



Reports in Digital Archaeology Number 4**March 2019**

The US Air Force CRM Program Meets the Challenges of Digital Data Curation: A Case Study Using tDAR

Francis P. McManamon, Rachel Fernandez, Leigh Anne Ellison, and Charlene Collazzi

Introduction

Many public agencies are responsible for managing large amounts of land, many historic structures, and a wide range of cultural and natural resources. The US Air Force (AF) is one such agency, with responsibilities including the stewardship of extensive archaeological and cultural heritage resources and the data and records related to them. Since the middle of the 20th century, these records and data have been created or collected as part of investigations carried out at AF installations. These investigations, mandated by laws and regulations, yield massive amounts of irreplaceable data, particularly on ancient Native American history and the lives of ethnic communities underrepresented in historical documents. These data are generated for the explicit purpose of benefiting the American public.

The AF data include paper and digital reports, correspondence, photos, maps, data sets, geospatial data, and other kinds of files in digital or paper formats. The paper versions of data occupy considerable space in offices and are difficult to locate, search, use, and share among staff, contractors, researchers, and the interested public. Unless the data and information are well-organized, findable, and accessible, simply locating and retrieving documents and data can be inefficient and require an investment of time from both AF staff and the person seeking the information. If accessing known documents is sufficiently difficult, a search will often be abandoned, and if the person doing the search does not know what documents exist and might be relevant, then the results of the prior work cannot be leveraged for ongoing decision-making or for other purposes.

Furthermore, since the reports and other paper records exist in limited numbers, maintaining a single copy of these records in one place raises risks for preservation and the potential for catastrophic loss, for example, if a storm destroys or inhibits access to records because the building holding them is damaged or destroyed. Increasingly, new projects include digital as well as paper records, but without an appropriate tool to manage the digital files, the media

McManamon, Francis P., Rachel Fernandez, Leigh Anne Ellison, and Charlene Collazzi. The US Air Force CRM Program Meets the Challenges of Digital Data Curation: A Case Study Using tDAR. *Reports in Digital Archaeology* Number 4, Center of Digital Antiquity, Arizona State University, Tempe, AZ, 2019.

containing the files may be simply stored on a shelf or printed, inhibiting data discovery, accessibility, and use, thereby adding to data management problems.

Meeting the Digital Data Management and Curation Challenges with tDAR

In 2010, AF cultural resources subject matter experts, Dr. James Wilde and Dr. Paul Green (now retired), began to consider whether and how use of tDAR (the Digital Archaeological Record; www.tdar.org; McManamon et al. 2010; 2017) could help the AF meet its responsibilities for data management and sharing. These responsibilities are required by the National Historic Preservation Act, the Archaeological Resource Protection Act, and the regulations of the National Archives and Records Administration (Cultural Heritage Partners 2012).

The Digital Archaeological Record (tDAR) is a digital archive and repository designed to curate archaeological and cultural heritage documents, images, datasets, and other digital files. Users of tDAR can search for these materials from archaeological and cultural heritage projects throughout the world. Users who want to preserve and share their data can easily upload a wide variety of digital file types in tDAR. Curating data in tDAR offers broader and easier access to data and sharing for future uses, such as resource analysis and evaluation for decision-making, background studies, public interpretation and outreach, project management, and research. In tDAR, the documents and other data are linked to metadata, which allows the information to be searchable by time period, culture, study type, location, and a variety of other kinds of information. Metadata and files in tDAR also are interactive. Records are updated easily and contacts for each file can be easily reached via the tDAR website's native email application for further information. After careful investigation, AF experts determined the ready access, preservation, and ease of use provided by tDAR perfectly met their needs.

In 2011, the AF Civil Engineering Center (AFCEC) and cultural resource management (CRM) program partnered with Geomarine Associates (Duane Peter and Michelle Wurtz-Penton, project managers) and the Arizona State University (ASU) Center for Digital Antiquity (www.digitalantiquity.org; Digital Antiquity), which develops and manages the tDAR repository, to develop methods and procedures that would address the access, management, and preservation of digital data at AF installations nationwide. CRM staffs at three AF bases (Dyess AFB in Texas, Shaw AFB in South Carolina, and Avon Park AF Range in Florida) agreed to host pilot projects that would test whether and how utilizing tDAR could help the AF meet its data management needs. At the beginning of the project, the bases involved did not possess a systematic method to sufficiently curate information from past archaeological investigations or the cultural resources on their bases. For example, in some instances the records, documentation, and data were scattered among several locations and in differing formats. In addition to this, much of the documentation was still curated only as paper copies, physical media, or in a non-preservation-standard software. Proper curation of data and records is essential to maintaining a link among the projects, physical artifact collections, including associated paper documents, and other efforts documentation. The digital archiving done in the pilot projects followed the guidelines and steps for digital data curation described by the

Digital Data Centre (DCC 2019) and the Archaeology Data Service and Digital Antiquity (ADS & Digital Antiquity 2013, 2019).

The pilot project created a digital archive in tDAR for the AF archaeological and cultural heritage data with sub-collections for each of the three bases, which provide organization, reliable access, and sustainable long-term preservation for their data (<https://core.tdar.org/collection/16304/us-air-force-archaeology-and-cultural-resources-archive>). Since the initial work was completed, over 20 additional installations, some of them parts of Joint Bases, have partnered with Digital Antiquity to create digital archives. **Figure 1** schematically displays the distribution and quantities of the AF digital records in tDAR across the US.

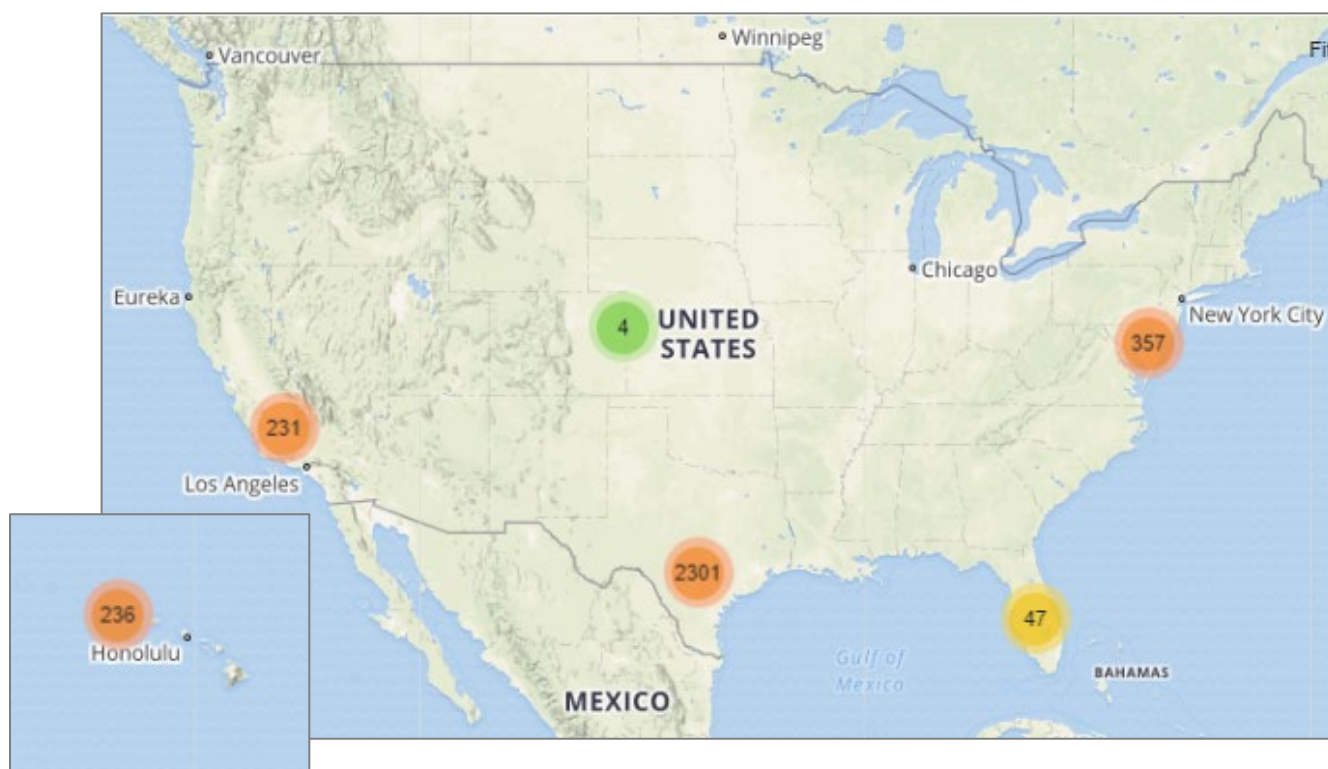


Figure 1. Schematic Map Showing the General Spatial Distribution and Numbers of Documents and other Digital Files for the AF Digital Archive in tDAR as of 2018

The methods, techniques, and activities for the digital data management and curation program were developed and tested in the pilot phase of the program. Developments included the creation of a framework for the AF digital archive in tDAR and a standard set of procedures for establishing digital archives at other AF facilities in need of similar data management and digital curation tools. The procedures that were effective during the pilot phase have been refined by subsequent work with additional AF installation and AFCEC CRM experts.

Procedures for Working with AF Installations

At the beginning of any new AF digital data management and curation project, an essential first step is consultation between Digital Antiquity curators and installation CRM staff to identify current and desired uses of the data. These meetings should lead to a number of initial activities:

- Identify and create a list of the projects, reports, other documents, maps and other images, and other materials that might be included in the data repository.
- Identify the documents, images, maps or other material that are not yet in digital formats and identify for possible scanning to create digital versions of these.
- Collaborate with Digital Antiquity staff to establish the specific administrative and substantive metadata categories, as well as any specific terms that installation and other AF staff want to include in the metadata drafted for the installation digital archive materials.
- Discuss and address any other topics or challenges related to the specific situation at the installation and the digital archive being planned. For example, what, if any, files or portions of files containing information, such as specific site locations, should be considered confidential or sensitive. Stricter access controls can be applied for these files once they are curated in tDAR.

In order to identify the archaeological and cultural heritage data and information related to the installation, a variety of activities are initiated. These activities include: checking the library and files at the installation, viewing files and collections related to the base at the state archaeological repository, reviewing bibliographies and “References Cited” sections of reports to identify documents that should be included in the digital archive, checking site files and reports archived by the SHPO, etc. From these activities, a list is created of the known materials to be included in the digital archive. When reports or other documents or data are available only on paper, scanning is needed to create digital files from the physical copies. The scanning effort may be substantial for bases that have existed for a long time. For installations with large numbers of archaeological and cultural heritage paper documents, images, or maps, a professional document scanning company may be suggested to digitize the backlog of physical files. Such companies, under specific scopes of work that reflect the needs at the installation, can use high-volume scanning equipment and are able to quickly, accurately, and securely digitize documents, images, maps, and other data. Such an arrangement may be preferable to saddling digital curators and AF staff with the labor-intensive and time-consuming task of scanning roomfuls of documents on equipment not designed to meet high-volume demands. With carefully composed scopes of work for the scanning task, utilizing these professional scanning services efficiently delivers the digital data to Digital Antiquity curators. This allows Digital Antiquity and installation CRM staff to focus their efforts solely on curation, expediting the project overall.

Once the documents and other types of data are in appropriate digital formats, these files are sent to Digital Antiquity. Digital Antiquity curators review the files to ensure the paper records have been scanned fully and that the digital files are an accurate copy, that the scanned files

are appropriately formatted, and that the scanned files will be technically accessible. They pass all files through an optical character reader (OCR) program so that information contained in these documents can be easily searched for once uploaded to tDAR. Curators also review the texts and illustrations to identify potentially sensitive information that might need to be redacted or designated as “confidential,” such as specific locations for sites vulnerable to disturbance or looting. When encountered, curators review this questionable material with the installation CRM staff, and then use Adobe Acrobat Pro’s redaction tools to remove confidential information as needed. In most cases, the redacted version of the document is uploaded and made available to registered tDAR users, but the full report, also uploaded, is marked as confidential and access to it is restricted to registered tDAR users specifically authorized by AF.

At Digital Antiquity, metadata records for each file or groups of similar files are drafted by digital curators. Metadata is a term used for

...descriptive information about documents, data sets, images, and other digital objects. Providing metadata enables discovery, accessibility, and usability of files in a digital archive by providing general and standardized fields and terms to describe files and the information they contain. Furthermore, metadata provides context for non-textual materials, such as data sets or images that may not be usable without this additional information. (ADS & Digital Antiquity 2013:46)

For example, metadata for a document includes the document title, date of creation or publication, author(s) name(s), a summary or abstract of the document content, and other information. For a spreadsheet of data, the metadata should include descriptions of the data represented in the rows and columns of the spreadsheet, descriptive keys to any codes used, descriptive information about how any measurements in the data set were made, and other information that would be needed by other researchers using the spreadsheet data (48-52).

Metadata for files deposited and curated in tDAR are organized into a number of categories, such as Title, Year Created, Description, Author/Creator, Investigation Types, Material Types, Cultural Terms, and more (see a complete list at: <https://tdar-arch.atlassian.net/wiki/spaces/TDAR/pages/557072/Data+Dictionary>). For the AF program, Digital Antiquity collaborated with AF staff to create a “Metadata Style Guide” for each installation or project. This ensured that the digital curators selected consistent terms for each project when drafting the metadata records, eliminating potential confusion for instances when more than one category term may be appropriate.

Upon completion of the metadata drafts, installation CRM staff and, if appropriate, AFCEC CRM experts are able to confidentially review the draft metadata records and uploaded files in the installation collection in tDAR. Through subsequent consultation with the Digital Antiquity curators, any necessary edits or changes to the files are made. When the final review is complete, Digital Antiquity curators are then ready to make these records active in tDAR. As the records’ status is changed from “draft” to “active,” the metadata pages become visible online and any files not marked as “confidential” are available for viewing or download by

registered tDAR users. Once uploaded to the repository, the files are stored, backed-up securely and systematically, curated, and preserved for long-term discoverability, accessibility, and use. Participation in the archiving process by installation CRM staff is essential to ensure that projects are planned, appropriately conducted, and result in a tool that assists the base CRM staff in management, research, public outreach, and resource protection activities for which they are responsible. A general description of the procedure is presented in **Figure 2**. The circled boxes indicate activities in which the installation CRM staff are most directly involved.



Figure 2. Generalized Diagram of the AF Digital Curation Procedures and Workflow for Legacy Data Projects

Managing AF Archaeological and Cultural Heritage Data and Information

Encouraged by the success of the pilot projects, the AF CRM program at the AFCEC and CRM installation staffs continued to partner with Digital Antiquity to create and implement digital data management programs at additional bases in 2015, 2017, and 2018. At present, over twenty additional digital archives have been established at AF bases and facilities. Four more digital archive projects are underway in 2019. **Table 1** lists these bases, joint bases, and installations. The table also summarizes the thousands of records that are included in their collections so far.

The US Air Force CRM Program Meets the Challenges of Digital Data Curation: A Case Study Using tDAR

USAF Base/Facility	State	Date Created in tDAR	Records			Types of tDAR Records						Notes	
			Draft	Active	Total	Documents			Images	Data sets	Projects		Total
						Citation	Full Text	Total					
Avon Park AF Range	FL	2012	-	43	43	-	32	32	-	1	5	43	
Dyess AFB	TX	2012	-	17	17	1	14	15	1	-	1	17	
Shaw AFB	SC	2012	-	81	81	-	58	58	9	1	13	81	
Multi-base/General Resources	NA	2014	1	10	11	-	11	11	-	-	-	11	
Bellows AF Station	HI	2015	148	-	148	50	97	147	-	-	1	148	
Hickam AFB	HI	2015	91	46	137	47	86	133	2	-	2	137	Includes: Hickam AFB and Wake Island
Kaena Pt Satellite Tracking Station	HI	2015	8	2	10	2	8	10	-	-	-	10	
Koke'e AF Station	HI	2015	1	1	2	1	1	2	-	-	-	2	
JB San Antonio	TX	2015	2,369	-	2,369	16	1,746	1,762	505	36	66	2,369	Includes: Fort Sam Houston; Camp Bullis; Lackland AFB; Randolph AFB
Travis AFB	CA	2015	19	6	25	6	17	23	-	-	2	25	
Vandenberg AFB	CA	2015	-	148	148	11	134	145	-	-	3	148	
Hanscom AFB	MA	2017	-	23	23	-	22	22	-	1	-	23	
JB McGuire-Dix-Lakehurst	NJ	2017	-	231	231	-	228	228	-	2	1	231	Includes: McGuire AFB; Fort Dix; Lakehurst AFS
New Boston AFB	NH	2017	-	7	7	-	6	6	-	1	-	7	
Niagara Falls AFB	NY	2017	-	3	3	-	3	3	-	-	-	3	
Rome AFB	NY	2017	-	9	9	-	9	9	-	-	-	9	
Westover AFB	MA	2017	-	11	11	-	11	11	-	-	-	11	
Los Angeles AFB	CA	est 2019	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	
Nellis AFB	NV	est 2019	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	
JB Charleston	SC	est 2019	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	
Robins AFB	GA	est 2019	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd	
TOTAL			2,637	638	3,275	134	2,483	2,617	517	42	94	3,275	

Table 1. List of AF installations' collections in tDAR (March 2019)

In tDAR, the metadata records and digital files are organized in collections and sub-collections that can be configured in a variety of ways. This flexibility enables AF CRM staffs to create an organizational format they find most useful and easiest to manage based on their individual needs. Features in tDAR also empower collection managers to rearrange the structure of a collection and to create alternative organizational structures, depending upon their management needs. The intent is to ensure that the digital materials are curated to meet the requirements of AF cultural resource staff to care for these resources, as well as to comply with federal laws and regulations governing preservation and management of archaeological and cultural heritage data, such as 36 CFR 79 (Cultural Heritage Partners 2012).

CRM installation staffs have found their tDAR collections useful in a variety of contexts. At Joint Base San Antonio, CR manager Arlan Kalina has provided contractors working on historical architecture resources access to reports, other documents, and images in selected sub-collections of the JBSA tDAR collection for their background research and comparative studies of historic structures at the installation. For CR program leader Kathy Couturier (2017) at Avon Park Air Force Range (APAFR), "...tDAR is a safe place to store digital documents, images, and other data outside of [less flexible]...IT systems." She also notes that for "...new contracts at APAFR we have to send [the contractors] our past survey work (34 years) which is impossible using [another] system. Giving a contractor access to our records, all in [tDAR], is convenient, safe and easy for both APAFR and the CRM firm doing the investigation for us parties." Dr. Adrienne Velasquez, CR manager for AF installations in the northeastern United States, recently completed a large project to curate tDAR hundreds of documents and other files from the bases in that region. She now is looking forward to being able to easily share with the

installation CR officials this historical and legacy data and information in ways not possible before.

In addition to serving a variety of data management functions, collections in tDAR also provide a back-up system for local or other agency digital archives. If these other archives are wiped out for whatever reason, the records and files in tDAR are available for a fresh start with very little effort. The appropriate sharing of some data and information also is possible using the tDAR collections. The tDAR metadata records are available publicly through the repository website. Access to the files themselves is possible for individuals who register at tDAR and agree to the terms of use of the data and information (<https://www.tdar.org/about/policies/terms-of-use/>). As mentioned above, files also can be marked confidential when deposited. In such cases, users who find the metadata record and seek to have access to the files may request, using the tDAR email feature, access from the individual or organization with stewardship responsibility for the data. For the AF data in tDAR typically this means CR officials at the installation and/or in a regional role and/or at the AFCEC.

To help measure the broader effect of these efforts, tDAR usage statistics are available for the AF overall collection. These show that in 2017 there were over 38,000 page views of metadata records and 924 downloads of files for the AF collection. In 2018, there were over 28,000 views and 1,283 file downloads. Our usage data does not retain personal information about the page view or download of specific tDAR registered users. Certainly, some are members of the general public. Other users are academic, CRM, and public agency archaeologists and cultural heritage experts. Because the AF is curating these data in tDAR, this wide range of individuals are able to discover, access, and use information of which they otherwise would be unaware. Curating the results of the research at AF installations makes available decades-worth of archaeological data and information for cultural, historical, social, scientific, and other uses. **Figure 3** summarizes the numbers of user page views and downloads of AF's tDAR collection since its creation in 2011.

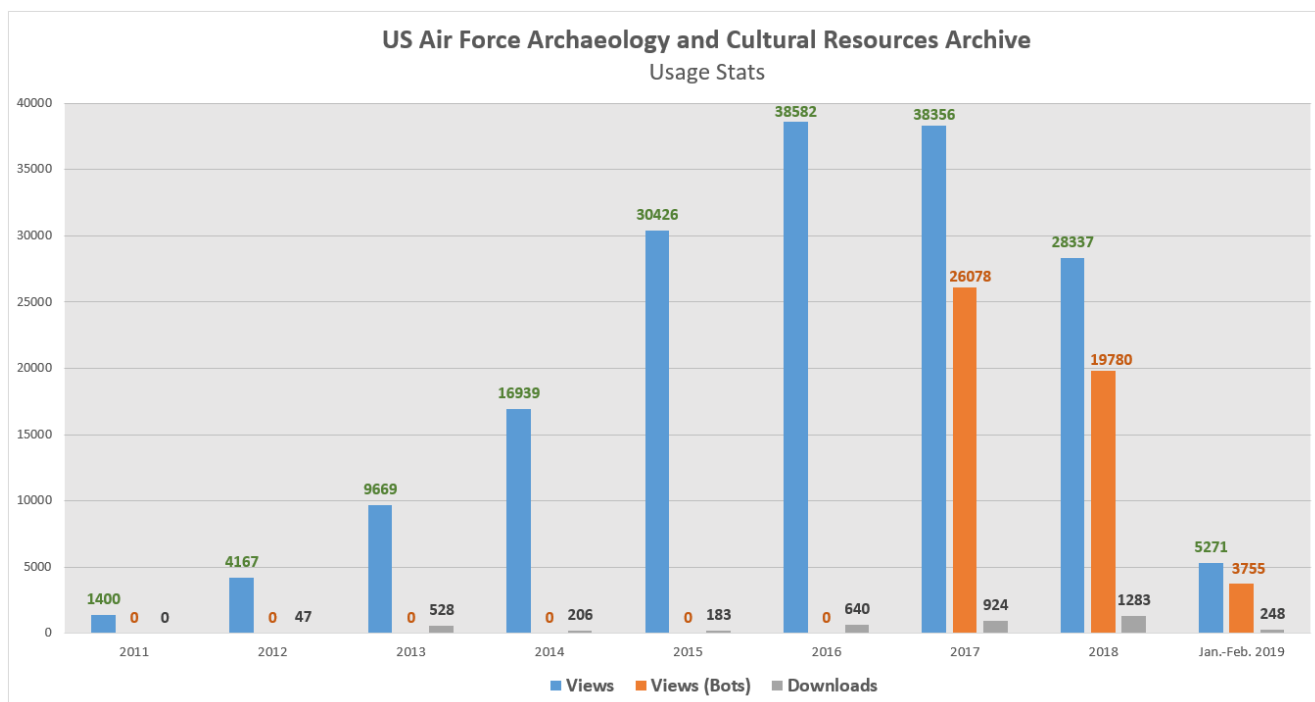


Figure 3: Page views and downloads for metadata records and files in AF tDAR Collection

Note: We started actively tracking search engines and bots accessing tDAR in April 2017. Totals include bot and search engine traffic to show consistency in numbers.

By utilizing tDAR, the AF contributes to data sharing among experts, as well as general outreach to the public about the archaeological and other cultural resources and data for which it is responsible. The AF also supports outside research that may be helpful for public interpretation, resource protection, or other management functions in the future. By making these data easily available, with appropriate access controls, the AF is meeting the Department of Defense policy of sharing information with the public.

Managing Data and Information from New Investigations

One important data management feature that the AF CRM program instituted in 2016 is the requirement that digital curation of documents and data created by current and future cultural heritage investigations and other research to be included within every project’s scope. Best practice in digital curation (e.g., ADS & Digital Antiquity 2013:9-22, 44-52; Michener et al. 1997:331) calls for the creation of metadata records and uploading of digital data, coordinated with appropriate review and approval by AF CRM experts, to a responsible domain digital data repository as one of the outcomes for all archaeological and cultural heritage investigations. Including the digital curation of project data as a built-in component of project funding and scheduling is a quick and easy way to ensure that the data and information from studies are added to a repository with no delay and no addition to a backlog.

Continuing use of tDAR to meet digital data management challenges and digital curation requirements positions the AF to comply with the existing United States statutes and

regulations regarding archaeological resources and historical property (Cultural Heritage Partners 2012). The AF digital curation program also complies with broad data sharing requirements of the recently enacted 2019 OPEN Government Data Act (<https://www.congress.gov/bill/115th-congress/house-bill/1770> and <https://www.datacoalition.org/open-government-data-act/>). Use of tDAR provides the AF CRM program with a means of avoiding the loss of legacy and new digital data, a current problem faced by public agencies and other organizations conducting publicly funded research (Kintigh 2018; Witze 2019).

This approach to the immediate curation of digital data created by new investigations is being instituted and proven effective by other federal agencies, as well. For example, the Phoenix Area Office and Lower Colorado River offices of the Bureau of Reclamation (Digital Antiquity 2013) have instituted such requirements in current and new contracts with CRM firms doing work for them. The terms of new contracts include a digital curation stipulation and guidelines for creating digital records and metadata for deposit into tDAR as part of the execution of the studies. As a result, future work carried out on the lands and facilities managed by the agency will be deposited automatically to tDAR, organized in a way that is immediately useful for agency internal functions, and preserved for long-term access and use by Reclamation staff, contractors, researchers, and the interested public.

There are other examples from the Bureau of Land Management and the Delaware Department of Transportation (McManamon 2018). Metadata creation by the same individuals who produce the digital records, in this case the CRM firms conducting the studies, is an effective way to produce accurate and detailed metadata for their tDAR records. This work builds the content of the agencies' tDAR archives and ensures that agency staffs' time and funds are spent on more critical tasks. By requiring contractors to enter the data, and placing this requirement in the contract scope, the AF would be able to reduce staff time and eliminate backlog growth by ensuring that new data is properly curated from the start.

References Cited

ADS & Digital Antiquity (Archaeology Data Service and the Center for Digital Antiquity)
2013 *Caring for Digital Data in Archaeology: A Guide to Good Practice*. Oxbow Books, Oxford and Oakville, California.

2019 *Guides to Good Practice in Digital Archiving for Archaeology*. A series of webpages by the Archeology Data Service and the Center for Digital Antiquity containing detailed guidance about creating, managing, and treating digital files in archaeology at: <http://guides.archaeologydataservice.ac.uk/g2gpwiki/>. Accessed 27 February 2019.

Cofield, Sara Rivers and Jodi Reeves Eyre
2014 *Evaluating a Cooperative Approach to the Management of Digital Archaeological Data*. Department of Defense Legacy Resource Management Program Project No. 13-711.

Department of Defense Legacy Program, Washington, DC.

<https://core.tdar.org/document/393996/evaluating-a-cooperative-approach-to-the-management-of-digital-archaeological-records>, accessed 7 February 2019.

Couturier, Kathy

2017 Using tDAR for Information Sharing and Resource Management at Avon Park Air Force Range, Florida. tDAR Blogpost, 20 October 2017.

<https://www.tdar.org/news/2017/10/using-tdar-for-information-sharing-and-resource-management-at-avon-park-air-force-range-florida/>, accessed 29 January 2019.

Cultural Heritage Partners

2012 *Federal Laws and Regulations Requiring Curation of Digital Archaeological Documents and Data*. <https://www.digitalantiquity.org/wp-uploads/2013/05/2013-CHP-Legal-Analysis-of-Fed-Req-for-Curation-of-Dig-Arch-Docs-Data.pdf>, accessed 7 February 2019.

Digital Antiquity

2013 Case Study: Using tDAR to Manage Legacy and New Archaeological Documents and Data, Phoenix Area Office, Bureau of Reclamation. Electronic document,

<https://www.digitalantiquity.org/wp-uploads/2011/07/Using-tDAR-to-Manage-Legacy-and-New-Archaeological-Documents-and-Data-the-Phoenix-Area-Office-of-the-Bureau-of-Reclamation-.pdf>, accessed 30 November 2016.

DCC (Digital Curation Centre)

2019 What is digital curation? Webpage at: <http://www.dcc.ac.uk/digital-curation/what-digital-curation/>, accessed 25 February 2019.

Kintigh, Keith W.

2018 America's archaeology data keeps disappearing – even though the law says the government is supposed to preserve it. *The Conversation* 17 October 2018. Electronic document:

<http://theconversation.com/americas-archaeology-data-keeps-disappearing-even-though-the-law-says-the-government-is-supposed-to-preserve-it-104674>, accessed 28 February 2019.

McManamon, Francis P.

2018 Making Data Easier to Discover, Access, and Use as Part of Section 106 Streamlining. tDAR Blogpost, 8 October 2018. Electronic document:

<https://www.tdar.org/news/2018/10/making-data-easier-to-discover-access-and-use-as-part-of-section-106-streamlining/>, accessed 4 March 2019.

McManamon, Francis P., Keith W. Kintigh, and Adam Brin

2010 Digital Antiquity and the Digital Archaeological Record (tDAR): Broadening Access and Ensuring Long-Term Preservation for Digital Archaeological Data. *CSA Newsletter* XXIII (2). Center for the Study of Architecture. Electronic document,

<http://csanet.org/newsletter/fall10/nlf1002.html>, accessed December 8, 2016.

- 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* 5(3):238-249.
<https://repository.asu.edu/items/45144>, accessed 7 February 2019.
- Michener, W.K., J.W. Brunt, J.J. Helly, T.B. Kirchner, and S.G. Stafford
1997 Nongeospatial metadata for the ecological sciences. *Ecological Applications* 7: 330–342.
Available at: [http://onlinelibrary.wiley.com/doi/10.1890/1051-0761\(1997\)007%5B0330:NMFTE%5D2.0.CO;2/abstract](http://onlinelibrary.wiley.com/doi/10.1890/1051-0761(1997)007%5B0330:NMFTE%5D2.0.CO;2/abstract). Accessed 20 Nov 2017.
- Witze, Alexandra
2019 Disappearing Digital Data. *American Archaeology* Spring, 2019:40-45.