How to Ensure Proper Curation of Digital Archaeological Records
Center for Digital Antiquity -- 2015

Background

In the United States, federal and other public agencies have legal responsibilities to care for archaeological collections resulting from investigations that they conduct or require. Digital data and records are a part of these collections and also must be curated properly (for a detailed legal analysis of this topic see Cultural Heritage Partners 2012). In addition, a series of Executive Orders, guidelines, and policies require federal agencies to make the results of research that they conduct or fund more easily accessible by the public (e.g., Holdren 2013).

All, or nearly all, of these archaeological research results are in digital formats. Digital files require different 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 the curation of physical collections. However, it is specialized and agencies are obligated 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 (Richards et al 2010; McManamon 2014; Kintigh et al. 2015).

Digital files have become a focus for curation because most data and information created by contemporary research in all subjects is created, stored, and most easily shared in digital formats. Even documents and other texts that are still published on paper are most commonly accessed and shared electronically as digital files. Digital data and documents are far easier to share than the same information in paper format. However, like paper records, although in different ways, digital files can be damaged or destroyed if handled inappropriately. Importantly, and unlike paper records, unless properly curated digital data can become obsolete and inaccessible rapidly.

In order to take advantage of the wealth (or "deluge" to cast it in a different light) of digital data, appropriate procedures are needed to care for the data after it is created (e.g., see Hey and Trefethen 2003; Lord et al. 2004; Seltzer and Zhang 2009; and, Faniel and Zimmerman 2011). The appropriate activities and procedures can be grouped under the general heading of "digital curation." Good digital curation is not simply a legal and regulatory requirement in many circumstances. By making digital data easily discoverable and more accessible, digital curation greatly enhances the ability of other researchers to test and build upon work that has been done by others. Replication or refinement of research results by subsequent studies is a hallmark of scientific knowledge. Studies invariably build upon what has been learned from research done in the past. By improving discoverability and access to existing data, digital curation also enables current research projects to avoid unnecessarily redundant studies and to build upon results that are available.

Agencies and other organizations need to adopt guidelines like those summarized in this guideline and the references that it cites in order to ensure that the archaeological data for which they are responsible is cared for properly.

Guidelines for Good Digital Curation

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 the agency 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. The agency guidance should contain a requirement that digital data and information produced by agency-related archaeological investigations be deposited in a digital archive or repository. Digital archaeological data includes: documents (e.g., field/lab notes, interim/final reports, specialist reports, correspondence, etc.), images (e.g., maps, drawings, photographs, etc.), data sets, geospatial data, scanned data files (e.g., 3D, LiDAR, etc.), and other types of digital files.

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 types of digital files. The second set of activities includes those undertaken by the digital archive or repository where the digital data are deposited.

For Digital Data Creators
Actions by digital data creators (depending on the agency, these may include agency archaeologists, contractors working for the agency, contractors working for other organizations that are required by the agency to conduct archaeological investigations, or some combination of these categories):

1. Plan for the creation and subsequent management of the digital resources as part of archaeological investigations.
2. Produce the digital resources as part of the project, creating administrative, substantive and technical descriptions of the digital objects, commonly referred to as "metadata."
3. Provide the means for others to access and make use of the digital files.
4. Evaluate the continuing importance of digital files during the project and select objects that merit long-term preservation for future uses, in consultation with the appropriate agency official.
5. Dispose of the digital files not selected for long-term preservation, in consultation with the appropriate agency official.
6. Deposit the digital files selected for long-term preservation in a digital repository where the data can be discovered, accessed (with appropriate controls), and preserved for future uses.

For Digital Data Repositories
A digital repository is one established and operated for the express purpose of providing access to and long-term preservation of digital data. Digital files related to archaeological resources or studies can be curated by the organization that generates the files or by a different domain or institutional digital repository, as so long as the required criteria for services are met. Such a repository is organized so that it can be sustained and function in its curation role indefinitely. A digital archive or repository 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. More detailed descriptions of the activities, policies, procedures, and standards of such archives or repositories are available from the Digital Curation Centre (2010) and the Center for Research Libraries and Online Computer Library Center (2007).

Actions by digital repositories that hold archaeological data should include:

1. 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.
2. Implement the policies and procedures necessary for the long-term public access and preservation of digital records (Cultural Heritage Partners 2012:2-11) as provided in federal curation regulations (36 C.F.R. Part 79) and records management requirements (e.g., 36. C. F. R. 1220), such as the following:
   a. Regularly and systematically checking the files in the repository to ensure that no deterioration has occurred.
   b. Taking actions to remedy deterioration if it is detected.
   c. Periodically migrate the digital files to new file types or in order to conform to new hardware and/or software standards.
   d. Regularly backing-up and storing e files in multiple locations for security.
3. Ensure cross-referencing between physical collections and digital records.
4. 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).
5. Enable data depositors to easily manage their data in the repository.

Procedures for Ensuring the Deposit of Data in an Appropriate Digital Curation Repository

In the United States, most Requests for Proposals (RFPs), Scopes of Work (SOWs), contracts, and other types of agreements for archaeological investigations undertaken, funded, or required by federal agencies require the curation of physical collections (including associated documents) that these investigations create. The ever-increasing amount of digital data generated by such projects (ADS and Digital Antiquity 2013:9-20) emphasizes the importance of effective curation for these data in digital repositories where they can be accessed, cared for, and preserved properly for future uses. This is essential because there may be no other record of the investigations and the archaeological resources investigated. Digital files also are far easier to share. They can be made available widely (if appropriate) via the Internet, which enables the public agencies responsible for the data to meet their requirement to make such information easily available for the public (Cultural Heritage Partners 2012).

To implement policies for proper digital curation, it essential that RFPs, SOW, contracts, and other agreement documents require that as part of archaeological projects, digital data that has been generated by the investigation is placed in an appropriate digital archive or repository.

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

Current and future RFPs, SOW, contracts, and other agreements must specify the requirements for digital curation. Digital data, such as field records, images, laboratory records, data sets resulting from field and laboratory analyses, and Geographic Information System (GIS) files and maps, should not be stored on CDs or other digital media as the sole or primary means of providing access and preservation within a physical curatorial facility that focuses on curating material remains. The digital records cannot be treated the same as paper records and artifacts. Such curation practices neither preserve digital data nor make it accessible. CDs and other digital media deteriorate over time. Data stored on these media are not readily discoverable or accessible

to users and will eventually become obsolete as computer hardware and software change (Digital Curation Centre 2010; ICPSR 2012; ADS and Digital Antiquity 2013).

The archaeologists responsible for the investigations that create the digital data should incorporate activities that are necessary to ensure that good data curation can be undertaken easily at the end of their projects. Such actions include: planning from the beginning of each project for the creation and management of the digital data; providing clear descriptions of the digital files (the "metadata"); using digital file formats that are standard and open source; evaluating the likely future importance of various types of data and select for curation those files that merit long-term access and preservation, in consultation with agency officials; and, finally, deposit in a digital repository the digital files selected for long-term preservation.

In order to ensure proper curation of the digital products from archaeological investigations, the data should be deposited in a repository that has appropriate staff expertise, computer hardware and software capacity, and established digital curation policies and procedures. Specific functions and services of such repositories include the following:

1. Overall Repository Functions
    a. Long-term preservation
       The repository has a sound business plan to ensure its continuation into the long term future.
    b. Access to the deposited files into the future
       The repository ensures that deposited files will be findable. The archival content of the repository should be discoverable and accessible not only within the repository itself, but also via general search engines, from outside of the repository. The repository should have in place procedures to ensure that deposited files will remain readable into the long term future.
2. Active File Maintenance Services
    a. Authentication
       The repository monitors all digital files and can guarantee that the file a user sees is that same as the file you deposited.
    b. Permanent identifier
       The repository assigns a unique identifier to each file which allows users to both access and accurately and uniquely cite the work (e.g., persistent URLs; digital object identifiers; or other).
    c. Maintain files in open and preferable formats
       The repository stores the original file, but also has policies and systems in place to create new derivative files (copies) in open, non-proprietary formats to ensure accessibility into the future.
    d. Rich administrative, descriptive, and technical metadata are linked to the deposited files
       The repository has an easy to use, but robust system of text, keywords, and other descriptive information to ensure that your files will be found when users perform a web search.
    e. Indexed by major search engines for discoverability
       The repository content should be indexed by major search engines, such as Google, to ensure that your files are discoverable.
    f. Collect file and project statistics and metrics
       The repository collects usage statistics and other metrics (e.g., page views, downloads) for deposited files to allow tracking of content usage.
    g. Provide options to manage privacy and confidentiality of files
       The repository provides services such as embargo dates, or marking files "private" to allow depositors to manage access to files in the event that some files require limited access.
    h. Regularly Check File Integrity

The repository has automated and/or human supervised systems in place to regularly open and review stored files, much like a museum periodically inspects physical objects to prevent decay.

3. Long-term Preservation Services and Planning
    a. Migrate file formats as they become obsolete
       The repository stores the original file you contribute but creates a derivative copy as needed to keep up with changing software and hardware.
    b. Plan for obsolete technology
       The repository monitors changing software and hardware systems and updates its systems to stay current. The repository can retrieve files stored on obsolete media and/or in obsolete software systems.

Agreement Documents, Scopes of Work, or Project Deliverables for Digital Curation

The following is example language on digital curation that can be included in an agreement, contract, RFP, or SOW for archaeological investigations or other types of cultural resource studies. This example requires use of tDAR (the Digital Archaeological Record). A modified version of this text could serve for seeking general archaeological repository services, in the event that the use of tDAR cannot be a specific requirement. However, if an agency is unable to require use of a specific digital curation repository, a more general reference to the required capabilities of the potential digital repository must be included in the solicitation. In such cases, officials drafting the RFP or proposed SOW must include sufficient description of the characteristics of an appropriate digital archive or repository. These characteristics are summarized in the two preceding sections.

1. [Name of entity conducting the archaeological work] shall deposit the 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, 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], approved, and made active.

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.

Using tDAR (the Digital Archaeological Record) for Digital Curation

The Digital Archaeological Record (tDAR) is a 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 (over 8,000 at present) and virtual visitors (about 50,000 page views and over 1,000 downloaded files per month). Content of tDAR includes nearly 30,000 documents, images, data sets, 3D/scan and geospatial data, as well as over 365,000 citation records incorporated from the National Archaeological Database [NADB] with enhanced metadata.

For archaeological data from the United States and many 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.

There are general-purpose, "institutional," digital repositories, including those operated by universities for data their faculty or students 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 or the disciplinary specific metadata 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 (Cultural Heritage Partners 2012).

References Cited

ADS (Archaeology Data Service) and the Center for Digital Antiquity
2013   *Caring for Digital Data in Archaeology: A Guide to Good Practice*. Oxbow Books, Oxford, England and Oakville, Connecticut. A website with the information in this publication (in a different format), plus more detailed descriptions and suggested procedures for additional digital file formats is at: http://guides.archaeologydataservice.ac.uk/. Accessed 19 November 2015.

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. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf. Accessed 21 February 2013.

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. http://www.digitalantiquity.org/CHPFinal.pdf. Accessed 21 February 2013.

Digital Curation Centre
2010   What is Digital Curation? Digital Curation Center, the University of Edinburgh, Scotland. http://www.dcc.ac.uk/digital-curation/what-digital-curation. Accessed 20 November 2015.

Faniel, Ixchel M. and Ann Zimmerman
2011   Beyond the Data Deluge: A Research Agenda for Large-scale Data Sharing and Reuse. *International Journal of Digital Curation* 6(1):58-69.

Hey, Tony and Anne Trefethen

    2003   The Data Deluge: An e-Science perspective. In *Grid Computing – Making the Global Infrastructure a Reality*, edited by F. Berman, A. Hey, and G. Fox, pp. 809-824. http://users.ecs.soton.ac.uk/ajgh/DataDeluge(final).pdf. Accessed 19 November 2015.

Holdren, John P.

2013   *Increasing Access to Results of Federally Funded Scientific Research*. Memorandum (22 February 2013) Office of Science and Technology Policy, White House, Washington, DC. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf. Accessed 16 November 2015.

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. http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/index.html. Accessed 19 November 2015.

Kintigh, Keith W., Jeffrey H. Altschul, Ann P. Kinzig, W. Frederick Limp, William K. Michener, Jeremy A. Sabloff, Edward J. Hackett, Timothy A. Kohler, Bertram Ludascher, and Clifford A. Lynch

2015   Cultural Dynamics, Deep Time, and Data: Planning Cyberinfrastructure Investments for Archaeology. *Advances in Archaeological Practice* 3(1):1-15.

Lord, Philip, Alison MacDonald, Liz Lyon, and David Giaretta

2004   From Data Deluge to Data Curation. Proceedings of the UK e-Science All Hands Meeting 2004. http://www.researchgate.net/publication/236870393_From_Data_Deluge_to_Data_Curation. Accessed 19 November 2015.

McManamon, Francis P.

2014   Digital Archaeological Data: Ensuring Access, Use, and Preservation. *Encyclopedia of Global Archaeology*, Claire Smith, general editor, pp. 2124-2128. Springer Science+Business Media, New York, NY. https://core.tdar.org/document/378289/digital-archaeological-data-ensuring-access-use-and-preservation. Accessed 19 November 2015.

Richards, Julian D., Tony Austin, and Catherine Hardman

2010   Covering the Costs of Digital Curation. *Heritage Management* 3(1):255-263.

Seltzer, Michael L. and Lei Zhang

2009   The Data Deluge: Challenges and Opportunities of Unlimited Data in Statistical Signal Processing. http://research.microsoft.com/pubs/78712/0003701.pdf. Accessed 18 November 2015.

Spindler, Robert P.

    2007   Digital Preservation. *Preservation Leaflets*. Northeast Document Conservation Center, Lowell, MA. https://www.nedcc.org/free-resources/preservation-leaflets/6.-reformatting/6.5-digital-preservation. Accessed 20 November 2015.