

Guidance to Ensure Proper Curation of Digital Archaeological Records

Background

Federal agencies have a legal responsibility to care for archaeological collections resulting from investigations that they conduct or require. Digital data and records are parts of these collections and must be curated properly (for a detailed legal analysis of this topic see Cultural Heritage Partners 2012¹). Digital files require different kinds of care and procedures than physical collections to ensure that they are properly preserved and accessible for appropriate uses. The nature of digital curation is not necessarily more complicated or expensive than physical collections. However, it is specialized and agencies need to take affirmative steps to ensure that the archaeological data about their resources and from their projects are deposited in an archive or repository where the expert care, principles, standards, and techniques of digital curation are followed. To ensure that digital archaeological files are properly cared for, agencies need to adopt guidelines like those described below.

Guidelines for Good Digital Curation

Agencies need to have guidance stating that digital information from archaeological investigations for which they are responsible are deposited in and cared for by a digital archive or repository. The following paragraphs describe the kind of activities, procedures, and standards that agencies should include in guidance for their professional staff and contractors who carry out or oversee archaeological studies, and of archives and repositories where they deposit their digital archaeological data^{2,3}.

The process of digital archiving for archaeological data can be divided into two general sets of actions. The first includes activities to be taken by the individual(s) or organization(s) that create the digital data, documents, image, or other kinds of digital files. The second set of activities includes those undertaken by the digital archive or repository where the digital data are deposited.

Actions by digital data creators (depending on the agency, these may include agency archaeologists and/or contractors working for the agency):

- Plan for the creation and subsequent management of the digital resources as part of archaeological investigations.
- Produce the digital resources as part of the project, creating administrative, substantive and technical descriptions of the digital objects, commonly referred to as “metadata.”
- Provide the means for others to access and make use of the digital files.
- As part of the project that created the digital files, evaluate their continuing importance and select objects that merit long-term preservation, in consultation with the appropriate agency official.
- Dispose of the digital files not selected for long-term preservation, in consultation with the appropriate agency official.
- Deposit in a digital repository the digital files selected for long-term preservation.

A digital repository is one established and operated for the express purpose of providing access to and long-term preservation of digital data. Such a repository is organized so that it can be sustained and

function in its curation role indefinitely. A digital archive or repository typically has a professional staff that carries out activities necessary to ensure the long-term preservation of and appropriate access to the digital files it curates. References with more detailed descriptions of the activities, policies, procedures, and standards of such archives or repositories are Digital Curation Centre² and Center for Research Libraries and Online Computer Library Center⁴.

Actions by digital repositories include:

- Upon deposit of the digital files, test the objects to ensure that they have the characteristics described by the creators and can be adequately preserved in the repository.
- Undertake actions, such as regular checks for file stability and retention of integrity; regular and frequent backups and storage in multiple locations for security and safety of files; etc., to ensure preservation of the digital files in the repository.
- Implement the policies and procedures necessary for the long-term preservation of digital Federal records as provided in Federal curation regulations (36 C.F.R. Part 79), such as:
 - regularly and systematically checking the files in the repository to ensure that no deterioration has occurred;
 - taking actions to remedy deterioration if it is detected; and,
 - periodically migrate the digital files to new file types or in order to conform to new hardware and/or software standards.
- Manage digital project data such as reports, data sets, photographs and other graphic images, GIS, and LiDAR and other remote sensing data.
- Ensure the cross-referencing between physical collections and digital records
- Keep the digital files stored in a secure manner.
- Provide the means for access and use of the digital files, within any constraints placed on use by the depositors (e.g., permits access to “confidential” digital data to be restricted).

Agency guidance on proper curation of its digital archaeological files should be strongly stated and apply to data from all of the archaeological activities in which it is involved. The guidance needs to include a description of the kinds of activities, procedures, and standards associated with appropriate digital curation such as provided in this section.

Procedures for Depositing Data in a Digital Curation Archive or Repository

To implement the policy for proper digital curation, agency managers and archaeologists, or archaeological contractors working for the agency, need to include in the scopes of work for their investigations requirements that at the end of the project, the digital data generated by the investigation be placed in an appropriate digital archive or repository.

The archaeologists responsible for the investigations that create the digital data also must undertake activities required for good data curation, as described in the policy section above. These actions include: planning for the creation and management of the digital resources; create the digital resources as needed for the administration, substantive, and technical parts of the project; provide clear descriptions of the digital files (the “metadata”); provide the means for users to access the files; evaluate the likely continuing importance and select those files that merit long-term preservation, in consultation with agency officials; and, finally, deposit in a digital repository the digital files selected for long-term preservation.

Most Requests for Proposals (RFPs), scopes of work (SOWs), contracts, and other kinds of agreements for archaeological investigations undertaken, funded, or required by federal agencies require the curation of collections that these investigations create. Given the ever-increasing amount of digital data generated by

these projects⁵, it is important to curate digital archaeological data in digital repositories where they can be accessed, cared for, and preserved properly.

Currently, most RFPs, SOWs, contracts, and agreements do not specify the requirements for digital curation. As a result, digital data, such as field records, images, laboratory records, data sets resulting from field and laboratory analyses, and Geographic Information System (GIS) maps, are stored on CDs or other digital media within a curatorial facility that focuses on curating material remains. The digital records are treated the same as paper records and artifacts. Such curation practices neither preserve digital data nor make it accessible; CDs and other digital media degenerate over time, are not readily accessible to users, and will eventually become obsolete as computer hardware and software change^{6,7}.

To ensure the accessibility and preservation of digital archaeological data, agency officials preparing RFPs, scopes of work, contracts, and other kinds of agreements need to include specific requirements to ensure that digital curation is an explicit project deliverable, along with the curation of artifacts and other materials generated by the project.

The following is example language on digital curation that can be included in a RFP, SOW, contract, or agreement. This example identifies the Digital Archaeological Record (tDAR) as the repository for digital curation. However, if an agency is unable to require use of a specific digital curation archive or repository, a more general reference can substituted. In such cases, officials drafting the scope of work, contract, RFP, or agreement will need to include sufficient description of the characteristics of an appropriate digital archive or repository. It is our hope that managers will also suggest to their contractors that tDAR would be an appropriate repository for these digital data.

Example of scope requirements for digital curation:

1. [Name of entity conducting the archaeological work] shall deposit all copies of digital data listed as deliverables for this project in [location of description of digital project deliverables in RFP, scope of work, contract, etc.], in tDAR, the Digital Archaeological Record repository (www.tdar.org).
2. [Name of entity conducting the archaeological work] shall thoroughly document all digital data with the following archaeological, administrative, and technical metadata, using the tDAR metadata creation and file upload web pages available at: <http://www.tdar.org/why-tdar/contribute/>.
3. [Name of agency/office] will not consider the project complete until the project's digital records in tDAR have been reviewed by [name of agency official and/or position title].
4. Any file containing information that is "confidential," for example as defined in Section 9 of the Archaeological Resources Protection Act (16 U.S.C. 470hh), or "restricted," as defined in consultation with [Name of agency/office] during the execution of this project shall be deposited in its complete form and marked in tDAR as confidential and shall also be deposited in a redacted, public form, with redactions of all confidential information identified.

Why Use tDAR (the Digital Archaeological Record)?

Since September 2011, tDAR (the Digital Archaeological Record) has been a fully functioning on-line digital repository for archaeological information with a focused and skilled professional staff. The Center for Digital Antiquity, a unit of Arizona State University with an independent Board of Directors, developed and maintains tDAR. tDAR has a growing number of registered users (2985 at present) and content (over 14,000 documents, images, and data sets and over 350,000 citation records enhanced and incorporated from the National Archaeological Database).

For archaeological data from the US and most international contexts, there is no viable alternative to tDAR as a disciplinary digital repository. At the University of York in England, the Archaeology Data Service (ADS) maintains an archaeological digital repository, but it includes only data from the United Kingdom (UK) archaeological contexts or data that are generated by UK researchers. ADS and tDAR do not compete and have partnered on several projects. In the US, the OpenContext web-publishing site, which once advertised itself as a digital repository, has substantially modified its services over the last few years; it now focuses squarely on data publication. It uses other organizations to archive data and, indeed, OpenContext has approached Digital Antiquity about using tDAR as the archive for some projects. OpenContext requires completely open access to the data it publishes and cannot protect confidential information, e.g. specific site locations, a function that is essential for public agency use.

There are general-purpose digital repositories, including those operated by universities for data their faculty create or utilize. However, many of these either do not accept or do not effectively document the more complex data types that archaeologists collect. Because of their general-purpose nature, these repositories cannot offer the functionality that tDAR provides for archaeological data. While they maintain standard technical metadata, they include only very general substantive metadata, seriously limiting both information discovery and reuse. tDAR, on the other hand, allows for the inclusion of detailed substantive metadata specifically tailored for archaeology and for the administrative and management needs of the federal agency. This metadata is essential for data discovery, reuse and preservation, especially for systematically recorded databases. tDAR structures information and provides a user interface designed for archaeologists and the managers of archaeological information.

Digital data that are not curated effectively are highly fragile, subject to complete loss due to media degradation, software and hardware evolution, and inadequate metadata. Few if any traditional artifact curation facilities are providing or capable of providing anything like the federally mandated level of digital data curation and access. tDAR was explicitly designed to fill this void; records in tDAR are preserved and made accessible in accordance with federal laws and regulations.

¹ Cultural Heritage Partners (2012). *Legal Analysis of Federal Requirements for Curation of Digital Archaeological Documents and Data*. Report to the Office of General Counsel, Arizona State University, Tempe, AZ. Accessed 21 February 2013 at <http://www.digitalantiquity.org/CHPFinal.pdf>.

² Digital Curation Centre (2010). What is Digital Curation? Digital Curation Center, The University of Edinburgh, Scotland. [<http://www.dcc.ac.uk/print/node/515> accessed, 8 Sept 2010].

³ Inter-University Consortium for Political and Social Research (2012). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle*, (5th Edition). Ann Arbor, MI.

⁴ Center for Research Libraries and Online Computer Library Center (2007). *Trustworthy Repositories Audit and Certification: Criteria and Checklist, Version 1.0*. Center for Research Libraries, Chicago, IL. [accessed 21 February 2013, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf].

⁵ Petrovic, V., A. Gidding, T. Wypych, F. Kuester, T.A. DeFanti, and T.E. Levy (2011). Dealing with Archaeology's Data Avalanche. *Computer* 44(7):56-60.

⁶ Hedstrom, Margaret (1998). Digital Preservation: A Time Bomb for Digital Libraries. *Computers and the Humanities* 31: 189-202.

⁷ Spindler, Robert P. (2007). Digital Preservation. *Preservation Leaflets*. Northeast Document Conservation Center, Lowell, MA.