Sheehan, Beth. 2015. Comparing Digital Archaeological Repositories:
tDAR Versus Open Context, Behavioral & Social Sciences Librarian, 34:4, 173-213, DOI:
10.1080/01639269.2015.1096155

COMPARING DIGITAL ARCHAEOLOGICAL REPOSITORIES: tDAR VS. OPEN CONTEXT

BETH SHEEHAN
University of Illinois at Urbana-Champaign, Urbana, Illinois

Short Title: "Comparing Archaeology Repositories: tDAR vs. Open Context"
Keywords: archaeology data, online data repositories, The Digital Archaeological Record (tDAR), Open Context, data repository comparison

Abstract: In the last decade, two online archaeology data repositories, Digital Antiquity's The Digital Archeological Record (tDAR) and the Alexandria Archive Institute's (AAI) Open Context, have emerged in the United States as key players in the development of technology and web platforms for preservation and public online access to archaeological research data. The need for these services has intensified since 2011, in the wake of U.S. federal open access mandates and funding agency data management plan requirements for grant applications. Through a comparison of selected features of tDAR and Open Context, this paper highlights similarities and key differences in their designs, data management policies and practices, and provides guidance to subject specialist librarians and others who advise archaeological researchers about how these differences can impact the suitability of each repository for specific data characteristics.

Archaeologists have long recognized that the data they collect during fieldwork is an essential and often irreproducible record of the subject of their study, and that subsequent research related to a specific site and its context within future studies relies upon the availability and accuracy of this data. Most archaeological research cannot be repeated to verify findings, as the act of collecting the data often requires the systematic removal of layers of soil and artifacts, and the documentation of these findings (photographs, coded spreadsheets, drawings, maps, and field notes) become the representation of a research environment and research subject that can never again be examined in the same way. Since "the [archaeological] archive remains the primary record of what was there before" (Aitcheson 2009, 67), preservation of archaeological research data is in essence the preservation of the archaeological site itself, and is critical not only for preserving, publicizing and verifying an individual researcher's findings, but also to enable future research, and for responsible stewardship of cultural and historical heritage. As noted by Jacobs and Holland (2007, 198), the aim of archaeological research to "reconstruct cultures on the basis of physical remains" is often achieved through collaborative comparison of multiple projects. Increasingly, researchers must often rely upon the digital record of the data collected to make this comparison, in order to build upon or refute previous findings and conclusions. In addition, archaeology is one of many disciplines which have been affected by the United States federal government funding agency requirements for data management plans, and initiatives which encourage (and sometimes mandate) public sharing of the data produced by government-funded projects. Agencies such as the National Science Foundation (NSF), the Institute of Museum and Library Sciences (IMLS), the National Endowment for the Humanities (NEH) and the National Institutes of Health (NIH) have all added data management plan requirements to their grant application procedures (for more details about these requirements, see

the “DMP Requirements” page of DMPTool.org [University of California Curation Center 2015]). There are many possible reasons why an archaeologist may need to make choices regarding the management of their research data, but regardless of the motivation, ultimately there is much to gain from sharing and preserving archaeology data in the long term, and making it discoverable.

Subject specialist librarians who work with archaeology researchers should begin to consider their familiarity with the variety of online sources of primary archaeology data, including archaeology-specific data repositories, to be just as critical and valuable as their familiarity with relevant subject area research article databases. This is due to the evolving role of the liaison librarian in supporting emerging data management needs of researchers, as well as the nature of archaeological research data as a key primary information source. A 2009 report published by the Association of Research Libraries asserts that “the biggest part of the ‘last mile’ problem is to get faculty-created datasets into a permanent home so the data can be found and used again. Liaisons can do the project analysis and management required…based on their analysis, librarians can determine the best home for the data” (Gabridge 2009, 17). Background disciplinary knowledge and relationships with researchers are the strengths of subject librarians and position them well to aid researchers in this manner, but they must also know what the repository options are, and how to match data characteristics with these options. This, coupled with the fact that sharing of archaeological research data is not only beneficial to the discipline but increasingly mandated, means that knowledge of current archaeology data sharing options, and the unique attributes, advantages and restrictions of each, is necessary to properly advise archaeology researchers who need help selecting a repository that best meets the needs of their research data, and to help researchers find information and data about past excavations that are

needed for their future research plans. In addition to assisting with data management planning for primary research data deposit, archaeology data repositories are critical sources of the contextual and historical information needed to inform archaeology-related research questions. Since a large portion of the records of archaeological research consist of data and excavation documentation that is not formally published in academic journals and is part of the "grey" scholarly record (Aitchison 2009), repositories which provide access to archaeology data and documentation must be part of the archaeologist's (and therefore archaeology liaison librarian's) arsenal of disciplinary research resources.

In the time since archaeological researchers began experimenting with computerized data banks in the 1960's (Scholtz and Chenhall 1976), digital databases of archaeological data have become common, often individually developed for specific projects as a means of providing structure for documentation of the artifacts or other objects of analysis collected. Efforts to create web-based archaeological data repositories with the purpose of preserving this data and sharing it with others have been seen in the U.S. and U.K. since the early 1990's. In the guidelines document "Data Management for NSF SBE Directorate Proposals and Awards," (National Science Foundation 2010), the NSF Directorate of Social, Behavioral and Economic Sciences names two archaeological data repositories as examples of organizations that store specific types of data for a defined disciplinary community. The repositories named by NSF are Open Context (URL: http://opencontext.org/) and the Digital Archaeological Record (tDAR) (URL: http://www.tdar.org/), but they follow this with a statement emphasizing that "NSF does not endorse the use of any specific repository" (NSF 2010, 2). So how can archaeologists determine whether they should deposit their data in an online repository, and evaluate their options? What should they look for when choosing a repository, and why?

This article seeks to aid librarians and archaeologists by comparing and contrasting attributes of two online archaeology data repositories, Digital Antiquity's The Digital Archeological Record (tDAR) and the Alexandria Archive Institute's (AAI) Open Context, and describing situations and types of data which would be optimal or suboptimal for each based on the repository's unique characteristics. Characteristics which differ significantly between the two repositories are called out as the major deciding factors, which can then be compared to significant traits of individual archaeological project data to determine which repository is most optimal.

Why Compare Open Context and tDAR?

Open Context and tDAR were chosen for comparison in this article because they are both online archaeological data repositories with a similar scope that provide public online access to data. They both contain data from archaeological sites worldwide, and are actively accepting new submissions of archaeological data deposits from many different types of archaeology research projects and researchers.

There are other online archaeological repositories, but they are often more limited in scope and are therefore not suitable for direct comparison. For example, the United Kingdom's Archaeology Data Service (ADS) (URL: http://archaeologydataservice.ac.uk/) is a long-established repository which was founded in 1996 by a consortium consisting of the Council for British Archaeology and eight British universities, is still actively accepting new data submissions, and has received the Data Seal of Approval indicating that it is a "trustworthy repository" (University of York 2015a). However, the ADS is primarily focused on and gives priority to archaeological data from research sites located in the British Isles (ADS 2014), so although they state that they will accept data from other regions in which UK researchers have an

interest, the scope is a bit limited by the UK focus. Since Open Context and tDAR accept a broader range of archaeological data without giving priority to any region, and are the only established online archaeological data repositories the author is aware of that currently share all of these common characteristics at the time of this study, they are ideal for comparison and consideration for the majority of archaeological projects.

Additionally, Open Context and tDAR have also become increasingly more visible in the archaeological community in the United States, through presentations and marketing at conferences, recognition by professional associations, and articles published in archaeological journals and newsletters. Since 2012, several posts on the "News" section of the tDAR website (Digital Antiquity 2015c) and the Open Context *Heritage Bytes* blog (AAI 2015d) indicate that representatives from each organization have given presentations about their repository at lectures, lightning talks, forums and exhibit booths at the Society for American Archaeology (SAA) Annual Conferences, and at other organizational conferences such as the Computer Applications and Quantitative Methods in Archaeology Conference and the Digital Humanities Conference. In addition, tDAR has been recommended by members of the American Anthropological Association (AAA). In September 2012, the chair of the Grey Literature subcommittee of the Resource Development Committee of AAA published a statement in *Anthropology News* encouraging archaeologists to deposit their data in tDAR (Hahn 2012). Open Context has also received national recognition; developer Eric Kansa received commendations at the White House's "Champions of Change for Open Science" in June 2013 (Bobley 2013).

In scholarly literature, existing articles about tDAR or Open Context that have been published in journals or newsletters generally describe the research and design work that the repository developers have done to create the repositories, and their approach to solving the

challenges of archaeological data management by creating systems which will address these obstacles. Such articles generally give details and information primarily about one of the repositories, and are usually written by people who have been directly involved in the development of that repository, and who may have an ongoing investment and personal stake in the success of the repository. For example, articles describing the development and features of Open Context have been published in *Near Eastern Archaeology* (S. Kansa, E. Kansa, and Schultz 2007), *Educational Technology* (S. Kansa and E. Kansa 2007), and *Archaeologies: Journal of the World Archaeological Congress* (E. Kansa et al. 2010). Additional articles authored or co-authored by Open Context developers make mention of Open Context within the context of discussion open archaeological data in general (E. Kansa 2012), and a study of the needs of archaeology data reusers (Faniel et al. 2013). Case study articles describing the integration and reuse of zooarchaeology datasets via Open Context, and co-authored by Open Context developers, have recently been published in the *Journal of Archaeological Method and Theory* (Atici et al. 2013) and the *International Journal for Digital Curation* (E. Kansa, S. Kansa, and Arbuckle 2014). tDAR's executive director, board members and staff have also published articles promoting tDAR, primarily in professional association newsletters such as the *SAA Archaeological Record* (McManamon and Kintigh 2010), and the *Newsletter of the Center for the Study of Architecture* (McManamon, Kintigh, and Brin 2010), as well as the journal *Heritage Management* (Kintigh and Altschul 2010). A brief, generally positive review of the tDAR repository and website was published in *Choice: Current Reviews for Academic Libraries* (Cleland-Sipfle 2013).

These articles, conference presentations, and endorsements by professional associations are evidence of increasing prominence and visibility of these repositories. It is also encouraging

to see developers and stakeholders collecting evidence of archaeological data needs and best practices to inform the design and functions of their repositories, and publishing the results of these efforts. However, none of these sources of information have provided a direct, comprehensive comparison of features of both repositories, to provide guidelines for researchers to help them decide which repository would be optimal for their data. As might be expected, articles and presentations by repository stakeholders tend to focus on their own repository's strengths and approaches, while providing only limited mention of other repository options, and discussion of cases in which a researcher might appropriately choose NOT to use their repository is often absent or only briefly addressed.

By contrast, this article makes direct comparisons between both repositories, and discusses both the merits and potential drawbacks of each. Each repository has unique features and approaches to data sharing, archiving and publication, and will be optimal in some cases but less optimal in others, depending on the goals and intent of the researcher, anticipated reuse needs, and parameters and characteristics of the project's data. This type of comparison enables the matching of individual project data attributes and requirements with repository qualities and attributes which can best meet these requirements, facilitating data repository selection for the purposes of data deposit.

It is worth mentioning that online archaeological data repositories are not the only options for sharing and preserving archaeological data. Some academic research institutions accept datasets in their institutional repositories, so archaeological data collected by researchers or projects affiliated with their institution could be deposited there. There are also multidisciplinary online repositories that accept data from many different disciplines. For example, Dryad (http://datadryad.org/) is a repository that is affiliated with specific research

journal publishers, and hosts datasets that supplement articles published in those journals. Figshare (http://figshare.com/) is a multidisciplinary repository which does not require affiliation with a specific institutional or publication affiliation, though they also partner with several publishers of archaeology research, including Taylor & Francis and Public Library of Science (PLOS) (Figshare 2015). Figshare in particular holds a fairly significant amount of archaeology-related research data (on June 18, 2015 an advanced search in Figshare for items assigned to the Category "archaeology" yielded 279 results; a search for items with the Tag "archaeology" yielded 1659 results) which demonstrates that it is meeting the needs of some archaeology researchers, and that it should be remembered as a source when searching for archaeology data for reuse or comparative purposes. These other options can be valid considerations, especially if there is a reason why the data from a specific project cannot be put in an archaeological repository (examples of project attributes which might prohibit deposit in both tDAR and Open Context are mentioned in the "Shared Characteristics" section of this article).

However, while some archaeological research data is collected and produced by academic researchers, much archaeological research data is also collected outside of academia, and not all of it results in academic publications. In fact, archaeological data is frequently collected through cultural resource management activities, conducted by contract archaeologists and sponsored by state or government agencies, museums, corporations and other non-academic institutions. An online archaeological data repository can bring together related data from many different kinds of institutions, agencies, and projects, providing richer context and greater discoverability and access to a much broader range of research. In addition, disciplinary data repositories have the opportunity to customize their features to better suit the unique needs of researchers and data types and uses in that discipline, whereas multidisciplinary and institutional

repositories serving many different research communities with many different types of subject matter cannot. Amber Seely (2005, 9-10) describes the ideal "comprehensive archaeological database," which should include images, maps, links to project or dig site websites, data about the excavation and lab analysis, and references with full text if possible to related articles, books, and gray literature reports. It should include data from museums and archives as well as academic projects, it should be "deeply indexed with specificity, exhaustivity, and a controlled vocabulary," and should be searchable by "site name, site number, geographic location, archaeologist, sponsoring institution, report title, material, culture, occupation date, publication date, and keyword" (10). Through the collaboration of archaeology professionals and data curation and information management specialists, both tDAR and Open Context have realized many aspects of this vision of an ideal archaeology database, with a level of specialization that multidisciplinary and institutional data repositories cannot hope to match. Since online archaeological data repositories have significant advantages to offer, a comparison of tDAR and Open Context options and consideration of their suitability is a vital component of archaeology data stewardship decisions.

Sources and Methods

The information about each repository was collected from website user documentation, developer documentation, help files, directions for depositors, articles and blog posts about the repositories, and through an exploration of the data files and search interfaces. When specific details about repository practices or policies were not immediately apparent from this documentation, repository managers and staff were contacted directly for clarification.

Table 1 illustrates the framework used as the basis for repository comparison. The *Repository Quality* categories "Sustainable," "Visible," "Long-Term Preservation," and

"Supports Data Use & Reuse," and the *Ideal Characteristics* identified within each category, are derived from factors which have been identified by the information science community as indicators of successful, trustworthy digital repositories, regardless of which discipline(s) they serve. Formal metrics and evaluative tools, such as the *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)* (OCLC and CRL 2007), and the 2011 revised version of TRAC released as the *Audit and Certification of Trustworthy Digital Repositories* (CCSDS 2011), which later became known as the *Trusted Digital Repository Checklist* and *ISO 16363* (CRL 2015) were consulted for guidance. However, this article is not a formal evaluation of these repositories using the TRAC and Trusted Digital Repository metrics, and should not be interpreted as official certification or endorsement by the agencies responsible for these audit criteria. In addition, the article "Scientific Data Repositories on the Web: An Initial Survey," (Marcial and Hemminger 2010) identifies characteristics which are indicative of successful data repositories, and these qualities were also considered in the creation of this framework. Finally, the "Practical Considerations" category listed in Table 1 includes additional characteristics which could affect the suitability of the repository for a particular type of project, such as file types accepted, cost to deposit, and options for restricting access to deposited data. Even if a repository is technically doing everything correctly in terms of reliability, sustainability, and adherence to preservation best practices, it does not necessarily mean that it is the best fit for every project. Researchers will need to compare the project needs and data parameters with the characteristics of each repository to determine the best choice for their data.

Shared Characteristics of tDAR and Open Context

Although the choice between tDAR and Open Context will usually be based on their differences, it is important to note that before making that decision, the researcher must first

determine whether an online disciplinary archaeological repository is appropriate for their project data, through an examination of the similarities shared by tDAR and Open Context. If any of the following characteristics are unsuitable for the project, the researcher is encouraged to explore one of the alternative data management options mentioned in the "Why Compare Open Context and tDAR" section of this article.

Sustainable

Long Term Funding Model, and Multiple Funding Sources. As noted by Marcial and Hemminger in their survey of scientific repositories (2010), characteristics of a successful and sustainable digital repository appear to be multiple sponsors, as well as an independent business plan and funding model geared toward providing ongoing support (2047). They observed that many scientific data repositories depend upon governmental funding during the initial development period, but this funding is often limited to this initial stage. Since ongoing funding for maintenance and support for curation and preservation in the long-term is usually not provided, repositories need the investment and support of multiple partners, and a sustainable funding plan, if they are to be entrusted with data stewardship in the long term.

Both tDAR and Open Context have received grant funding from multiple agencies for initial and ongoing development and maintenance, but have also partnered with additional institutions and agencies. Both also charge a fee to deposit, which is part of a long-term sustainable funding model (the success of which is in part contingent upon continuous additions of new projects, and the willingness and ability of depositors to pay for these services).

tDAR development and ongoing operations have been funded by grants from the National Science Foundation (NSF), the Andrew W. Mellon Foundation, the National Endowment for the Humanities (NEH), and the Higher Funding Council for England acting

through the Joint Information Systems Committee (JISC). It is run by Digital Antiquity, a non-profit organization established in 2009 which is currently housed at Arizona State University, with representatives from many institutions (including University of Arkansas, Arizona State University, the Pennsylvania State University, the SRI Foundation, Washington State University, and the University of York) on the Board of Directors and Professional Advisory Panel (Digital Antiquity 2015a). This diversity of funders and support of multiple grants, as well as the participation and investment of personnel from many different institutions, is a positive indicator of stability. In addition to ongoing grant support, in June 2013 Digital Antiquity announced they had reached a four year agreement of support with the Society of American Archaeology (SAA). SAA will provide annual financial support to Digital Antiquity and tDAR, and promote the repository to archaeology researchers. In return, tDAR waives the deposit fee for student SAA members, up to three free file uploads per year for the four year period (Brin 2013b). This is evidence that Digital Antiquity is aware that diversifying funding by recruiting financial support from professional associations is critical, and that the visibility and opportunity to recruit early career professionals that such an agreement provides sets them up for long-term growth. Although the future of Digital Antiquity as an organization is not completely decided (according to the Digital Antiquity website's "About" page, "at the end of its initial development, the organization (and tDAR) may be established as a 'stand-alone' not-for-profit organization or be incorporated into an appropriate non-profit, such as a professional association" [Digital Antiquity 2015a]), the endorsement from AAA and partnership with SAA show that they have the connections and support which could make this possible. In addition, they have made provisions to ensure that there is a contingency plan for stewardship of tDAR data. Digital Antiquity has a formal Memorandum of Agreement with Arizona State University's libraries and institutional

repository, so that in the event that Digital Antiquity is dissolved or no longer capable of sustaining the repository, responsibility for preserving and maintaining tDAR will be assumed by the ASU library (Ellison 2012).

Open Context has also received developmental and operational funding from NSF and NEH, as well as the Encyclopedia of Life, the American Council of Learned Societies, the Alfred P. Sloan Foundation, the Institute of Museum and Library Services, and the William and Flora Hewlitt Foundation (AAI 2015h). It is operated by the Alexandria Archive Institute (AAI), a collaborative organization which has been registered as an official non-profit organization in the state of California since 2001 (AAI 2015e). AAI includes diverse institutional representation on the Board of Directors and Advisory Board (including UCLA, Harvard University, University of Chicago, UCSF, UCSD, Boston University, Smithsonian Museum of Natural History) (AAI 2015l). They have also partnered with the California Digital Library at the University of California, an award-winning digital library and repository system, which currently provides data archiving and preservation support for Open Context data (AAI 2015i).

Both tDAR and Open Context are following in the footsteps of the aforementioned, long-operating Archaeology Data Service (ADS). The ADS was established in September 1996, and was originally one of five discipline-specific repositories established in the United Kingdom by the Arts and Humanities Data Service (AHDS), which was funded by the Arts and Humanities Research Council (AHRC) and the Joint Information Systems Committee (JISC). Although the AHDS lost its funding in 2008 and no longer exists, the ADS has continued to thrive in large part due to the investment of many additional partners, including ongoing funding from AHRC and the Natural Environment Research Council (both of whom also require funding recipients to deposit their data in ADS), and project funding from JISC and the European Union (University

of York 2015a; 2015b). The ADS is an example of an archaeological data repository that has outlasted its initial development stage and has proven to be a sustainable and reliable repository.

On the other hand, one of the earliest U.S.-based archaeological data repositories, the Archaeological Data Archive Project (ADAP), is no longer operational today. ADAP ceased in 2002 after ten years in operation. Eiteljorg (2002) describes several factors contributing to the abandonment of the ADAP, including a loss of grant funding due to budget cuts. ADAP was also limited in the size and number of data contributions, which resulted in decreased utility for future studies and comparisons of the data, since the scope of data was very broad but with shallow coverage of each data type and content area. According to Eiteljorg (2002), the data was returned to the depositors, and the ADAP website no longer provides access to the data archive. While this does not necessarily mean that the data was lost, it again became the responsibility of the individual researcher to preserve it and ensure that it could be accessible in the future.

Marcial and Hemminger (2010) also note that holdings size and scope is another discriminating factor in the sustainability (or lack thereof) of repositories, in addition to sustainable funding. Since tDAR and Open Context have both moved to an operational funding model which relies in part on a one-time fee for deposit to cover the costs of their services, the continued longevity and success of both repositories may be reliant upon buy-in and continued use by the archaeological community. It remains to be seen whether both will be able to simultaneously survive, or whether dividing the discipline between the two will jeopardize one or both of them. The Alexandria Archive Institute states on their website that Open Context is intended "to complement those [efforts] of leaders like the Archaeology Data Service and Digital Antiquity" (AAI 2015f), an indicator that they are not intended to be competitors, but rather to provide complementary services that serve slightly different needs for archaeology data

publication, and this could be the way in which both could successfully survive. However, although they are not meant to be mutually exclusive services, the reality is that depending on the goals of the researcher, the resources available to them (money, time, personnel and expertise) for data preparation and deposit costs, and the data characteristics, it is likely that researchers may need to make a choice rather than utilizing both. Therefore, the approach of each repository, and their strengths and limitations, may yet have implications for sustainability. However, since both repositories have contingency plans and external partner support for archiving of contributed data which mitigates the risk of data loss, worries that the repository may lose financial support in the future should not deter archaeologists from depositing in either of these repositories, because that could become a self-fulfilling prophecy. The immediate benefits of depositing in one of these repositories, and the potential benefits to the future of archaeological research as these repositories grow and develop, are strengthened and supported with each new deposit.

Visible

Publicity and Awareness within the Disciplinary Community. The more visible a repository and its contents are, the more likely that data will be discovered, used and reused, and the more likely that it will continue to receive new data submissions, thereby continuing its relevance and perceived value to the scientific community. If the repository relies on deposit fees for sustainable funding (which is true of both tDAR and Open Context), visibility and growth can also directly impact sustainability.

In addition to the conference presentations and publicity, journal and newsletter articles, professional association endorsements and nods by funding agencies mentioned in the "Why Compare Open Context and tDAR" section of this article, both repositories are discoverable via

re3data.org, an online directory of data repositories, which can be used by researchers to locate sources of data for citation and reuse. Both tDAR and Open Context are also listed as "Recommended Repositories" by the *Journal of Open Archaeology Data* (Ubiquity Press 2015).

Although tDAR seems to have a visibility advantage due to its formal affiliation with SAA and the endorsement by AAA, it is very common to see both tDAR and Open Context mentioned in listings of sources of archaeological data, and at national conferences of professional associations.

Evidence of Visibility and Data Use. Both tDAR and Open Context provide a means of tracking and demonstrating evidence of the impact of and interest in an individual's archaeological research project, to supplement tenure cases or to use when advocating for funding support for related research projects.

Open Context provides detailed and immediate statistics, including the actual number of views on each artifact object record and project record. This information is publicly available. In 2015, tDAR added features which allow data contributors with administrative access the ability to see the views and downloads statistics for their contributed resources (Brin 2015b). tDAR also provides annual aggregate statistics in their public news blog (tdar.org/news), indicating the top ten most viewed records, most downloaded files, most used site name keywords, most used people (researcher names), most used institution names, and more (see, e.g., "Taking a Look Back at 2014," Brin 2015a).

Long Term Preservation

Preservation Commitments and Contingencies. Ensuring that data will be preserved for the long term is in the best interest for the field of archaeology and the subjects of archaeological study, so selecting a repository that will provide adequate preservation services is critical.

tDAR's stated mission is to preserve and provide access to archaeological data in perpetuity. The preservation statement on their website (Digital Antiquity 2013c) lists the steps and precautions they take, which are in keeping with accepted best practices in digital preservation. These include: systematic checking for file deterioration; support for open file types; routine on-site and off-site backups; maintenance of original file and preservation file copies, with migration to new file types if necessary due to changing technology. The contingency plan and agreement to transfer to ASU's digital repository in the event that tDAR ceases is a necessary precaution. Researchers selecting tDAR as their repository can feel confident that tDAR has appropriate mechanisms in place to plan for long-term digital preservation. Since tDAR is responsible for performing their own digital preservation and maintenance, ASU's agreement to assume responsibility for the data in the event of repository closure is critical. If this occurs, it should be noted that future data access and preservation may be limited by the capacity and funding of ASU's digital repository.

Open Context, by contrast, warns that they are not a repository in the traditional sense, in that they do not directly guarantee preservation, though they do provide metadata and tools which will be helpful for preservation in other repositories which have that as their mission (AAI 2015i). In fact, they encourage researchers to deposit their data in other repositories in addition to Open Context, which is made possible by the fact that researchers are encouraged to utilize Creative Commons licensing for their Open Context data. Although Open Context is more like a publishing platform than a long-term preservation repository, they have partnered with the California Digital Library (CDL), which is a leader in the field of data curation and preservation. Data deposited in Open Context is also automatically deposited with the University of California Curation Center's Merritt preservation repository as a "dark archive," so although the data

cannot be found and accessed publicly through CDL's website, CDL maintains a preservation copy of the data in accordance with their preservation practices and procedures [Perry Willett, personal communication]. This means that if Open Context experiences technical issues or data loss, they should be able to recover the data from the CDL archives; if Open Context ceases to operate, the data is not permanently lost.

Regardless of which repository is considered, if a researcher is affiliated with an academic institution that has an institutional repository, they should be encouraged to check with their institution's library or institutional repository to determine whether depositing their data in *both* an archaeological data repository and the institutional repository would be possible. This would provide an additional safeguard in the event that the archaeological data repository is eventually not sustainable. However, both repositories do have adequate mechanisms in place to protect against data loss and provide for preservation, though tDAR accepts greater responsibility for long-term preservation in their mission statement than Open Context.

Supports Data Use & Reuse

Scope (format and subject of data). Selecting a repository that contains data of similar format and subject matter is one way to increase the likelihood that the repository is well equipped to process and manage the researcher's data, and to increase the chances that other researchers looking to reuse existing data of that type will choose to look in that repository, and will discover and potentially reuse the data. This is one advantage of selecting an archaeological repository such as tDAR or Open Context, as opposed to a multidisciplinary or institutional repository. Both tDAR and Open Context primarily contain data which describe prehistoric and historic archaeological artifacts (including ceramic, bone, metal, wood, textiles, stone, etc.) and

architectural features, photographic and map images of artifacts and excavation sites. This information is deposited in the form of project reports, summaries, articles, images and descriptive data spreadsheets of artifacts, excavation/survey site features and methods, maps and other related data. Both tDAR and Open Context are particularly strong in faunal zooarchaeological data.

There are other types of data which can also be collected during the course of archaeological research and subsequent analysis, including DNA analysis, botanical analysis, chemical analyses, etc. These types of data are not commonly found in tDAR or Open Context at this time, although in general tDAR may be better equipped than Open Context to handle this variety of data formats and types, due to the way that Open Context processes and displays data contents. Other specialized repositories may be better suited to manage and process this data in a way that makes it more discoverable and usable. For example, mitochondrial DNA sequencing data obtained from archaeological specimens may be more appropriate for a repository such as NCBI's GenBank (http://www.ncbi.nlm.nih.gov/genbank/), or other NCBI genetic databanks like the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) or Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi). Resources such as re3data.org can help researchers locate other appropriate disciplinary repositories. For greater interdisciplinary visibility and discovery, archaeologists could consider depositing the report of excavation and other archaeological contextual data in an archaeological repository, even if other types of specialized data are deposited elsewhere, and include cross-referencing links to related data records in both repository entries and in publications derived from this data.

Public Access to Descriptions of Data and Projects. In both repositories, the basic descriptions of all deposited projects are publicly available online. This includes details such as

the summary of the project, names of researchers/authors, culture and temporal keywords and generalized geographic location, and the number of files and types of data included. Free public access to information about archaeological research projects increases the visibility and impact of the archaeological data projects, improves the likelihood of data discovery and reuse, and provides a means of making the results of publicly-funded research available. It is also in keeping with many of the ethical principles and goals of responsible archaeology research, as outlined by professional associations such as the Society for American Archaeology (SAA 1996) and the World Archaeological Congress (WAC 2015). At the same time, this means that even if the contents of the dataset are not publicly viewable (tDAR offers the option to restrict public access to data file contents, as described in a later section of this article), the depositor of the data must still have the authority and permissions to publicly share descriptive information about the project.

If the archaeological researcher investigating data deposit options is not the "owner" of the project data, or if they are unsure whether they have the authority to make the decision to share project information, they should consult with the sponsoring or affiliate institution, government agency, project managers and advisors, and other stakeholders to obtain necessary permissions before sharing project information or data. The researcher should also consider the possible impacts (potentially positive or negative) to communities or individuals within the studied cultural group, if project information is shared publicly. Ownership of archaeological data and the right to share it is complicated by differences between countries and cultures and their laws (for a thorough discussion of the cultural, legal and ethical considerations relevant to archaeology data sharing, see Alleen-Willems [2012; chapter 7]).

Persistent Identifiers. Both tDAR and Open Context supply persistent identifiers to data records. A persistent identifier is a unique, alphanumeric label assigned to a digital publication (article, document, record, etc.) which never changes, and can be used to locate the publication online, and included in citations referencing the publication. Persistent identifiers are also useful in tracking citation history and demonstrating impact.

tDAR assigns Digital Object Identifiers (DOIs) to each data record which is active and contains at least one attached file/document (deVos 2012). Open Context assigns Archival Resource Key (ARK) identifiers to individual items of data [individual artifacts/features], and DOIs to projects and datasets" (AAI 2015i). In general, this means that both repositories use the DOI system for persistent identifiers to entire data files/records, which is becoming commonplace for electronic articles, and digital libraries and archives. However, Open Context takes it a step further in granularity, so that an ARK identifier can be used to provide persistent access to an individual artifact or feature and its associated attributes and values as a stand-alone information object, rather than limiting the persistent linking to the entire dataset (or datasets) which are the sources of these values (see Willett and Kunze [2015] for more information about ARKs).

Practical Factors

Archaeological Research Site Location Security Protection. Archaeological site privacy is often a major concern, and is an example of a type of sensitive data that is a concern for researchers in this discipline. Disclosing the exact location of dig sites could encourage looters or other disturbances. To address this issue, both tDAR and Open Context provide mechanisms to facilitate masking of specific site coordinates. While completing the "Spatial Terms" section of the project metadata web form in tDAR, depositors assign relevant geographic keywords, and

draw a box on a map to indicate the approximate latitude/longitude range which includes the research site(s). The depositor decides how general or specific to be when choosing geographic keywords, and how broad or narrow to make the map box (boxes smaller than 1 square mile are automatically widened by tDAR). Depositors can also enter the exact geographic coordinates of relevant site(s) in tDAR, but this is not required (Brin 2013a). Open Context reminds depositors to consider the sensitive nature of site location data, and they will work with depositors to ensure that the required geographic location information is published with "reduced precision." They also recommend including information in the data records to notify users that location information has been generalized (AAI 2015c). In both repositories, depositors are responsible for redacting or generalizing any specific location details contained within the contents of the data files prior to deposit, if necessary.

Since tDAR and Open Context were both developed specifically for archaeology data, they include built-in mechanisms and documentation procedures to provide sufficient location identification information to establish the crucial geographic context of archaeological data, without compromising research site security. However, if public disclosure of generalized location information would still compromise site security or ongoing research, or pose privacy risks for current residents/stakeholder communities, sharing the project data publicly in these repositories may not be advisable. Researchers should be encouraged to contact the repositories to discuss the amount of public location data required for deposit, if they are concerned about the risks.

Cost to Depositor. Both tDAR and Open Context charge researchers a one-time fee to deposit data. This can be a business model strategy for long-term sustainability (Marcial and Hemminger 2010), and evidence of this can be seen in the success or decline of archaeological

repositories historically. The successful ADS has had a "Charging Policy" in place since the late 1990's (ADS 2007), while the now-defunct ADAP did not charge depositors a preservation fee for their data (Eiteljorg 2002). However, this also means that depositors must be willing and able to pay for the cost of data access, maintenance and preservation services provided by the repository.

In both repositories, the deposit cost is generally dependent upon the number and size of files. tDAR sells file space (up to 10 MB per file) on a sliding rate scale, which as of July 14, 2015, starts at $10(USD) per 1 file/10 MB for up to 99 files, while 100+ files are $5(USD) each (Ellison 2015). This represents a recent, significant drop in tDAR's cost per file; prior to July 14, the cost given on the pricing page (Digital Antiquity 2015d; when accessed on March 17, 2015) was 50USD per 1 file/10 MB, 10-49 files was 40USD per file, 50-99 files was 30 USD per file, etc. These prices do not include support for file description or content review; depositors upload the files and create the metadata descriptive documents on their own. For more complex projects requiring customized services and support, tDAR also offers curation services at additional cost per hour: "file checking, metadata drafting, basic quality control" at $90 per hour; "consulting, planning, programming, management" at $180 per hour (Digital Antiquity 2015d).

Open Context states that their "publication fees vary between $250 and $6000 depending on the complexity and size of the contributed database and related content" (AAI 2015c). Open Context processes and provides descriptive content within the contents of datasets upon deposit (more details on this process later in this article), and therefore their publication fees are covering more than just file storage space. Depositors are paying for the work that Open Context will do to enhance the value and usability of the datasets, and depositors can elect to pay more for additional optional web services. However, this also means that the cost is dependent upon

complexity of files and amount of processing required, in addition to number and size of files, so an advance estimate of cost requires completion of an estimate web form and communication with Open Context editors (AAI 2015b). Since both repositories will provide advance estimates to researchers based on anticipated project data, researchers who want to compare price quotes will need to do so with the specifics of an individual project in mind. Generalized comparisons of price are difficult since the cost models are based on different criteria, and it is dependent upon the type of files and amount of data curation support needed. tDAR's reduced prices can make basic file self-deposit more affordable, but the data (and depositor) would not receive the same level of documentation and support for consistent curation and description that Open Context provides.

After obtaining an estimated cost for their ideal repository, researchers may include data stewardship costs in grant or project proposal budget requests. Repository managers may also be able to offer suggestions for funding assistance or discounts for certain types of projects or society membership status. For example, Open Context currently offers a 25% discount for deposits of data resulting from a doctoral dissertation improvement grant (AAI 2015b), while student members of SAA are currently eligible to upload three free files (up to 30 MB) annually to tDAR as part of their membership benefits (SAA 2015).

If funding support for data deposit is not available, or the cost estimates exceed the available budget allotment, then depositing in one of these repositories may not be a viable option. However, best practices in archaeological data collection, documentation, storage and management should still be followed as closely as possible, to maximize the chances that the data will still be available and suitable for deposit at a later date, should funding become available. Following similar guidelines, even if the data is deposited elsewhere, may also

facilitate future comparison and compatibility of this data with the holdings of tDAR and Open Context. The *Archaeology Data Service/Digital Antiquity Guides to Good Practice* (ADS and Digital Antiquity 2015) provide sound guidelines for archaeological data management.

Data Preparation and Processing Required Prior to Deposit. Both repositories require the depositor to complete data preparation tasks prior to deposit, such as assigning descriptive metadata, organizing data by the repository's conventions, file conversion to accepted formats if necessary, etc. The amount of effort required for data preparation before deposit will vary depending on the procedures and methods used by the researcher during data collection and description and organization, and the requirements of the chosen repository. Depositors must be willing and able to provide all necessary descriptions and prepare the files for deposit according to the procedures required by the repositories, and should carefully read the repository's deposit instructions to ensure that they will be able to prepare the data adequately. This process will be less cumbersome if researchers choose an intended repository prior to the start of a project, and use the repository's guidelines to shape data collection, documentation and file organization decisions.

tDAR provides detailed step-by-step instructions and guidance for depositors via their documentation wiki website (tDAR 2013). The "Planning your Data Contribution" and "Creating and Editing Resources in tDAR" sections of this webpage indicate that researchers will need to make decisions about how to organize/group their files and documents, and will complete web forms which guide the researcher through the process of assigning metadata (descriptive information) to the deposited files and projects. It includes a mechanism for "inheriting" metadata, which helps to streamline this process and reduce the amount of time required. For example, if the researcher is uploading 100 digital photo files taken during excavation at a dig

site, and also uploading a spreadsheet of ceramic sherd measurements and attributes describing artifacts collected during the same excavation, the researcher can create a "Project" in tDAR, enter the descriptive information about the site once (site number, geography, site type, temporal information, cultural keywords, excavation methods, investigator/institution names, etc.), and then assign the "Resources" (digital photo files and spreadsheet) to that Project, selecting the option to inherit the Resource metadata from the Project metadata. The records for the photo files and spreadsheet would automatically include all of the descriptive information from the Project record, and the researcher need only enter the additional descriptive information specific to the resource, or which differs from the Project information (for example, spreadsheet column and row label information, or summary description of photo contents). When depositing datasets, such as an excel spreadsheet of artifact attributes, depositors complete web forms which document and describe the contents of each column. i.e. Column type (numeric, text, coded values, measurement, etc.), Column name/label, Column category (chosen from predefined lists: for example, a "Cut Marks" column documenting cut marks on animal bones would be category "Fauna: Modification"), Narrative description if desired (example: "'Soil Color' column describes the color of soil in the associated excavation layer as observed by the excavating researcher, using the standardized alphanumeric classification system of the *Munsell Soil Color Charts*").

Open Context has very specific requirements not only for the description of the data, but also for the way that dataset content itself is organized and presented, due to the way that they process and publish the data (AAI 2015c; "Data Preparation" section). For example, if a coding system was used when documenting data attributes, the coding needs to be removed and replaced with "intelligible text." To give this example in archaeological terms, when documenting the

presence or absence of cut marks on bones, some might choose to use a code to facilitate data entry and analysis, such as entering a "1" for the presence of cut marks on a bone or a "0" for absence of cut marks. However, under Open Context's data preparation guidelines, code should be replaced with text, so in a column for "Cut Marks," "1" would need to be changed to "present" and "0" changed to "absent." This is different from tDAR and many other social science repositories which accept coded datasets as long as an explanatory codebook file is also included in the deposit. The reason for this difference is that unlike tDAR, Open Context does not simply present a downloadable excel file and associated codebook for viewing and use; as a data publishing platform, Open Context editors extract and publish the contents of the datasets from within files, and publish each item of data at the object level, displaying all of the object's descriptive information on the webpage record for that object. Using the cut mark coding example, rather than displaying "Cut Marks: 1," which wouldn't make sense to the viewer, or which could be incorrectly interpreted as a numeric count of the marks, Open Context would display "Cut Marks: present." This makes the data more visible at a glance without analysis and interpretation by the user, but can increase the amount of data preparation work required from the depositor.

Open Context also describes requirements for providing narrative descriptions of each field, creating a table to link image files with the object(s) they depict, checking for and correcting inconsistencies in spelling or terminology used, etc. Meeting these requirements will require less effort on the part of the depositor during data preparation if the decision to deposit project data in Open Context is made BEFORE data collection and documentation begins. This would allow the PI or project manager to develop and implement procedures and practices that are tailored to Open Context's guidelines, reducing the amount of processing required. If the

choice to deposit in Open Context is made during or after data collection and documentation, more time and effort to clean up and restructure or edit the data content to meet Open Context's guidelines may be required. Open Context's website indicates that their editors will work with depositors to aid in corrections for consistency, decoding, disambiguation, etc.

Further study of the experiences of researchers who have deposited data in these repositories is needed before making direct comparisons about the relative ease or difficulty of data deposit between tDAR and Open Context. In the absence of this information, researchers should be aware that both repositories do require some level of data preparation and effort by the depositor to describe and clean up data, and they will need to review and compare the deposit process documentation provided by both repositories to determine how much processing and preparation would be required for their specific project, and decide whether they are able and willing to complete the necessary processing to meet a repository's deposit requirements.

Licensing and Copyright. Potential depositors need to consider any existing restrictions on their ability to share and publish their data due to licensing or copyright ownership. Sharing data in tDAR or Open Context does not mean transfer of "ownership" of that data to the repository, but in both cases the repositories have very specific requirements about the licensing terms of deposited data and files, and depositors must have the authority to share the project data under the required license terms.

tDAR's Terms of Use Agreement (Digital Antiquity 2013e) indicates that data content is subject to the terms of the Creative Commons Attribution 3.0 Unported License, unless otherwise indicated with respect to a particular file. The Contributor's Agreement (Digital Antiquity 2013b) indicates that depositors can designate alternative license terms for specific files, and since tDAR allows restriction of public access to content of data files, depositors have

the ability to apply more restrictive terms for licensing and reuse if they need to due to constraints imposed by publishers or institutional/organizational sponsors. They may still be able to archive the data with tDAR for preservation purposes, without releasing full access to the general public for reuse.

All data content in Open Context is publicly available for viewing and download, and must either be in the public domain, or under a Creative Commons license. Datasets containing primarily factual information "those that mainly include physical measurements and adhere to widely used conventions in nomenclature and recording" are licensed under Creative Commons-Zero; Datasets containing interpretive content should be deposited under Creative Commons Attribution license (AAI 2015c). Since more restrictive licensing terms are not permitted, and all data content in Open Context is public in order to align with their mission to provide open access to data, some projects may have ownership or licensing issues requiring greater restrictions than can be offered by Open Context.

More information about what each of the Creative Commons license types mean can be found at http://creativecommons.org/licenses/ (Creative Commons 2015). Factors which could affect a potential depositor's ability to designate these licenses include intended journal publisher's requirements (some journal publishers require data to be deposited in a specific place, under their terms), whether or not they were the primary investigator on a project, whether they have permission from the institution or agency that sponsored the research, permission from the government of the research location, etc. This can only be determined on a case-by-case basis, and experts in licensing and copyright law in the United States and in the research site country may need to be consulted. If there is a need for more restrictive licensing or permissions, tDAR's documentation seems to indicate that they provide more flexibility and allow more

restrictive licensing choices than Open Context. However, regardless of the repository chosen, the "owner" of the research data should always be involved in the decision to deposit and share data, and in discussions of licensing and copyright.

Differences between tDAR and Open Context: Decision Points

The following "Decision Points" represent major differences between tDAR and Open Context that can affect the choice that the researcher makes when depositing their data. The researcher should evaluate each Decision Point's comparison of repository features, discussion and recommendations within the context of their individual research project and data parameters, to determine which ones will have the greatest impact for their particular situation. The points are not listed in any particular priority order, as this may vary greatly between projects. In some cases, the researcher may decide that one specific Decision Point recommendation eliminates one of the repositories as a viable option, while in other situations the researcher may find that the results of the Decision Point recommendations are split. When both repositories are indicated for different reasons, the researcher will need to decide which points are most important for the utility and future use of their data, as well as their research goals and intentions for the deposit, and which ones can be compromised for the benefit of the other priorities. Alternatively, since neither repository requires transfer of copyright, theoretically the researcher could consider depositing in both, though preparation and submission of data in multiple repositories does result in added cost and time commitments.

For quick future reference during archaeology research data consultations, Table 2 ("Differences between tDAR and Open Context") contains a brief summary and comparison of each repository's qualities with respect to the Decision Points, and recommendations for consideration that can be used to aid in determining which repository may be ideal for a specific

archaeology data project. The "Summary and Conclusions" section of this article also summarizes the unique characteristics of each repository. The discussion below provides additional context and implications of these unique differences between tDAR and Open Context.

Access Restrictions. Archaeological data sets may include some sensitive information, affecting the ability to publicly share the data. This can include information which could compromise site security, such as site location or documentation of items perceived as having high monetary/ commercial value, as well as culturally-sensitive subject matter such as images of human remains, or documentation of objects with ceremonial or religious significance that are not supposed to be seen or known by others outside of that culture (Alleen-Willems 2012, 50-51). For many reasons, including sensitive material, ownership and copyright issues, publisher restrictions, and status of ongoing research, some researchers may need or want the ability to restrict access to the contents of data files to certain users or embargo access for a specified time period. On the other hand, researchers may be required or strongly encouraged to provide public, open access to all. Ensuring that the repository has the appropriate options to allow for data access while still meeting required parameters and constraints is critical.

tDAR offers several different options for data access restriction. Although the upper level project description information (metadata) is always public, the Contributor's Agreement states that the contents of datasets may be restricted to certain users indefinitely, or embargoed for public access for up to four years post-deposit. The researcher can also allow users to contact them directly to request access permissions to restricted content (Digital Antiquity 2013a, Section C). Anyone can browse or search the public metadata in tDAR, but all users must

register for a free tDAR account before they can view or download data contents. However, it should be noted that even when content is marked "confidential" or "embargoed," tDAR staff reserve the right to access to all files for preservation purposes (Digital Antiquity 2013d), and this may need to be disclosed to stakeholders prior to deposit.

Open Context, by contrast, does not permit any restrictions of public access to deposited data content. All data deposited in Open Context must be made openly available to the general public, and no user registration is required to view data contents. They note that this is in keeping with the American Library Association's guidance for protecting patron rights to confidentiality and academic freedom, and it also means that the contents of Open Context data can be indexed by commercial search engines (AAI 2015j). Choosing Open Context can indicate a commitment to removing all access barriers to the data and to protection of user identities and confidentiality, and may also lead to greater discoverability of data contents via commercial search engines.

File Types Accepted. Both tDAR and Open Context are well suited to accept common types of text files such as .pdf, .doc/docx, dataset/database file types such as .xls/.xlsx, .accdb/.mdb, .csv, and image files such as .tiff, .gif, .jpg. Open Context prefers datasets to be submitted as a Microsoft Excel (.xls) file due to the way they process the data; they later convert the submitted files into non-proprietary, "open" formats (AAI 2015c; "Data Formats and Structures" section). tDAR accepts proprietary file types, and retains both the original file version as well as a converted, non-proprietary archival copy for preservation purposes (Digital Antiquity 2013d). A full list of tDAR accepted formats is available at http://www.tdar.org/why-tdar/contribute/ (Digital Antiquity 2013f) A full list of Open Context's accepted formats is available at http://opencontext.org/about/publishing (AAI 2015c).

Despite these similarities, there is an important difference between the two repositories with regards to certain specialized file types, which are becoming increasingly relevant and useful for archaeological research. Remote sensing files, 3D scans and point clouds, and geospatial data are accepted by tDAR, which has the ability to retain and present this type of data in a way that allows for future manipulation and analysis of the data. For example, in the tDAR documentation wiki there is a description of how 3D sensory datasets archived in tDAR can be viewed and used (Brin 2014). On the other hand, due to the way that Open Context processes and publishes archaeological data, these types of files are accessioned as static digital objects, and therefore they are not shared or maintained in a manner that facilitates data manipulation and visualization. Open Context editors will openly refer researchers to tDAR or other repositories for deposit of these file types (Eric Kansa, personal communication; AAI 2015b).

Data Record Organization, Retrieval, Description and Comparative Integration. One of the most immediately noticeable differences between tDAR and Open Context is in the way that datasets and their contents are processed and made available to users. The two repositories differ greatly in their approach, which directly affects the way that data is viewed, searched, and accessed. In tDAR, data project and file records and metadata (descriptive information about the data) are assigned and searchable at project/dataset/file level, NOT at individual artifact (dataset contents) level. By contrast, in Open Context records and metadata are assigned and searchable at both the project/dataset/file level, AND at the individual artifact (dataset contents) level.

A typical record in the tDAR interface will include a file containing the data content, which may be downloaded in order to access the content of the data, and if the file contains a dataset (such as a spreadsheet or .csv file), a chart is provided to indicate the structure of the

data, column headings, type of data, and indication of whether it is associated with a coding sheet or ontology. This presentation, which requires a download in order to view the contents of a dataset (the descriptive values for individual ceramic sherds, for example), is quite different from the presentation of data in Open Context. Open Context records are displayed at a more granular level of the artifact or individual data content, so for example, rather than downloading a spreadsheet of sherd measurements from a project in order to view and manipulate them, the user can see a separate record for each individual sherd, including measurements, photos, etc. The data and contents of the datasets are described and assigned metadata at a much more detailed level in Open Context, which allows for browsing and searching within the contents of many datasets at once, without requiring individual downloads of each source dataset (although downloads of source datasets are also available).

There are advantages and drawbacks to each approach. Open Context more deftly facilitates browsing of the actual data content, and integrates all attributes associated with a specific artifact in a more visible way, since it presents data contents immediately in the browsing/searching results screen. This allows objects and data from multiple different projects which share a desired feature to be extracted from within all of the different project files, and displayed in one uniform group of results. Open Context's application of descriptive content is also more uniform across all data contents. The granularity and consistency at which they apply metadata to individual parts of data contents results in a comparative data framework that is integrated into the ingest and data record creation process, which means that the work of preparing data for comparison and integration is done for the researcher (Faniel 2013), and is consistently applied to all data contents in this platform.

While this makes simultaneous browsing of dataset contents of multiple projects easier, at this point the only formats and mechanisms provided for the user to download the resulting data values require an understanding of how to use Atom, JSON or KML feeds to retrieve the facet data of summarized query results (AAI 2015g; "Faceted Search Results and Representations"), which is beyond this author's realm of understanding (and perhaps beyond some archaeology researchers as well), and may require the assistance of technical specialists for data retrieval and manipulation. If Open Context were to develop a built-in mechanism that was visible from the search/browse results screen (such as an "Export these results" link, with format choices that are familiar and compatible with common data analysis software packages), this would be a more intuitive way to produce an integrated, customized file for manipulation and comparative analysis. So although Open Context is designed in a way which can facilitate this kind of data retrieval and flexible reanalysis, it remains a largely theoretical advantage at this time unless one has the technical expertise to interpret and manipulate the Atom, JSON or KML feed formats. Open Context is still a relatively new platform and the developers frequently make updates to add new features and functionality; so in that sense the design of Open Context is poised to realize this advantage, and its potential usefulness will grow as more data is deposited and made available for comparative query results analysis. If the structure and contents of a project's data are compatible with Open Context's model, it can add value and improve discoverability and browseability of data values if deposited in Open Context.

By contrast, tDAR does not readily facilitate browsing of dataset contents and integration of data from other datasets and projects within their search/browse interface, and unlike Open Context, it is not designed to provide the same level of uniform structure and description of the contents of datasets when the data is deposited. This makes it more difficult to find and compare

other relevant data at the initial browsing retrieval phase, and requires the user to download and examine the contents of each relevant dataset, in order to determine how to compare the unstandardized contents. However, tDAR does provide other options which can be applied to the data for comparative analysis, post-deposit. Depositors and researchers can map dataset contents to a common "ontology", and/or filter search results to display only data which has already been mapped to an ontology. An ontology is "a formal description of the concepts and relationships that are relevant for a given domain" (Khosrow-Pour 2013:674), and by mapping the contents of different archaeological datasets to the same ontology (for example, mapping the contents of a dataset documenting cut marks on animal bones to a faunal butchering ontology), it allows researchers to more easily identify and compare related data, even if the original researcher used different terms or methods of describing the data. tDAR also provides a built-in data integration tool, which can be used to combine and compare the contents of different datasets within tDAR. This combined data can also be exported and analyzed using common quantitative statistical software, such as SAS, SPSS, or R (Brin 2015b). This represents an impressive step forward in archaeological data synthesis and reuse, and allows agile integration of disparate data sets. However, it is limiting in that the onus is placed on the researcher (the depositor and/or the researcher desiring to reuse the data) to develop ontologies and apply them to existing data sets within tDAR. Since it is reliant upon the creators of ontologies, and is not uniformly and consistently applied to all deposited datasets, this also means that there is not an ontology for every type of data and every dataset in tDAR; some tDAR data have an associated ontology and are ready for integration, but others do not. Application of tDAR ontologies and the data integration tool are also limited to data that is already in tDAR, and they are not required to use

existing standards of description, so this does not readily facilitate future integration with data that is located in other repositories or undeposited.

Open Context's answer to the issue of standardizing the descriptive metadata used to describe archaeology data is that they have begun implementing "Linked Open Data". According to linkeddata.org, Linked Data is a means of using standardized descriptive terms and links which can connect previously disconnected datasets across the web. By applying Encyclopedia of Life biological taxa, and place names from the Pleiades Gazetteer, across all relevant datasets, they are standardizing the descriptive information applied to data both within AND outside of Open Context. Essentially, this means that the data in Open Context can be connected to all data from archaeology or any other discipline that also uses these standard descriptive terminologies. They are also working on mapping their data to the CIDOC Conceptual Reference Model (CRM) ontology (AAI 2015k). The CIDOC CRM "is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to," and it is now an ISO standard (International Council of Museums 2014). Unfortunately, in its current implementation, the Open Context interface does not intuitively connect to other Linked Data (e.g., clicking on a Linked Data term in a data record does not retrieve a list of all other Open Context data records that apply that term, and also does not retrieve related data from outside Open Context), but they have set up the framework that will be necessary to make this a useful feature in the future. Also, since they use external existing descriptive standards such as CIDOC CRM, this means that the potential for future data comparison with external archaeology data sources, and even with data from other disciplines and types of sources that have adopted the same standards, is much greater with Open Context than with tDAR. Although tDAR is currently more agile and intuitive for data reanalysis

and reuse of some types of archaeology data, Open Context has set up a framework which, with a few new local interface developments and overall advancements in the field of Linked Data, could ultimately have greater impact in the realm of cross-repository and cross-disciplinary data integration.

As the expert on their data and how it needs to be used, the archaeology researcher depositing the data is the most qualified to judge which approach to data record organization, retrieval, and comparative integration will be most beneficial for an individual project. Factors to consider include:

- The structure, type, and number of data files
- The subject of the data: Does it map to an existing tDAR ontology? If not, does the depositor have the time and expertise to create a new ontology? Would the subject matter of the data benefit from the Open Context Linked Data classifications?
- Anticipated future uses of the data: Will integration with other relevant datasets be important? Is an integrated file combining dataset content that can be readily exported for reanalysis most desirable, or is it more important that dataset contents are integrated in a browseable, visual way during a repository search?

Editorial Control. Open Context has adopted an approach to editorial control and publication of archaeological data which is analogous to academic journal publishing models. Open Context screens their submissions, is selective about accepting projects and data, and provides some level of quality control. They state that all submissions must pass an editorial review and be approved

according to three basic standards: “methodological soundness and data quality, quality of documentation, and suitability for wider reuse” (AAI 2015c). Optional external peer review is also available, and data which have received peer review approval are indicated as such. Therefore, acceptance of a project could be presented as evidence of the quality of an archaeologist’s research output. Additionally, future data users can be reasonably sure that the data contained is sound and includes sufficient documentation to facilitate interpretation. This approach mirrors the publishing standards for articles in peer-reviewed journals, and is consistent with Open Context’s stated goal of providing a platform for data publishing. At the same time, acceptance of a project’s data is not guaranteed, and this model may not be well suited for deposit of data which is fragmentary or has less thorough documentation.

tDAR takes a different approach, and does not explicitly state that they review data for soundness, though minimum requirements for data documentation, such as completion of all required descriptive fields, must be met in order to complete a submission. The accession policy also states that they reserve the right to “review and remove files and metadata records that do not comply with [tDAR’s] policies” (Digital Antiquity 2013a). The instructions to depositors and accession policies indicate that editorial control and review of documentation is not exercised at the point of deposit, and prior approval is not required to submit data. While this does not provide the same level of prestige for the researcher compared to Open Context, and some data contents may have less documentation and contextual information than data in Open Context, tDAR’s approach allows for a wider reach and for inclusion of a greater variety of projects and sources of data. This means that tDAR can provide preservation services and public access to historical pieces of the archaeological record that might otherwise be lost, even if they lack thorough documentation by today’s standards. For example, tDAR contains a record entitled

"1933 and 1934 Casa Grande Ruins Excavation Photographs", which contains photographs from Civil Works Administration (CWA) excavations conducted at Casa Grande National Monument by Russell Hastings during these years. The record description explains that the photographs contain images of artifacts and pit houses, but "the photographs are poorly documented and do not list provenience of the excavations." Examination of the image files confirms that each photograph includes only basic archival information (classification and numbering system used by the institution holding this set of photonegatives), and no documentation of archaeological site provenience other than that they were taken at Casa Grande during the 1933/1934 excavations (1933 (tDAR ID: 372679) ; doi:10.6067/XCV82N50Z6). Although this limits the usefulness of individual photographs for reinterpretation and contextual inferences, these images have been digitally preserved and made publicly available, making it possible that contextual information might be discoverable in the future (through comparison to other documentation or reports generated by the project, for example), or that they could find utility in other ways (such as documentation of the excavation practices used by this project, comparison of features to other similar excavation projects, etc.). Also, inclusion in tDAR has the added benefit of supplying more descriptive information than was contained in the original archival photograph files and documentation, since the tDAR record contains information such as the culture (Hohokam) and temporal keywords (Sedentary through Classic Period). This in turn facilitates connections with other relevant collections in tDAR. For example, clicking on the "Sedentary through Classic Period" or "Hohokam" links in the record leads to a list of other projects with the same subject headings.

tDAR is also more inclusive in that they will accept deposits of documents (reports, white papers, etc.) and media files without associated primary datasets, while document and media files

in Open Context must be associated with a deposited dataset (AAI 2015c). tDAR's stated goal is to provide a means for perpetual preservation and access of archaeological data, and its structure and foundations support this. This may also make it attractive to contract archaeology firms, museums and archives who may prioritize guaranteed storage, access and preservation of their work, over peer review and editorial approval. However, if a researcher would benefit from publishing their data with some measure or certification of quality such as a peer review, particularly those who hope to include data publication output for purposes of promotion and tenure in academic research, Open Context has the procedures in place to accommodate this.

Current Holdings Size/ # of Contributors. Due to the differences in data record organization and granularity (as described earlier), the total number of Open Context records are not equivalent to the total number of tDAR records. This makes it difficult to do direct comparisons of holdings size, number of depositors and depth vs breadth of coverage. For example, in tDAR a theoretical .xls spreadsheet documenting attributes of 500 artifacts would have one dataset record, while the same dataset in Open Context would have 500 records, one for each artifact. Since we cannot easily compare raw numbers of records between tDAR and Open Context, it is more manageable to provide an overview of the relative comparative size and number of contributions in each repository by comparing the number of "projects" and discrete "datasets". The holdings size and number of contributors are important for predicting sustainability, usefulness of database for data discovery and reuse.

tDAR: As of January 13, 2015, tDAR had 774 datasets (excluding "citation only" records, which have descriptive information but do not include the dataset content held elsewhere outside of tDAR). Of those datasets, 200 were "restricted access" (meaning dataset content cannot be

downloaded or viewed unless specific permissions are granted or until after embargo, although the tDAR record with descriptive information about the dataset and project is visible), while 574 were "publicly accessible files." tDAR had 741 projects, though in tDAR not all projects will contain datasets in a way that can be compared equivalently to Open Context projects (some contain only documents such as a report, map images, or a 3D modeling file, etc.), and multiple datasets could be associated with one project. These numbers were determined using the number of dataset and project records, as displayed by tDAR's search and browsing screens. It was not possible to limit search results to display only the number of project records which also had associated datasets.

Open Context: As of January 13, 2015, Open Context had 54 projects, and by definition all of these projects were comprised of records derived from one or more associated datasets. It was not possible to easily determine how many dataset files (as would be equivalent to tDAR's datasets records) were contributed and processed in each project.

Even with the limitations of inequivalent comparisons of records, unless the projects in Open Context averaged more than 10 datasets per project, it seems likely that tDAR contained more data resulting from a greater number of independent contributors and discrete archaeological research projects. It is not possible to draw conclusive reasons for this difference without further study of archaeological researcher attitudes and repository selection behaviors. Additional data about comparable rates of growth of each repository over time might also be a better predictor of the implications for sustainability, and the potential for exponential increase in utility and buy-in by the disciplinary community of researchers. It also bears mentioning that while 100% of dataset contents in Open Context are publicly visible, approximately 25% of the tDAR datasets counted on January 13, 2015 were restricted (contents cannot be viewed or downloaded by most

people), so while tDAR may contain more data overall, a significant amount of the data is not publicly available to view.

Both Open Context and tDAR have partnered with national and state agencies that conduct archaeological research, to facilitate bulk deposits of data and data records from an entire agency, in addition to deposits from individual academic researchers. tDAR has imported records from the U.S. Department of the Interior National Park Service's National Archaeological Database (NADB) (Ellison 2011), although a quick browsing session of NADB records in January 2015 revealed that many are empty, "citation only" records. For those records, the actual data and reports described are not attached or deposited in tDAR as digital files, but are housed elsewhere. Open Context has partnered with the National Association of State Archaeologists in an NSF-funded initiative to create a Digital Index of North American Archaeology (DINAA), and is in the process of adding data from many state agency archaeological records, including data from Kentucky, Illinois, Iowa, Indiana, Georgia, Florida, South Carolina, and Missouri (AAI 2015d; "Digital Index of North American Archaeology (DINAA)" section). Both of these initiatives add diversity to each database's holdings and can facilitate better discoverability of and connections between different types of archaeological research, and also may improve the likelihood that these stakeholder agencies will take steps in the future to advocate for continued support and growth of the repository in which they have entrusted their data.

Geographic Coverage. Depositing archaeological data in a repository that contains a large amount of data from the same geographic region (or another region with relevant cultural significance to the project) is beneficial for discovery and reuse of that data, since researchers may be more likely to turn to that repository for data about that region, and will be more likely to

find significant connections and relationships between data from projects conducted in similar regions. However, geographic coverage in a repository actively accepting new data submissions is a moving target, since every new deposit adds to the repository's holdings for that region, and a repository that has little data from a given region today may have a much larger amount of data from that region tomorrow. Therefore the following comparisons of geographic coverage are presented in generalized terms (top ten countries of archaeological data origin with the most holdings in each repository, ranked in descending order by the number of records) based on observations of repository data holdings in early January 2015. This ranking is based on the reported number of "records" associated with each country in the repository, relative to the number of records in that same repository for other countries. It should be noted that this is not the same as the number of discreet projects, dig sites, depositors, etc., since in both repositories one project collection may have multiple associated records, depending on the number of associated files.

tDAR holdings by country, ranked in descending order of quantity

1. United States of America
2. Mexico
3. Canada
4. Iceland
5. Tanzania
6. United Kingdom
7. Greece
8. France & Iran (tied)

9. Italy

10. Iraq

Source: Digital Antiquity 2015b: http://core.tdar.org/browse/explore, accessed January 6, 2015

Open Context holdings by country, ranked in descending order of quantity

1. United States of America

2. Turkey

3. Jordan

4. Iran

5. Germany

6. Italy

7. Israel

8. Cyprus

9. United Kingdom

10. Palestinian Authority

Source: AAI 2015a: http://opencontext.org/sets/, accessed January 6, 2015

As time passes, researchers will need to re-evaluate data holdings coverage. Researchers may wish to look for geographic coverage at a more specific level- at the state/province/regional level for example. There may also be some archaeological projects for which other kinds of coverage (types of artifacts, excavation methods used, time period, etc.) will be more important than geographic coverage, and the researcher should evaluate whichever attribute will have the most relevance and greatest impact on the future discovery and use of their data.

For UK archaeology site data, the ADS is recommended. Data from sites with relevance to UK archaeology (for example, data from UK colonial occupations or trade routes in other regions) could be beneficial to put in tDAR, since tDAR and ADS work together and have created the Transatlantic Archaeology Gateway project to allow for cross-searching of their records (ADS 2015). Otherwise, look for a repository with large holdings of data from similar geography and cultural coverage (or other relevant attributes) for optimal discovery and comparison. Currently both repositories have significant North American holdings, with tDAR edging out Open Context in the overall number of North American projects and diversity of holdings originating from this region, despite the large number of tDAR records from the NPS database which describe NPS data but do not actually include the data itself.

Many of Open Context's North American project collections are from their DINAA project arrangement with participating state agencies, so data from other excavations in these states may benefit from deposit in Open Context. Many DINAA records currently contain site descriptions and partial data but not the full original source data, though some of these site files were marked as forthcoming on the Open Context "Projects" page (AAI 2015m), and may be expanded upon in the future. Open Context has particularly strong coverage of the Middle East (especially Mesopotamian zooarchaeology data), with several different types of datasets and excavation projects from Turkey, Jordan, and Iran.

Summary and Conclusions

Both tDAR and Open Context share qualities that make them attractive solutions for archaeological data sharing and publication. They are both gaining visibility in the disciplinary research community, particularly in the United States. They have implemented funding models which set the stage for diversified and long-term financial support, provided that stakeholder

researchers are willing and able to deposit data, and that the repositories continue to maintain the strong partnerships with archaeological state agencies and professional organizations that they have begun to establish. Both repositories can provide public access to the research data contents, and assign persistent identifiers to provide a means of consistently tracking citations and locating the original data used in future publications. In addition to facilitating public sharing of the products of archaeological research, both repositories have taken steps to facilitate the preservation of this data. Both provide the ability to track views and downloads of the data, giving the researcher evidence of the impact of their work. Additionally, since both repositories were designed specifically with archaeological data in mind, they provide support for needs specific to archaeological research, such as the masking of site location information, as well as descriptive properties and controlled archaeological vocabulary terms which provide the context necessary to explore and interpret the data.

Despite these similarities, tDAR and Open Context differ in several ways, and these differences must be compared to the specific parameters and attributes of an individual research project and its data, and the goals or intentions of the researcher which are driving the decision to deposit and share the data, to determine which repository is the optimal choice.

Attributes unique to tDAR which may be seen as positive deciding factors for some projects are:

- Ability to set access restrictions (embargoes or restrictions to specific individuals) for the viewing and downloading of data
- Support for specialized data types such as remote sensing, 3D scans and GIS data
- Ability to view and download data in original file format

- Partnership with the Society for American Archaeology (including discounts for SAA student members)
- Ability to assign licenses more restrictive than the Creative Commons Attribution default if needed
- Larger total number of holdings and diversity of geographic coverage (with primary strengths in North America and Western Europe)
- Connections with the UK's ADS including the Trans-Atlantic Gateway Project
- Support for ontological integration and comparison of similar datasets within tDAR.

Attributes unique to Open Context which may be seen as positive deciding factors for some projects are:

- Peer review and editorial control for quality of content, akin to academic publishing models
- Contents of all deposited data are open and publicly available
- Search, browse and view descriptive and contextual data at the artifact level from multiple different datasets and projects simultaneously, within the Open Context interface
- Strong geographic coverage in the Middle East and increasingly greater coverage of North America through DINAA state agency deposits
- Standardization of archaeological data description and integration of other relevant standards using Linked Data.

Further study of the choices and opinions of researchers and institutions who have deposited archaeological data in tDAR and/or Open Context (or who have chosen not to deposit in either) will provide additional insights about the perceived strengths and weaknesses of each, and the factors which have had the greatest influence on the choice of a data repository/publication platform. In addition, since both repositories differ greatly in their interfaces, search/browse retrieval designs, approach to data description and integration tools, another aspect of these repositories well suited for additional research is the perspective of users who have used tDAR and/or Open Context to search for, access and reuse data, as their experiences will illuminate the features which either facilitate or hinder archaeological data discovery and reuse.

Librarians can use the charts and lists of features provided in this article to help them ask the right questions about data projects and make informed recommendations to archaeology researchers, in order to optimize the discoverability, usefulness and preservation of their data. Knowledge of these repositories and their holdings is also essential when helping archaeology researchers locate existing data sets for reanalysis.

The usefulness of the disciplinary data repository evaluation and comparison process illustrated in this article is not limited to the field of archaeology. Subject specialist librarians in many fields can use this comparative framework to become familiar with the major strengths and shortcomings of the existing disciplinary data repository options available to the researchers they advise. Awareness of disciplinary data repositories, and the ability to compare them and explain them to researchers, should be considered an integral component of a subject specialist librarian's skill set and knowledge base.

References Cited

1933. “1933 and 1934 Casa Grande Ruins Excavation Photographs.” The Digital Archaeological Record (tDAR), accessed February 3, 2015, http://core.tdar.org/image/372679/1933-and-1934-casa-grande-ruins-excavation-photographs, doi:10.6067/XCV82N50Z6.

AAI (Alexandria Archive Institute). 2015a. “Browse [Results list].” Open Context website. Accessed January 6, 2015. http://opencontext.org/sets/.

——. 2015b. “Data Management, Publishing, & Budgeting.” Open Context website. Accessed June 23, 2015. http://opencontext.org/about/estimate.

——. 2015c. “Data Publication Guidelines for Contributors.” Open Context website. Accessed July 7, 2015. http://opencontext.org/about/publishing.

——. 2015d. *Heritage Bytes.* Accessed March 17, 2015. http://ux.opencontext.org/blog/.

——. 2015e. “History.” Alexandria Archive Institute website. Accessed June 23, 2015. http://alexandriaarchive.org/about/history/.

——. 2015f. “Impacts: Expanding Participation and Reach in Research.” Alexandria Archive Institute website. Accessed May 5, 2015. http://alexandriaarchive.org/impacts/.

——. 2015g “Open Context API/Web-Services Documentation.” Open Context website. Accessed July 20, 2015. http://opencontext.org/about/services.

——. 2015h. “Open Context Sponsors, Funders, and Collaborators.” Open Context website. Accessed June 23, 2015. http://opencontext.org/about/sponsors.

——. 2015i. “Open Context’s Applications and Uses.” Open Context website. Accessed June 23, 2015. http://opencontext.org/about/uses/.

——. 2015j. “Open Context’s Privacy Policy” Open Context website. Accessed July 20, 2015. http://opencontext.org/about/privacy.

——. 2015k. “Open Context’s Technologies.” Open Context website. Accessed July 20, 2015. http://opencontext.org/about/technology.

——. 2015l. “People with Open Context” Alexandria Archive Institute website. Accessed June 30, 2015. http://alexandriaarchive.org/about/people/.

——. 2015m “Projects and Collections.” Open Context website. Accessed January 6, 2015. http://opencontext.org/projects/.

ADS (Archaeology Data Service). 2007. *Charging Policy*, version 4. Accessed July 9, 2015. http://archaeologydataservice.ac.uk/advice/chargingPolicy.

—. 2014. “2.1.1 Geographic Range.” *Collections Policy*, 6th edition. Accessed June 16, 2015. http://archaeologydataservice.ac.uk/advice/collectionsPolicy.

—. 2015. “Welcome to the TAG portal.” Transatlantic Archaeology Gateway website. Accessed July 20, 2015. http://archaeologydataservice.ac.uk/TAG/intro.jsf.

ADS (Archaeology Data Service) and Digital Antiquity. 2015. *Archaeology Data Service/Digital Antiquity Guides to Good Practice*. http://guides.archaeologydataservice.ac.uk/. Accessed July 9, 2015.

Aitchison, Kenneth. 2009. "Standards and Guidance in Archaeological Archiving: The Work of the Archaeological Archives Forum and the Institute for Archaeologists." *Grey Journal (TGJ)* 5 (2): 67-71. Accessed June 9, 2015, http://search.ebscohost.com/login.aspx?direct=true&db=lls&AN=42849017&site=ehost-live.

Alleen-Willems, Russell. 2012. "Designing the Digital Archaeological Record: Collecting, Preserving, and Sharing Archaeological Information." Order No. 1511231, M.A., Northern Arizona University. http://search.proquest.com/docview/1020616419?accountid=14553

Atici, Levent, Sarah Whitcher Kansa, Justin Lev-Tov, and Eric C. Kansa. 2013. Other People's Data: A Demonstration of the Imperative of Publishing Primary Data. *Journal of Archaeological Method and Theory* 20(4):663-81.

Brin, Adam. 2013a. “Creating or Editing a Project.” Creating & Editing Resources in tDAR section, tDAR Documentation Wiki. Last Modified January 21, 2013. https://docs.tdar.org/display/TDAR/Creating+or+Editing+a+Project.

——. 2013b. “Society for American Archaeology (SAA) and the Center for Digital Antiquity (Digital Antiquity) Announce Collaborative Program.” tDAR News. Accessed June 23, 2015. https://www.tdar.org/news/2013/06/society-for-american-archaeology-saa-and-the-center-for-digital-antiquity-digital-antiquity-announce-collaborative-program/.

——. 2014. “Working with 3D Sensory Objects.” tDAR Documentation Wiki. Last Modified August 19, 2014. https://docs.tdar.org/display/TDAR/Working+With+3D+Sensory+Data+Objects.

——. 2015a. “Taking a Look Back at 2014.” tDAR News. Accessed July 2, 2015. http://www.tdar.org/news/2015/01/taking-a-look-back-at-2014/.

——. 2015b. “tDAR Software Update (Munsell).” tDAR News. Accessed June 30, 2015. http://www.tdar.org/news/2015/01/tdar-software-update-munsell/.

Bobley, Brett. 2013. “NEH grantees honored as White House Open Science Champions of Change.” NEH Office of Digital Humanities Featured Project website, accessed May 7, 2015,

http://www.neh.gov/divisions/odh/featured-project/neh-grantees-honored-white-house-open-science-champions-change.

CCSDS (Consultative Committee for Space Data Systems). 2011. "Audit and Certification of Trustworthy Digital Repositories." CCSDS. Accessed May 12, 2015, http://public.ccsds.org/publications/archive/652x0m1.pdf.

Cleland-Sipfle, K. 2013. "The Digital Archaeological Record (tDAR)." *Choice: Current Reviews For Academic Libraries 51*(2):234.

Creative Commons. 2015. "About the Licenses." Creative Commons website. Accessed July 15, 2015. http://creativecommons.org/licenses/.

CRL (Center for Research Libraries). 2015. "ISO 16363 / TDR." Center for Research Libraries. Accessed June 18, 2015. http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/iso16363.

deVos, Jim. 2012. "Test Plan – DOI Values." tDAR Functional Testing Plan section, tDAR Developer Documentation Wiki. Last modified January 27, 2012. https://dev.tdar.org/confluence/display/DEV/Test+Plan+-+DOI+values.

Digital Antiquity. 2013a. "Accession Policy." tDAR website. Accessed July 17, 2015. http://www.tdar.org/about/policies/accession-policy/.

——. 2013b. "Contributor's Agreement". tDAR website. Accessed March 17, 2015. https://www.tdar.org/about/policies/contributors-agreement/.

——. 2013c. "Preservation." tDAR website. Accessed July 20, 2015. http://www.tdar.org/why-tdar/preservation/.

——. 2013d. "Preservation and Curation Policy." tDAR website. Accessed July 17, 2015. http://www.tdar.org/about/policies/preservation-and-curation-policy/.

——. 2013e. "Terms of Use". tDAR website. Accessed March 17, 2015. http://www.tdar.org/about/policies/terms-of-use/.

——. 2013f. "Upload & Contribute to tDAR." tDAR website. Accessed July 17, 2015. http://www.tdar.org/why-tdar/contribute/.

——. 2015a. "About." Digital Antiquity website. Accessed March 17, 2015. http://www.digitalantiquity.org/about/.

——. 2015b. "Explore tDAR." tDAR website. Accessed January 6, 2015. http://core.tdar.org/browse/explore.

——. 2015c. "News." tDAR website. Accessed March 17, 2015. http://www.tdar.org/news/

——. 2015d. “Pricing.” tDAR website. Accessed July 17, 2015. https://core.tdar.org/cart/add.

Eiteljorg, Harrison. 2002. “The Archaeological Data Archive Project Ceases Operation.” *CSA Newsletter* 15 (2). Accessed October 20, 2014, http://csanet.org/newsletter/fall02/nlf0201.html.

Ellison, Leigh Anne. 2011. “National Archeological Database Added to tDAR.” tDAR News. Accessed July 17, 2015. https://www.tdar.org/news/2011/07/national-archaeological-database-added-to-tdar/.

——. 2012. “Agreement to sustain tDAR’s Digital Records.” tDAR News. Accessed June 23, 2015. https://www.tdar.org/news/2012/10/agreement-to-sustain-tdars-digital-records/.

——. 2015. “Digital Antiquity Announces per File Price Reductions for Digital Preservation in tDAR!” tDAR News. Accessed July 20, 2015. http://www.tdar.org/news/2015/07/digital-antiquity-announces-per-file-price-reductions-for-digital-preservation-in-tdar/.

Faniel, Ixchel, Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel. 2013. “The challenges of digging data: A study of context in archaeological data reuse.” In *JCDL ’13 Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 295-304. New York: ACM. DOI: 10.1145/2467696.2467712

Figshare. 2015. “Figshare for Publishers.” Figshare website. Accessed June 18, 2015. http://figshare.com/services/publishers

Gabridge, Tracy. 2009. "The Last Mile: Liaison Roles in Curating Science and Engineering Research Data." *Research Library Issues: A Bimonthly Report from ARL, CNI, AND SPARC*. 265: 15-21. http://publications.arl.org/rli265/16.

Hahn, Robert A. 2012. “Resource Development Committee Launches the Anthropology Portal.” *Anthropology News* 53: 24.

International Council of Museums. 2014. “What is the CIDOC CRM?” The CIDOC Conceptual Reference Model website. Last updated May 12, 2014. http://www.cidoc-crm.org/.

Jacobs, Paul, and Christopher Holland. 2007. “Sharing archaeological data: The distributed archive method.” *Near Eastern Archaeology* 70 (4): 197-200.

Kansa, Eric C. 2012. “Openness and archaeology's information ecosystem.” *World Archaeology* 44(4):498-520.

Kansa, Eric C., Sarah Whitcher Kansa, and Benjamin Arbuckle. 2014. “Publishing and Pushing: Mixed Models for Communicating Research Data in Archaeology.” *International Journal of Digital Curation* 9(1):57-70. DOI: 10.2218/ijdc.v9i1.301

Kansa, Eric C., Sarah Whitcher Kansa, Margie M. Burton, and Cindy Stankowski. 2010. "Googling the Grey: Open Data, Web Services, and Semantics." *Archaeologies: Journal of the World Archaeological Congress* 6(2):301-26.

Kansa, Sarah Whitcher, and Eric C. Kansa. 2007. "Open Content in Open Context." *Educational Technology* 47(6):26-31.

Kansa, Sarah Whitcher, Eric C. Kansa, and Jason M. Schultz. 2007. "An Open Context for 'Near Eastern Archaeology'." *Near Eastern Archaeology* 70(4):188-94.

Khosrow-Pour, Mehdi. 2013. "Ontology." In *Dictionary of Information Science and Technology,* 2nd edition. doi:10.4018/978-1-4666-2624-9.

Kintigh, Keith W. and Jeffrey H. Altschul. 2010. "Sustaining the Digital Archaeological Record." *Heritage Management* 3(2):264-74.

Marcial, Laura Haak and Bradley M. Hemminger. 2010. "Scientific Data Repositories on the Web: An Initial Survey." *Journal of the American Society for Information Science & Technology* 61 (10): 2029-48.

McManamon, Francis P. and Keith W. Kintigh. 2010. "Digital Antiquity: Transforming Archaeological Knowledge into Data." *SAA Archaeological Record* 10 (2):37-40.

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 23(2). Accessed October 20, 2014.
http://csanet.org/newsletter/fall10/nlf1002.html.

NSF (National Science Foundation). 2010. "Data Management for NSF SBE Directorate Proposals and Awards." Accessed November 12, 2013.
http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf.

OCLC (Online Computer Library Center) and CRL (Center for Research Libraries). 2007. "Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)." OCLC and CRL. Accessed July 22, 2014.
http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf.

SAA (Society for American Archaeology). 1996. *Principles of Archaeological Ethics*. SAA website. Accessed July 20, 2015.
http://www.saa.org/AbouttheSociety/PrinciplesofArchaeologicalEthics/tabid/203/Default.aspx.

——. 2015. "Benefits of Joining SAA Today." SAA website. Accessed July 9, 2015.
http://www.saa.org/AbouttheSociety/Membership/MembershipBenefits/tabid/159/Default.aspx.

Scholtz, Sandra and Robert G. Chenhall. 1976. “Archaeological Data Banks in Theory and Practice.” *American Antiquity* 41 (1): 89-96.

Seely, Amber. 2005. "Digging Up Archaeological Information." *Behavioral & Social Sciences Librarian* 24 (1): 1-20. http://dx.doi.org/10.1300/J103v24n01_01

tDAR (The Digital Archaeological Record). 2013. tDAR Documentation Wiki website. Accessed July 9, 2015. https://docs.tdar.org//display/TDAR/Documentation+Home.

Ubiquity Press. 2015. “About.” *Journal of Open Archaeology Data* website. Accessed June 30, 2015. http://openarchaeologydata.metajnl.com/about/.

University of California Digital Curation Center. 2015. “DMP Requirements.” DMPTool. Accessed June 9. https://dmptool.org/guidance.

University of York. 2015a. “About Us.” Archaeology Data Service website. Accessed June 16, 2015. http://archaeologydataservice.ac.uk/about.

——. 2015b. “About Us: Funders.” Archaeology Data Service website. Accessed June 30, 2015. http://archaeologydataservice.ac.uk/about/funders.

WAC (World Archaeological Congress). 2015. “Codes of Ethics.” World Archaeological Congress website. Last updated March 3, 2015. http://www.worldarchaeologicalcongress.org/about-wac/codes-of-ethics.

Willett, Perry and John Kunze. 2015. “ARK.” Microservices section, University of California Curation Center: Curation Wiki. Last modified March 11, 2015. https://wiki.ucop.edu/display/Curation/ARK.