant note, the increasing level of communication between California archaeology and the rest of the academic world has been expanding in multiple dimensions in recent years, and the posting of California papers online is part of this expansion. As President Allen (2009) has noted, California archaeologists are contributing increasingly to national and international journals, such as *American Antiquity*. As a result, the isolation of California research from the rest of the archaeological

community, as I discussed in an SCA paper a couple of decades ago, “The Bleak and the Grey” (Chartkoff 1987), is increasingly becoming a phenomenon of the past.

PROBLEMS AND RISKS

Nevertheless, online publication poses some serious problems of preservation and accessibility which can, in turn, affect the survival of archaeological knowledge. This risk poses issues of both operational effectiveness and ethical integrity that need to be examined more than they have been. Electronic publication allows for enormously widespread distribution and impact on the scholarly community, but it also introduces a fragility for survival that is at a significantly higher level of danger than when information is published on paper.

Internet-based sources, to cite one problem, can be extremely vulnerable to future inaccessibility. As software evolves, we cannot be assured that material online that is currently accessible will remain accessible. We already have older programs that are no longer accessible with current software. In addition, access depends upon the survival of the networks. Loss of key servers, much less entire networks, can result in the loss forever of entire bodies of information (Eli D. Chartkoff, personal communication 2009).

A VECTOR ANALOGY

To illustrate, let me use an example from my own early involvement with desktop computers. This example is not at the level of universal law, but it does provide a useful illustration of some of the kinds of problems that can emerge.

Our first home desktop computer was made by the Vector Company of Thousand Oaks. We had a Vector IV that was made in 1980. Although not designed to deal with the Internet, it was perfectly functional as a desktop writing machine. It had, by today’s standards, a very small memory, and required the use of the old 4½-in. floppy disks to hold files. It had its own software system, designed by its own company engineers. We used it quite productively for a number of years, but later acquired an Apple desktop computer which also was Internet-accessible. Subsequent machines include a Gateway and a Dell.

Our Vector computer is no longer operational, though we have preserved it for a museum donation. Even our later Apple computer has broken down and has been discarded for recycling. When we want to access files we created on the Vector machine, we can no longer do it. Our newer computers cannot read the Vector software and cannot accept the old floppy disks. Fortunately we printed off all our important files while the Vector was still working, but if we had had irreplaceable files on that machine, their contents would have become lost forever. Are there machines still in operation that might still be able to read the Vector’s old disks? Possibly, but we have never found any that could, nor have any of the highly skilled computer specialists at Michigan State University. It should be appreciated that those disks are only 30 years old or less. What would be the case a hundred years from now? If we had had unique documents on them, such as site records or analysis papers, they would simply be lost.

A comparable problem emerged with our Apple computer, which as of this writing was purchased only 15 years ago. It became dysfunctional two years ago. No computer technician we contacted was able to recover files stored in its hard drive. Those files also are now lost forever, and some of them were created only three or four years ago.

These experiences illustrate the short life span of both physical technology and program design in the computer world. The continual evolution of computer technology and programming design illustrate the short life span of computer-based records. What was radically innovative 20 years ago is dysfunctionally antique today, and the same is most likely to be the case for today’s newest breakthroughs 20 years in the future.

Is it possible for old files to be updated as newer technology emerges? It certainly is, but doing so requires the continuation of institutions that are willing and able to support doing the needed work. As we archaeologists know only too well, sociocultural institutions are all-too-fragile, and one cannot count on them being able to persist and function for endless centuries into the future.

Would it be possible in the future for a scholar to extract files from ancient computers and make the contents available to the then-current academic community? It certainly could be possible; we cannot predict what future technological innovation will be. How probable would it be, however, that all the work contained in, say, the *Proceedings of the Society for California Archaeology* would be discovered, recovered, and made available? The probabilities can be incredibly low. We could not predict what papers and monographs would be rediscovered in that way, much less how relevant they would be for the furthering of the advancement of knowledge in the contexts of those future times, compared to those that were forever lost.

COMPARISON WITH LIBRARIES FOR SECURITY

It is appropriate to compare the security of information online over time with the security provided by library housing of the same materials. Libraries serve as repositories of printed materials. A library can be destroyed, with the loss of everything in its collections, so libraries are not inherently fully secure repositories. When, however, hundreds of copies of a manuscript are distributed among hundreds of libraries, the probability of future access to the manuscript increases immensely. Most online documents do not get printed off, bound, and shelved by libraries, so this level of back-up security of long-term preservation is not provided. Though libraries are less accessible than online sources, they are more durable (Mary Ver Plank, personal communication 2009). Think of the use by archaeologists of texts and carvings from ancient Egypt, Mesopotamia, and Greece, for example. Will there be reports available regarding what we have learned through the archaeological study of California's prehistory in another 2,000 or 3,000 years? If not, it is appropriate to consider what contributions are really being made to the growth of human knowledge through archaeological research.

SOME ISSUES IN PRODUCTIVITY AND ETHICS

This potential problem of the loss or permanent inaccessibility of online materials poses issues for both the advancement of knowledge and the ethics of archaeological practice. Let us think for a moment about some of the central purposes and values that motivate the practice of archaeology. As archaeologists, we seek to learn about the past through the recovery and analysis of physical remains that reflect patterns of past behavior. Archaeology can contribute enormously to our understanding of the historic past, since the written historical record is invariably incomplete. For the prehistoric past, which constitutes the overwhelming majority of the human past, the archaeological record is especially precious, because essentially there is no other comparable record. Thus we share an appreciation of the potential value of the practice of archaeology.

When construction projects damage or destroy archaeological sites, or when looters or pothunters mine the sites to gather antiquities, they destroy all the contextual information associated with the recovered remains. What gets destroyed is what could teach us about that part of the past, so it is lost forever. Archaeological studies or excavations also destroy the parts of the physical remains of sites that are being examined, but the remains and their contexts get observed, recorded, and studied, as do both the recovered remains and the recovered information being preserved, so the knowledge is not lost. Furthermore, both the collections and the records can be studied in the future, to allow reevaluation, to apply new technologies to the analysis of the data, and to do new comparisons.

Think about what we learn when we are able to reanalyze old collections using new technologies. Just as one example, today we can do trace element analysis of residues on prehistoric milling tools to identify what foods were actually processed by those tools, something that could not have been done when the tools were first recovered. We have already seen the revolutionary growth and changes in

knowledge gained from the analysis of stone tools by the application of microscopic edgewear analysis, obsidian hydration, and trace element analysis. If these old collections did not exist or were not accessible, such radical improvements in knowledge could not be gained.

Thus it is relevant to think about the consequences if something happens to destroy what has previously been recovered. Then knowledge has not been gained, and the work done in field excavation contributes no more than does pothunting. In such circumstances, how does the ethical standing of archaeologists compare with that of pothunters? Preservation of information is not only a matter of functional concern. It also is a matter of the ethical standards that justify what we do.

CONTRIBUTIONS AND COMPARATIVE FRAGILITY

In the practice of archaeology, it is not only that data collection is done systematically and that records are made, but that what is recovered becomes both preserved and shared. Museums preserve collections and their associated documents. But the knowledge gained is shared with the research community through the writing of reports and analyses and the sharing or spreading of them through publication.

Publication increases exponentially the capability of recovered information and ideas being able to survive stochastic events. If the only records happen to be in a particular museum repository (such as a collection catalogue), and the museum burns down or is otherwise destroyed, the information is lost forever. If the information is in 500 different locations worldwide, the probabilities of its survival improve by millions of times. At the same time, the dissemination of the contents of published work brings the data and ideas to many other creative minds, which vastly enriches understanding and the development of stronger models and ideas. The growth of knowledge in general benefits enormously from publication.

Think, however, about survival of knowledge over longer periods of time. Let us say, simply as a hypothetical model, that at some point 500-1,000 or more years in the future, our society, or current civilization more generally, no longer exists. Going back to my Vector analogy, what happens to the survival and stimulating influence of knowledge should it become inaccessible?

SOME POSSIBLE IMPROVEMENTS

Professional library archivists already are discussing these general issues, because they already have become concerned about the potential loss of valuable materials that exist only online. For those of us in archaeology, we need to consider options that (1) are economically affordable, and (2) that do not rely on the eternal continuation of technological help. Here are a few ideas that I believe should be discussed and considered across the profession.

When journals are posted online, it could be valuable if a limited number of hard copies of each issue also were printed and distributed among appropriate repositories. By "limited numbers," I would suggest a minimum of 25-50 copies. These printed copies should be filed in archival libraries and other appropriate repositories, preferably spread widely around the country and the world, not just around California. A single hard copy might well be preserved for many years, but it also runs the risk of being destroyed by any of a wide variety of stochastic events. Each additional copy raises the probability significantly that at least one copy would survive. Multiple copies, distributed in widely spread geographic and political settings, make it highly probable that at least some copies will be available for many generations. In this way, it will be very probable that what is discovered and learned today can make a comparatively permanent contribution to the growth of knowledge about the past, which is, of course, our intention to help justify what we are doing in research and scholarship today.

REFERENCES CITED

Allen, Mark

2008 From the President: A New Journal for the Society for California Archaeology! *Society for California Archaeology Newsletter* 42(3):4, 14.

Chartkoff, Joseph L.

1987 The Bleak and the Grey: Critical Issues of Publication and Scholarship in California Archaeology. Paper presented at the 21st annual meeting of the Society for California Archaeology, Fresno.