Data Management Plans for Archaeological Research – 2017

Introduction

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 to 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 (ADS and Digital Antiquity 2013; Richards et al 2010; McManamon 2014; Kintigh et al. 2015).

Careful curation of digital files is important because most data and information created by contemporary research in many 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, 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 (Center for Digital Antiquity 2015).

For these reasons, agencies, foundations, and other funders of archaeological research, commonly require a data management plan (DMP) as part of proposals seeking funding (e.g., National Science Foundation 2010). The DMP template provided in this guide will help you complete your data management plan. For example, you

can use the template to address the requirement for such plans for NEH and NSF proposals, or as part of a CRM or public archaeology proposal to address requirements to provide for the curation of digital data generated by the project.

A Data Management Plan Template for Projects Using tDAR

Below is a template that describes how data that are proposed to be collected or created for particular projects will be curated when placed in the tDAR repository. The estimated number of files and overall file size should be filled in when the template is used as part of a proposal for a grants program or in responding to Requests for Proposals.

Data created by this project will be deposited for long-term access and preservation in the tDAR digital repository. We estimate up to xxx files and xx GB of data being added to tDAR as part of the project. Within tDAR we will organize the data (documents, data sets, images, and other types of digital files) as projects and collections with sub-collections, as appropriate. These organizational tools will enable both project researchers and outside investigators to easily access and report on their digital research products.

Project datasets and records will be uploaded and thoroughly documented with technical and semantic metadata using tDAR's interactive Web forms. Standard metadata categories used by tDAR include numerous cultural, geographic, temporal, material, bibliographic and other archaeological fields. Dataset columns are individually documented in terms of their type and content; nominal fields are documented value by value.

In tDAR all digital files are stored in the original submission format and, as needed, in a preservation format to maintain stability though long-term migration. Preservation formats are determined in consultation with the Library of Congress and the Archaeology Data Service procedures. Files may also be copied into a dissemination format to facilitate usability. tDAR's digital objects are thoroughly documented by administrative and technical metadata for preservation, descriptive metadata for effective resource discovery, and detailed semantic metadata needed to permit sensible scientific reuse of the data. Web forms guide data contributors through a comprehensive process of metadata entry and file upload.

tDAR's descriptive metadata includes general and bibliographic components incorporating Dublin Core and Library of Congress's MODS metadata standards. It also includes fields covering information unique to archaeology, such as: site types, investigation types, and detailed, column-by-column, and table-by-table information documenting datasets. tDAR's administrative and technical metadata utilize components of the Library of Congress's PREMIS metadata standard for capturing, technical, preservation, and rights information.

Access to tDAR content is provided through a Web interface with basic and advanced (including spatial) search capabilities. tDAR content is indexed by Google and other major search engines. Excepting legally protected confidential data and data temporarily embargoed by their contributors, all tDAR data are freely available over the web. Files deposited in tDAR are all

accompanied with DOIs and persistent URLs. Data may be reused, redistributed, or transformed, subject only to the provision of appropriate credit to the data creators and indication of any changes made (Creative Commons Attribution 3.0 Unported License). All data downloads include appropriate citation information.

tDAR is served by Linux virtual machines operating within Arizona State University's Server-On-Demand facility. Each virtual machine has fully resilient data storage on a redundant disk behind enterprise-class Network Attached Storage filers. Daily backups are maintained in a separate building to provide protection and rapid restore from any data loss and are retained for two-weeks. Regular offsite (bi-weekly, and quarterly) backups are also performed. Additional secure, encrypted offsite backups of tDAR use Amazon Web Services' Glacier Cloud Storage.

Should Digital Antiquity at any time cease to exist, the ASU University Library has agreed to continue to provide access to tDAR data. tDAR operates under the organizational umbrella of the Center for Digital Antiquity, a multi-institutional organization (based at ASU) designed to ensure the long-term financial, technical, and social sustainability of tDAR. It is governed by an independent Board of Directors and supported by an external Professional Advisory Panel. Start-up funding for tDAR has been provided by the National Science Foundation (0433959 and 0624341) and the Andrew W. Mellon Foundation.

Additional guidance on good digital curation practices and data management can be found in Inter-university Consortium for Political and Social Research (2012).

References

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 Digital Antiquity
2015   *How to Ensure Proper Curation of Digital Archaeological Records, Center for Digital Antiquity – 2015*. Center for Digital Antiquity, School of Human Evolution and Social Change, Arizona State University, Tempe. https://www.digitalantiquity.org/wp-uploads/2011/07/How-to-Ensure-Proper-Curation-of-Digital-Archaeological-Records1.pdf. Accessed 23 March 2017.

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.

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.

National Science Foundation

2010   *Data Management for NSF SBE Directorate Proposals and Awards*. National Science Foundation. https://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf. Accessed 24 March 2017.

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.