Member Node Description: The Digital Archaeological Record

Version 1.0   3/4/16   Adam Brin and Frank McManamon

General

Name of resource:	The Digital Archaeological Record (tDAR)
URL(s):	http://www.tdar.org
Institutional affiliation(s):	The Center for Digital Antiquity, School of Human Evolution and Social Change, Arizona State University
Primary geographic location:	Tempe, Arizona, USA
Project Director & contact info:	Francis P. McManamon (480) 965 1369 fpm@digitalantiquity.org
Technical Contact & contact info:	Adam Brin (480) 965 1278 abrin@digitalantiquity.org
Age of resource:	Since 2010
Funding support:	Grants, contracts with organizational clients, and individual user payment
Proposed Unique Identifier:	urn:node:TDAR

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):

tDAR is a repository for archaeological information spanning all seven continents and covering all periods of human prehistory and history. The repository was developed and is maintained by the Center for Digital Antiquity (Digital Antiquity) at Arizona State University (ASU). The repository includes information about a wide range of archaeological topics including specific investigations and resources, historical information, theory, method, and technique. Much of the information is from unpublished and "grey literature" sources—information collected as part of archaeological investigations that are not published in tradition venues, sometimes produced only for submission to public agencies or presented at conferences. In some cases, this information comprises the only detailed record of investigations of archaeological sites and excavations, research that is not possible to replicate or repeat.

Types of data (complex objects, text, image, video, audio, other):

tDAR contains a wide-range of data types including: documents, data sets, GIS data, images, and 3D scans or sensory data. We also maintain a number of "supporting" data types including projects, coding sheets, and ontologies.

Data and metadata availability (rights, licensing, restrictions):

All metadata is un-licensed, but effectively CC0. Files within tDAR are, by default CCBy 3.0, unless otherwise stated in the metadata. Users have the ability to restrict access to files within tDAR, but not the metadata.

Option for embargo (yes/no, duration):

Users have the ability to mark files as embargoed for a period of 6 months, 1 year, 2 years, or 5 years. Users can also mark files as confidential and perpetually restrict access. All metadata is always accessible, however. For archaeological data, protecting the site location (from potential looters or

vandals) is a critical issue. Thus, confidentiality of such data and other sensitive cultural information is an issue.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):
tDAR contains tens of thousands of documents, datasets, images, and other archaeologically significant materials.

Please describe recent usage statistics, if known, including information on annual data product downloads, annual number of users, annual number of data products used in publications:
We do not track number of data products used in publication. 14,565 files were downloaded in 2015.

User interactions

How does a user contribute data? (what can be deposited, how are data prepared, are specific software required, documentation/support available)
tDAR was designed to be self-service. Users visit the web site, register, and then complete an online form that allows them to specify administrative, technical, and descriptive metadata for the resource they are creating. As part of filling out the metadata form, a user can upload the appropriate file(s). Documentation is available both inline (built into the form) and on http://dev.tdar.org/confluence/. Users can deposit materials that meet our contributor's agreement (https://www.tdar.org/about/policies/contributors-agreement/), generally documents, data sets, images, and other materials related to archaeology, archaeological resources, and related topics.

How does a user acquire / access data?
tDAR provides a variety of methods to discover and access materials from basic and advanced search functionality to various methods for browsing the repository's contents. Once a user has identified materials, they can download materials once they've agreed to our terms of use.

For data sets stored within tDAR, users also can perform more advanced research using built-in data integration tools. These tools allow for synthetic research by allowing the merging of data sets with different collection standards and coding sheets into a new data set that combines their data and reconciles different data collection standards.

What user support services are available (both for depositing and accessing/using data)?
Digital Antiquity has data curators who are available to assist with the curation and discovery of data. We also have technical and other support staff who can assist in troubleshooting issues.

How does the resource curate data at the time of deposit?
tDAR is designed to attempt to validate all submissions at the point of deposit. The web-based forms validate and require appropriate metadata based on the type of resource, or options chosen by the end-user (e.g. files marked as confidential require a "contact", and data sets may require descriptions for columns). Separately, the application opens and validates each file upon upload to ensure that the data are readable and do not have any issues (e.g., equation errors in excel files).

Technical characteristics and policies

Software platform description, incl. data search and access API(s):
tDAR is a digital repository for archaeological data. It was designed and built by Digital Antiquity to solve the discovery and curation problems around archaeological data and to provide a platform for synthetic data integration. The repository software itself is written in Java and is built around a

Pairtree Filestore that stores all metadata and data files. The application is built using Spring, Struts 2, Solr, Hibernate, and aims to provide a self-service interface for archaeological data curation. The application is built around common repository and web-standard APIs including OAI-PMH, UnAPI, OpenSearch, CoINS, and others. We also maintain locally developed APIs for search, data ingest, and other common tasks.

APIs: https://dev.tdar.org/confluence/pages/viewpage.action?pageId=199786510

Service reliability (including recent uptime statistics, frequency of hardware refresh, if known):

Preservation reliability (including replication/backup, integrity checks, format migration, disaster planning):
Digital Antiquity aspires to have full OAIS compliance within tDAR in the next year. We maintain offsite backups and run weekly integration checks of the entire repository. Full copies of the tDAR application and data are maintained off-site in case of a disaster, and documentation is available to setup an instance of the application. Although, we are currently not performing format migration, we have limited the types of files we accept to those that can be migrated through methods we've already identified.

User authentication technology (incl. level of create/modify/delete access by users):

Data identifier system and data citation policy, if available:
tDAR uses internally minted IDs for all components of tDAR. Resources (documents, data sets, etc.) have a DOI minted through EZID when a file is associated with the resource, and it is made "public" within tDAR. All users must agree to a terms of use that required proper citation of any material they user (https://www.tdar.org/about/policies/terms-of-use/).

Metadata standards (including provenance):
The metadata model we use extends Dublin Core and MODS metadata standards to include detailed metadata necessary for discovery and curation of archaeological data. A data dictionary is available at: https://dev.tdar.org/confluence/display/TDAR/Data+Dictionary

Capacity/services to DataONE

At what functional tier will you initially be operating? (see http://bit.ly/MNFactSheet for definitions)
- ☒ **Tier 1: Read only, public content**
- ☐ **Tier 2: Read only with access control**
- ☐ **Tier 3: Read/write using client tools**
- ☐ **Tier 4: Able to operate as a replication target**

If you can host data from other member nodes, what storage capacity is available?

Can you provide computing capacity to the broader network? If so, please describe.

Other Services

What other services or resources (such as expertise, software development capacity, educational/training resources, or software tools) can be provided of benefit to the broader network?
Digital Antiquity has strong expertise in analyzing, describing, organizing, and preserving archaeological data. The tDAR software includes unique tools to help integrate qualitative data stored within the repository. We also have software development with strong affinities in server-side Java programming, and client side browser application development.