Digital Archaeological Data in the Wild West: the challenge of practising responsible digital data archiving and access in the United States

Christopher Nicholson, Rachel Fernandez and Jessica Irwin

Summary

Archaeology in the United States is conducted by a number of different sorts of entities under a variety of legal mandates that lack uniform standards for data archiving. The difficulty of accessing data from projects in which one was not directly involved indicates an apparent reluctance to archive raw data and supplemental information with digital repositories to be reused in the future. There is hope that additional legislation, guidelines from professional organisations, and educational efforts will change these practices.

1. Introduction

Though we are well into the 21st century, responsible digital archiving of archaeological data in the United States is not common practice. Digital archiving of cultural resource management reports in State Historic Preservation Offices, where they are often available by request though perhaps at a cost, is common; however, digitally archiving the datasets and other supporting materials that went into the creation of those documents is not. Though a vocal minority advocates for responsible digital archiving practices (Kansa and Kansa 2013; 2018; Kansa et al. 2019; Kintigh and Altschul 2010; Marwick et al. 2017; McManamon et al. 2017a; 2017b; Wells et al. 2019; Wilshusen et al. 2016), the US is facing a digital curation crisis (Kintigh and Altschul 2010; Majewski 2010; Thomson 2014; Witze 2019). Most archaeologists are simply not engaged in preserving the digital data that document their archaeological investigations. Here, we focus primarily on digital archiving practices related to datasets derived from analysis and research (i.e., artefact databases and catalogues, analysis records, feature logs, GIS maps and models, coding sheets, etc.) and other supporting digital materials
(i.e., images, 3D objects, etc.). We briefly discuss the landscape of digital curation attitudes, federal laws and regulations related to data preservation and archiving, current practices, and a vision for the future of digital archiving and data access in the US.

2. Digital Curation in the Academy and Cultural Resource Management

In the US, as in many other countries, the dichotomy in practices between academic archaeologists and cultural resource management (CRM) professionals is stark. As a generalisation, those in the academy conduct archaeological research that pertains to their topical/geographic area of interest, which is often rooted in a theoretical perspective. In contrast, CRM archaeologists are typically tasked with conducting compliance work in advance of development on federal land or on development projects that involve federal monies or consent. Federal laws and regulations require that these projects focus on identifying historic properties in a particular area and determine the potential effect of a project on those historic properties. Each approaches fieldwork, artefact analysis and report/manuscript writing from different perspectives. However, based on the number of American archaeologists who do not digitally archive materials, it is reasonable to say that the vast majority of archaeologists in the US are largely not concerned with what happens to digital data and information once a project is complete. This is likely the result of 1) how many American archaeologists learn the practice of archaeology with an emphasis on fieldwork and not on laws and regulations, report writing, or the realities of day-to-day practical cultural heritage management, and 2) the lack of enforcement of the legal requirements that are in place to protect archaeological information.

To begin with, archaeology education curricula across the US rarely include topics on the importance of, and how to go about, archiving, preserving, and reusing existing datasets to answer a variety of questions (Agbe-Davies et al. 2014). Archaeology undergraduate students receive some instructions on how to generate archaeological data through basic excavation, survey, and some artefact analysis (typically at field schools), but are not exposed to data analytics and statistics until their graduate careers. Even then, few receive instruction on best practices for data preservation and later reuse. Though taught to create data, graduate students rarely receive formal guidance related to its long-term preservation. Much of this is left to a do-it-yourself approach, which leads to academic archaeologists commonly submitting their datasets in the form of supplemental data in publications or saving their data in unstable and inaccessible digital media. The more data we produce, the more this lack of proper training and standards becomes a concern, especially as the prospect of losing digital information often becomes reality (Kintigh and Altschul 2010; McManamon et al. 2017a; Michener et al. 1997).

In professional practice, digital archiving and access to archaeological datasets is complicated by overlapping federal and state jurisdictions that make access opaque. So much so that in 2019, the Society of American Archaeology organised a task force to create guidelines and recommendations for accessing data from the variety of institutions that are required to hold these data (Wells et al. 2019). Submissions to these institutions are complicated by myriad factors, including 1) inadequate incentives that promote good archiving practices (Kansa and Kansa 2013), 2) financial and labour costs
of archiving, 3) Government agencies have yet to explicitly require curation of archaeological analytical documentation and supporting materials in a qualified digital archive and there are no simple means to assess penalties for non-compliance with current curation standards, and finally 4) not complying with ethical guidelines laid out by professional organisations. Despite these impediments, both real and perceived, there is a move by some CRM firms and governmental agencies responsible for cultural heritage management to use publicly available digital repositories (e.g., the Digital Archaeological Record [tDAR], OpenContext, and Digital Archaeological Archive of Comparative Slavery [DAACS]) to archive, and make accessible, their digital documents and datasets (Bates et al. 2020; Galle et al. 2019; Kansa et al. 2018; McManamon et al. 2017b; 2019; McManamon and Kintigh 2016; Schlanger et al. 2020).

3. US Federal Laws and Regulations

It has been estimated that 90% of archaeological investigations in the United States are done pursuant to Section 106 of the National Historic Preservation Act (Advisory Council on Historic Preservation 2018) with the vast majority of these conducted on federal lands, which account for about 28% of all land in the US (Congressional Research Service 2020). In 2012, Cultural Heritage Partners PLLC, a legal firm that specialises in cultural heritage law and policy, acting on behalf of Arizona State University, conducted a legal analysis of federal laws and regulations that pertain to data archiving (Cultural Heritage Partners 2012). They concluded that, under existing law, federal agencies must ensure the long-term preservation of and accessibility to objects and physical records from federally mandated archaeological investigations. In addition to being accountable for site preservation, these agencies are entrusted to make archaeological records and data available to the public, with appropriate controls in place to protect these archaeological assets (Cultural Heritage Partners 2012). To comply with these requirements, some federal agencies have their own curation facilities, such as the Curation and Management of Archaeological Collections (CMAC) for the US Army Corps of Engineers. Other federal agencies, such as the Bureau of Land Management, curates their artefacts and data in state-sponsored facilities. However, these facilities are primarily artefact repositories and few are staffed or equipped to curate digital data responsibly (Watts 2011).

While these efforts primarily account for the curation of physical objects, the needs of digital data are often not met. The primary federal requirements for the management of archaeological collections and associated records (including digital data) are set forth in the National Historic Preservation Act (1966) (NHPA), the Archaeological Resources Protection Act (1979) (ARPA), and regulations pursuant to those statutes including 36 CFR 79, the Curation of Federally-Owned & Administered Archaeological Collections, and the regulations put forth by the National Archives and Records Administration for the management of federal records (36 CFR 1220.1-1220.20). Here we briefly discuss the NHPA, ARPA, and 36 CFR 79.

The NHPA established permanent institutions, such as the State Historic Preservation Office, in each US state, and created a clearly defined process for federally mandated historic preservation in the United States. In 54 USC §306131(a)(1)(c) of the NHPA, professional standards for data management are laid out. They state: 'Records and other data, including data produced by historical research and archaeological surveys and excavations are permanently maintained in appropriate data bases and made available to potential users…' This language serves as a cornerstone in the long-term
preservation of information and data derived from archaeological investigations and is as important today as it was when the law was enacted in 1966.

ARPA recognises that archaeological resources are an irreplaceable part of America's heritage and that they were becoming endangered as a result of the escalating commercial value of artefacts from archaeological sites on Federal and Tribal Lands. Regarding the preservation of information and data, Section 4, paragraph (3) states that 'the archaeological resources which are excavated or removed from public lands will remain the property of the United States, and such resources and copies of associated archaeological records and data will be preserved by a suitable university, museum, or other scientific or educational institution'; recognising that archaeological records, and data, must be preserved in an appropriate repository.

Finally, NARA regulations require the protection of digital archaeological records from destruction, including from technological obsolescence. These regulations require agencies '… to counteract hardware and software dependencies of electronic records whenever the records must be maintained and used beyond the life of the information system in which the records are originally created or captured'. Focusing curation efforts on digital storage media alone also fails to meet the standards for digital curation expressed in 36 CFR 79 for three reasons (Cultural Heritage Partners 2012; Schreibman et al. 2015):

1. data are at risk of being lost because the physical digital media is subject to degradation,
2. treating the physical digital media as artefacts to be preserved renders the data inaccessible to the vast majority of potential users, and
3. the format of digital information will become unusable owing to software and hardware advances.

These laws are an important part of cultural heritage management and history, and understanding how they impact archaeological investigations is important for the long-term preservation of these irreplaceable resources. Unfortunately, the implementation of these laws and regulations has been far from uniform.

There is no doubt that there is language in the aforementioned federal laws and regulations that mandate digital archaeological documents and data, generated by federal projects, should be deposited and curated in a trusted repository, one with the technological capacity to truly provide a long-term solution to preserving digital files. Federal agencies responsible for curating artefacts are also responsible for curating data, not just the media on which they reside, to ensure that they are available for future reuse and research.

4. Current Digital Archiving Practices

While many archaeologists in Europe recognise a potential 'Digital Dark Age' (Wright 2020), many American archaeologists (academic and CRM), are still reluctant to work toward archiving and preserving digital information and datasets in repositories for reuse. This perceived reluctance in archiving digital data is demonstrable with User and Contributor data from the tDAR repository (the Digital Archaeological Record). tDAR is a digital repository for archaeological data, whose use, development, and maintenance are overseen by Digital Antiquity, a research centre at Arizona State University. By
means of user and contributor data gathered when an individual signs up to either upload or download information from tDAR, we can evaluate some digital data archiving patterns among professional archaeologists in the US.

Table 1: The number of tDAR User and Contributors and their self-identified affiliation compared to the estimated total number of archaeologists (n = 7,600) in the United States (as of 06/01/2020).

<table>
<thead>
<tr>
<th>Profession</th>
<th>Users</th>
<th>Contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher Education Faculty</td>
<td>974</td>
<td>22</td>
</tr>
<tr>
<td>CRM archaeologists</td>
<td>633</td>
<td>11</td>
</tr>
<tr>
<td>Public Agency archaeologists</td>
<td>549</td>
<td>10</td>
</tr>
<tr>
<td>Native American/Indigenous researchers</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2,322</strong></td>
<td><strong>43</strong></td>
</tr>
<tr>
<td><strong>Per cent of American archaeologists</strong></td>
<td><strong>30.6%</strong></td>
<td><strong>0.6%</strong></td>
</tr>
</tbody>
</table>

Using the number of tDAR users and contributors, and their self-identified affiliation (Table 1), along with an estimate of 7600 professional archaeologists in the US (https://www.careerexplorer.com/careers/archaeologist/job-market/), we examine the percentage of professionals using at least one digital repository (tDAR) to archive datasets and information. As of August 2020, there are 43 professional/student contributors (0.6%), though there are 2322 (31%) professional/student archaeologist users. At least in regards to the tDAR repository, most professionals use the service to find and access information, but are less inclined to contribute.

While tDAR is not the only repository where archaeological data can be properly archived, if the data are being archived elsewhere, then accessing, let alone finding, this material has been an issue (Robinson et al. 2019; Wells et al. 2019). This may be related to the fact that digital curation of archaeological documents and datasets derived from Section 106 compliance work tends to be regionally specific. Further, State Historic Preservation Offices (SHPOs), are charged with maintaining inventories of historic properties and evaluating those sites for nomination to the National Register of Historic Places. Unfortunately, many of these offices are severely understaffed when it comes to handling and managing large volumes of digital information, with only 40% of SHPOs having their own IT staff, while 7% have no IT support at all (National Conference of State Historic Preservation Officers, pers. comm. 2020). Some SHPOs have fee-based retrieval systems that allow them to generate some revenue to offset the costs related to maintaining these digital tools and resources, but these fees go to directly maintaining Section 106 data.

SHPOs vary greatly with respect to requiring digital documents and data archiving (including geospatial files) as part of their formal submission process. Like the Federal artefact repositories, few are staffed or equipped to curate digital data responsibly. Some SHPOs are also their state’s curation facility for artefacts and data. About 18 of
the 59 SHPOs (including US Territories) have agreements with local universities to serve as the state’s digital archive for archaeological materials, but the degree to which this extends to digital curation is unknown. Those SHPOs without curation responsibilities or facilities rely on the federal agencies to fulfil their requirements for curation of digital data and artefacts.

![Figure 1: a) tDAR file repository size, in megabytes from 2014 to 2019, b) the number of downloaded resources annually from tDAR from 2010 to 2019, c) the number of datasets curated from 2014 to 2020](image)

Despite these inconsistencies, there has been some growth in this area. Since 2014, resources archived in tDAR, for example, have grown from just under 0.6 to more than one terabyte (TB) and the number of datasets from 613 to 1901 (Figure 1). And, even more importantly, these resources are being reused! From 2017 to 2019, on average 25,000 resources were downloaded annually from tDAR (Figure 1); the number of citations, as measured by Google Scholar search results, increased from 11 in 2012 to 60 references in 2019. The growth in deposited, downloaded, and reused resources is promising, though more needs to be done not only to comply with federal laws and regulations, but also to extend our knowledge of the human past and improve the general management of our cultural heritage (Kintigh et al., 2014). Though there is some evidence to support the idea that it is becoming more common, there is still a tremendous amount of ground to cover.
5. Future Directions

Beyond being a useful tool for organising, preserving, and providing access to digital archaeological information, online digital repositories can serve as critical tools for federal and state agencies in complying with existing federal regulations. Using such regulations, archaeological organisations, practitioners and repositories in some cases have instituted reasonable practices to ensure digital data access and long-term preservation. Across the sciences, datasets and other supporting materials associated with publications commonly go either unarchived, unsubmitted when requested, or the ability to access those data, which are often in the form of supplemental data, is lost (Maunsell 2010; Miyakawa 2020; Richards 2015; Sholler et al. 2019; Vines et al. 2014). It is our estimate that the vast majority of datasets and supporting materials that make up the digital archaeological record, produced from compliance and academic archaeology, go digitally uncurated. And that needs to change.

In order to combat this trend, we advocate, first, for the adoption and institution of the FAIR principles (findable, accessible, interoperable, and reusable) by federal and state agencies charged with overseeing the Section 106 process (Boeckhout et al. 2018; Jacobsen et al. 2020; Koster and Woutersen-Windhouwer 2018; Wilkinson et al. 2016; 2019). Furthermore, we contend that there are specific, actionable items that can be done and/or instituted in the US, sooner rather than later, to expand preservation of and access to our digital cultural heritage. These include:

1. Incentivise archiving practices in the academy by educating anthropology faculty and departments/schools on the advantages of including data archiving in a similar (maybe not equal) manner as a technical report in the tenure process.
2. Through outreach efforts, organisations like Digital Antiquity or OpenContext, can promote data preservation standards, techniques and methods, archival principles, and reuse techniques at the undergraduate and graduate level. This can be done by including data curation as part of their core curriculum for graduate and undergraduate levels.
3. Professional archaeological organisations can help lobby federal agencies to mandate that all companies submit data to repositories as part of their final deliverables, thus becoming a standard part of the contracting and bidding process in the Section 106 compliance process.
4. Professional archaeological organisations (Society for American Archaeology, Society for Historical Archaeology, American Cultural Resources Association, and regional organisations) can provide ethical and well as practical guidelines and emphasise their ethical guidelines related to data archiving and give their constituents options for archiving (Wells et al. 2019).
5. SHPOs and THPOs (Tribal Historic Preservation Office) that lack sufficient cyber-infrastructure can partner with existing digital repositories to archive and share, when appropriate, data from federally mandated archaeological investigations.
6. Grant-funding agencies (National Science Foundation, National Endowment for the Humanities, Wenner Gren, etc.) should implement a ‘scoring’ system whereby those individuals/groups who have been awarded grants and archived their digital materials in the past, or have strong data management plans, are scored higher. The inverse of this scenario is also a possibility.
7. All publishers should require submission of project data in repositories, and not just as supplemental data.
6. Conclusion

Though there is a positive shift in attitude towards digitally archiving archaeological data and information from archaeological investigations, we have a long way to go. Archaeologists from the academy and CRM are not taking advantage of digital repositories for a variety of reasons, but awareness of issues surrounding the loss of digital archaeological data is increasing. Through the implementation of the FAIR Principles, including through implementing the seven recommendations listed above, it is our hope that archaeologists in the US will take action and vigorously work toward the long-term preservation of the digital documentation of our irreplaceable archaeological resources.

Acknowledgements

We would like to thank Karyn de Dufor of the Nevada State Historic Preservation Office for comments and suggestions, as well as the National Conference of Historic Preservation Officers for sharing information about digital archiving practices within their organisation. We would also like to thank Keith Kintigh, Sarah Schlanger and Cynthia Herhahn for their valuable input. All assertions and errors are the responsibility of the authors.


Cultural Heritage Partners 2012 Federal Laws and Regulations Requiring Curation of Digital Archaeological Documents and Data, Arizona State University, Tempe, AZ.


Kansa, E.C. and Kansa, S.W. 2013 'We all know that a 14 is a sheep: data publication and professionalism in archaeological communication', Journal of Eastern Mediterranean Archaeology & Heritage Studies 1(1), 88-97. https://doi.org/10.5325/jeasmedarcherstu.1.1.0088


Majewski, T. 2010 'Not just the end game anymore', Heritage Management 3(2), 167-88. https://doi.org/10.1179/hma.2010.3.2.167

Maunsell, J. 2010 'Announcement regarding supplemental material', The Journal of Neuroscience 30, 10599-600.


McManamon, F.P., Doershuk, J., Lipe, W.D., McCulloch, T., Polglase, C., Schlanger, S., Sebastian, L. and Sullivan, L. 2017a 'Values-based management of archaeological resources at a landscape scale', Advances in Archaeological Practice 4, 132-48. https://doi.org/10.7183/2326-3768.4.2.132


Richards, J.D. 2015 'Preservation and re-use of digital data: the role of the Archaeology Data Service', Antiquity 71, 1057-59. https://doi.org/10.1017/S0003598X00086014


