

Digital Data Curation and Access:

Why You and Your Organization Should be Actively Involved

An ACRA White Paper -- 2019

**Duane Peter, Keith Kintigh, Francis Pierce-McManamon,
and Leigh Anne Ellison**

Federal requirements for appropriate management of archaeological data are set forth in the National Historic Preservation Act (NHPA), the Archaeological Resources Protection Act (ARPA), regulations regarding curation of data associated with those statutes (36 C.F.R. 79), and regulations promulgated by the National Archives and Records Administration (36 C.F.R. 1220.1-1220.20) that apply to all federal agencies. The National Archives regulations require that “[a]gencies must create and maintain authentic, reliable, and usable records and ensure that they remain so for the length of their authorized retention period” (36 C.F.R. § 1220.32). In the case of associated archaeological records, that retention period is set by the NHPA, which mandates that the associated records be permanently maintained (16 U.S.C. § 470h-4(a)(2)).

Federal law also imposes an affirmative duty upon the heads of federal agencies to establish safeguards against the destruction of digital archaeological records not otherwise scheduled for destruction. As machine readable materials, digital archaeological records meet Section 3301 of 44 U.S.C. Chapter 33’s definition of “records”:

“records’ includes all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of data in them.”

Most are aware that the curation of archaeological and historical data and associated records has been a challenge for quite some time. Adequate curation of objects and associated records has received significant attention in the past two decades; however, professional archaeologists within academia, the Cultural Resources Management (CRM) industry, and federal and state agencies are keenly aware that we need to do more, if we are to successfully preserve our heritage, advance evidence-based understandings of the human past, and fulfill an obligation to serve as stewards of the archaeological record of human history (e.g., [Society for American Archaeology 1996](#)).

Archaeologists increasingly recognize that addressing our most pressing social and intellectual problems (Kintigh et al. 2014) demands comparative and synthetic research ([Altschul et al. 2018](#)). Archaeology uniquely provides access to the long term social and

environmental dynamics that must be understood to address important contemporary problems. Archaeology's long-term perspective is essential because key processes operate so slowly that they are only perceptible over centuries or millennia. An archaeological perspective illuminates and enables the synthesis of information concerning the initial conditions and the outcomes of thousands of completed social "experiments" in the past.

The needed synthetic research critically depends upon access to the vast corpus of archaeological, ethnographic, and historical data that the CRM industry has generated over the last 50 years. If we are to leverage those data to advance our knowledge of the past and for the benefit of the public, the data *must* reside in repositories where they are easily discovered, accessed, used, and preserved for future uses.

"Preserved for future uses" is the critical concept, for most current curation facilities are not equipped to migrate the data to new media as digital technology changes. Responsible digital data repositories that will enable synthetic research well into the future must exhibit these key characteristics:

1. Deposited data are broadly discoverable using commonly available Internet search engines, as well as through the repository website.
2. Deposited data are easily accessible online, but controlled access for confidential information (e.g., precise site locations) also is possible.
3. Deposited files are linked permanently to rich, disciplinary-appropriate administrative, descriptive, and technical metadata, which enable their re-use.
4. The repository operates under a sustainable financial and/or contingency plan that ensures long-term access and preservation of its content.
5. Deposited files are regularly checked and their integrity maintained.
6. The repository has an appropriate plan to provide long-term access and preservation as hardware and software change.
7. The repository maintains robust backups of repository content and software, including full off-site backup.

A review of curation facilities, both traditional and digital data repositories, reveals that only one digital data repository in the United States, The Digital Archaeological Record (tDAR), presently exists that meets the required characteristics for long-term digital access and preservation. tDAR was established 10 years ago and was designed specifically for this purpose ([McManamon et al. 2017](#)). The Mellon Foundation and the National Science Foundation have invested millions of dollars in tDAR to develop its powerful discovery and access capabilities, and its tools to facilitate synthesis and enhance the usability of the data ([Kintigh et al. 2018](#)). tDAR provides free access to its content through an easy-to use interface. That same interface allows individuals anywhere to upload datasets, documents, images, and GIS files and to richly document them so the data can be effectively used *and properly credited*. It allows data contributors to fully control access to sensitive content (primarily site locations) and to upload redacted content for public use. By default, tDAR obscures site locations in publicly accessible metadata. tDAR collects modest one-time

deposit fees (typically \$5/10MB file) to support the file's permanent preservation and access.

In addition to supporting the kinds of synthetic research described above, depositing data, documents, and images in tDAR enables their easy access and use in the day-to-day background and comparative research needed for CRM reports and proposals, as well as for academic papers. It also provides resource managers with ready access to information that is critical to their work.

Unfortunately, however, only a tiny fraction of the data archaeologists has generated over the last 50 years has been deposited in tDAR. This is despite the fact that a legal analysis conducted by Cultural Heritage Partners, PLLC, a DC-based law firm ([Cultural Heritage Partners 2012](#)) demonstrates that US laws and regulations regarding archaeological data generated as part of public undertakings *require* that these data be publicly available and made useable for legitimate purposes.

While tDAR has eliminated the technological obstacles to data sharing that we once faced (Kintigh 2006), today the key problems are (1) to transform archaeological practice so that normal investigation and research workflows provide for the consistent deposit of data from new and ongoing projects into tDAR or similar digital repositories, and (2) to develop ways to make available in tDAR the huge corpus of archaeological legacy data (un- or under published data).

What Can You, Your Academy, Your Firm, or Your Agency Do?

- ❖ As a responsible professional, each of us can work to ensure that all research is reported and made easily and widely available. Each of us should encourage our organizations or firms to ensure that all data are available for future research efforts.
- ❖ Archaeologists in academic institutions should train students about the importance of effectively curating the results of their research, best data practices related to long-term curation, and making their data easily and widely available.
- ❖ Archaeologists conducting research in academic institutions should deposit in tDAR data they generate from their research so that the data can be used in future research and resource management by other archaeologists in CRM, the academy, and government.
- ❖ CRM firms should deposit in tDAR (whenever possible) data they generate so they can be used in research and resource management by archaeologists in CRM, the academy, and government. By doing this, the firms will highlight the quality of their work in the competitive environment and their commitment to enhancing the sustained value of the archaeological record.

- ❖ CRM firms should deposit in tDAR archaeological data from legacy projects they have conducted and to work with project sponsors to obtain any necessary permission to do so.
- ❖ CRM firms should work with private clients to ensure that, insofar as possible, the archaeological data they collect are made available for appropriate use (e.g., in tDAR), consistent with our responsibilities as stewards of the archaeological record.
- ❖ Federal and state agencies at all levels should fulfill their legal responsibilities to make digital archaeological data (broadly construed as datasets, documents, images, etc.) generated as part of public undertakings easily discoverable, accessible, usable, **and preserved for future use.**
 - For new projects, contractual documents such as permits and RFPs must require the deposit of data in a digital repository that provides the necessary preservation and access.
 - As a condition of signing off on project completion, agency processes should ensure the proper completion of the data deposit.
 - For legacy projects, agencies should develop systematic plans to move their legacy data into a digital repository and should fund their incremental completion.

What Are the Benefits?

Having digital data from thousands of projects in one or a few institutions that will provide access for the future will provide the following efficiencies:

- ❖ Reduction of time and travel in conducting background research,
- ❖ Better knowledge of previously collected data will reduce the collection of redundant data,
- ❖ Reduction of some of the need for costly mitigation; and
- ❖ Enhanced understanding of the social and environmental dynamics that have contributed to the present human condition.

The achievement of such efficiencies and the use of those data in ways that the public sees as benefits will satisfy the goals envisioned by the creators of the NHPA, ARPA, and associated regulations.

References Cited

Altschul, Jeffrey H., Keith W. Kintigh, Terry H. Klein, William H. Doelle, Kelley A. Hays-Gilpin, Sarah A. Herr, Timothy A. Kohler, Barbara J. Mills, Lindsay M. Montgomery, Margaret C. Nelson, Scott G. Ortman, John N. Parker, Matthew A. Peeples, and Jeremy A. Sabloff
 2018 Fostering Collaborative Synthetic Research in Archaeology. *Advances in Archaeological Practice* 6(1): 19-29. [DOI:10.1017/aap.2017.31](https://doi.org/10.1017/aap.2017.31)

Cultural Heritage Partners, PLLC (CHP)

- 2014 Legal Analysis of Federal Requirements for Curation of Digital Archaeological Documents and Data. Submitted to: Office of General Counsel, Arizona State University. <https://www.tdar.org/wp-content/uploads/2013/01/CHP-Legal-Analysis.pdf>

Kintigh, Keith W. (Ed.)

- 2006 The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71(3): 567-578.

Kintigh, Keith W., Jeffrey H. Altschul, Mary C. Beaudry, Robert D. Drennan, Ann P. Kinzig, Timothy A. Kohler, W. Fredrick Limp, Herbert D.G. Maschner, William K. Michener, Timothy R. Pauketat, Peter Peregrine, Jeremy A. Sabloff, Tony J. Wilkinson, Henry T. Wright, and Melinda A. Zeder

- 2014 Grand Challenges for Archaeology. *American Antiquity* 79(1): 5-24

Kintigh, Keith W., Katherine A. Spielmann, Adam Brin, K. Selçuk Candan, Tiffany Clark, and Matthew Peeples

- 2018 Data Integration in the Service of Synthetic Research. *Advances in Archaeological Practice*. 6(1): 30-41. [DOI:10.1017/aap.2017.33](https://doi.org/10.1017/aap.2017.33)

McManamon, Francis P., Keith W. Kintigh, Leigh Anne Ellison, and Adam Brin

- 2017 tDAR: A Cultural Heritage Archive for Twenty-First-Century Public Outreach, Research, and Resource Management. *Advances in Archaeological Practice* 1-12. DOI: [10.1017/aap.2017.18](https://doi.org/10.1017/aap.2017.18)

Society for American Archaeology

1996 Principles of Archaeological Ethics.

<http://www.saa.org/AbouttheSociety/PrinciplesofArchaeologicalEthics/tabid/203/Default.aspx>