



This PDF is a simplified version of the original article published in Internet Archaeology under the terms of the Creative Commons Attribution 3.0 (CC BY) Unported licence. Enlarged images, models, visualisations etc which support this publication can be found in the original version online. All links also go to the online original.

Please cite this as: Kansa, E. and Kansa, S. Whitcher 2026 Open Context in a Changing Context: Data Publishing, Interoperability and Governance, *Internet Archaeology* 71. <https://doi.org/10.11141/ia.71.3>

Summary

This paper describes the rationale and strategies behind [Open Context](#), an online data publishing service for archaeology. Launched in 2006, Open Context has published nearly 2.4 million records from 200 projects and sub-projects, representing the work of more than 1,600 scholars working across the world. Open Context's current data dissemination and curation efforts reflect our attempts to navigate a landscape filled with diverse and evolving challenges. Responses to professional, institutional, and sustainability issues, some of a global character and some of a United States specific character, have helped shape Open Context's technological and semantic choices. Though focused on United States specific experiences, we hope that this in-depth discussion of how Open Context responded to these multifaceted concerns will help advance international community wide conversations about information architectures, governance, and social factors in archaeological informatics.

1. Introduction

This paper describes the rationale and strategies behind [Open Context](#), an online data publishing service for archaeology. Launched in 2006, Open Context has published nearly 2.4 million records from over 200 projects and sub-projects, representing the work of more than 1,600 scholars working across the world. Open Context's current data dissemination and curation efforts reflect our attempts to navigate a landscape filled with diverse and evolving challenges. Responses to professional, institutional, and sustainability issues, some of a global character and some of a United States specific character, have helped shape Open Context's technological and semantic choices. Though focused on United States specific experiences, we hope that this in-depth discussion of how Open Context responded to these multifaceted concerns will help advance international community wide conversations about information architectures, governance, and social factors in archaeological informatics.

2. Background and Institutional Context

Until the political upheavals of 2025, including the Trump administration's terminations and severe cuts to many federal agencies, the United States invested over \$1 billion per year to comply with historical and archaeological protection measures required by federal law (Altschul and Klein [2022](#)). This level of public investment by itself far exceeds the total combined 2024 budgets of the Institute of Museum and Library Services (IMLS \$294.8 million), the National Endowment for the Humanities (NEH \$20 million), and the National Science Foundation's (NSF) Social, Behavioral, and Economic Sciences program that includes archaeology (roughly \$290 million). Since the mid-1960s, CRM (cultural resource management) activities have recorded more than one million archaeological sites, conducted more than one million field studies, excavated over 100,000 sites, and curated more than one billion artifacts and associated records (Ortman and Altschul [2023](#)).

Even before the 2025 crisis, the sustained treatment of archaeological and other cultural heritage data has been a topic of frequent discussion (Kansa [2009](#), [2016](#); Kansa *et al.* [2005](#); Kansa *et al.* [2018](#); Kintigh [2006](#); McManamon and Kintigh [2010](#)) that has resulted in new policies and revised ethical



statements by professional or trade organizations such as the American Cultural Resources Association (ACRA [2019](#)) and a Register of Professional Archaeologists (RPA) effort in progress. Moreover, as archaeological field research has become more expensive and subject to political and other restrictions, there is an increasing demand for the reuse of existing data to inform research, instruction, and conservation policies. Coupled with these internal pressures for data reuse are closely related external pressures to increase accountability and transparency in research, notably in open access (e.g. The White House [2013](#); G8 Open Data Charter ([2013](#)); OSTP [2014](#); OMB [2015](#)); research replicability through journal requirements for provision of supporting data (Nature [2013](#); Vines *et al.* [2013](#)); and effective data management that leads to effective data discovery, access, reuse, and long-term preservation (e.g. NSF and NEH Data Management Plans).

Investments in archaeological data curation and preservation served multiple purposes. Archaeological sites and landscapes globally face tremendous conservation pressures from development, the impacts of climate change, economic factors that lead to neglect and looting, vandalism, as well as intentional or unintentional damage from armed conflict. Curating data about archaeological resources plays an important role in conservation planning and mitigation efforts. Some archaeological research methods, especially excavation, are destructive, so the curation of both excavated materials and excavation records directly serves conservation. Secondly, archaeological data curation can and does open new research opportunities. The [Coalition for Archaeological Synthesis](#) (CfAS) sponsors several high-profile data driven research initiatives. Data published by Open Context itself has led to significant research outcomes that still continue in zooarchaeology (Arbuckle *et al.* [2014](#); Kansa *et al.* [2014](#)) and modeling the impacts of climate change driven sea level rise on coastal cultural heritage (Anderson *et al.* [2017](#)). Open Context's emphasis on curatorial investments in data are intended to promote reuse (see Yoon and Lee [2019](#)).

Altogether, there is a clear and urgent need to dedicate a fraction of the public investment in cultural heritage research and conservation to data management. Internationally, archaeological data management policies and practices vary greatly (Geser *et al.* [2022](#)). Where archaeological data management services exist, they typically have university related or public-sector (government) organizational homes. For example, the [Archaeology Data Service](#) (ADS), which serves as the leading archaeological data repository in the United Kingdom, is a not-for-profit organization at the University of York. [tDAR](#) (the Digital Archaeological Record), the leading archaeological data repository in the United States has a similar organizational model. It is managed by a nonprofit organization, the [Center for Digital Antiquity](#), housed at Arizona State University. In France, the national government manages archaeological data in several systems reviewed by Marx, Rossenbach and Bryas ([2021](#)). The Netherlands, Sweden, Israel, and other nations also have archaeological data management systems largely provisioned through government agencies (Geser *et al.* [2022](#)).

Digital preservation services have somewhat large capital and organizational requirements. They require computing facilities, expert staff, and institutional ties with funding agencies and other authorities that oversee archaeological activities (see Novák *et al.* [2023](#) for a review of the European context). Repositories also require institutional stability and permanence, which in turn require sustained access to financial resources. Given these requirements, it's not surprising that most archaeological data repositories grow out of already well-established university or government agency organizational settings. Unfortunately, the recent political upheavals of the United States illustrate potential vulnerabilities for digital repositories operating in such institutional settings. We will return to this point below.

The high barriers to entry in founding and running a digital repository mean that, in the absence of continued public funding, repositories often have to charge for their services. Because of cost and



other barriers, many practitioners in the archaeological community do not deposit to dedicated archaeological repository services, though some archaeologists use more general purpose repositories with lower deposit fees, such as Zenodo or even commercial services such as FigShare or Sketchfab (the sale of Sketchfab has triggered considerable concern by many depositors of cultural heritage 3D models) While cheap and accessible, unfortunately, general purpose repositories such as Zenodo lack curatorial services to assist in the creation of metadata, especially metadata specific to the domain of archaeology. Lack of domain-specific metadata impedes research because relevant resources, though archived, are difficult to discover and can have sparse documentation.

Open Context has a somewhat contrasting and peculiar institutional context. Open Context was launched in 2006 by two (then) "early career researchers" (Sarah Witcher Kansa and Eric Kansa), who also co-founded the [Alexandria Archive Institute](#) (AAI), a not-for-profit organization based in California, USA. The AAI continues to develop and manage Open Context, running it largely independently of universities (with the notable exception of some university library services), in contrast to most other archaeological data infrastructure.

Open Context's different institutional circumstances and history play a fundamental role in shaping the services that it provides. Because Open Context is not integrated with authorities that oversee archaeological work, it has little leverage to influence funders or other authorities to mandate repository data deposition or other policy requirements. Most importantly, because Open Context is run by a small institution, it lacks the organizational capacity to offer data archiving services independently. With access only to short term grants and fees for professional services (see below), Open Context focuses on data dissemination services and prioritizes ensuring long-term data access. While its services facilitate data archiving, Open Context relies upon other organizations to provide long term preservation services. We elaborate on this approach in the next section.

3. Data Curation and Externally Provisioned Preservation Services

Archaeological data management involves a much broader set of concerns than digital repository services. The Open Context editorial team dedicates time and effort to improve data quality and interoperability with workflows to review, edit, clean, annotate and integrate datasets provided by contributing researchers. Data published through these workflows becomes integrated and dynamically accessible and searchable on Open Context's Web-based platform. To meet long term data preservation needs, Open Context archives data using more general purpose digital repositories (chiefly, the California Digital Library's [Merritt](#) repository and [Zenodo](#)), essentially to avoid putting all eggs in one basket.

Recent information management literature emphasizes the need for expert curatorial investment to enable and promote data reuse (Marsolek *et al.* [2023](#); Hemphill *et al.* [2022](#)). To meet these needs, Open Context's "high touch" approach toward data publication attempts to provide domain expertise to data curation in archaeology. At the same time, while Open Context provides domain expertise for data curation, its use of general-purpose repository services helps it to avoid the higher capital costs associated with launching and maintaining a dedicated digital preservation repository.

This helps fill important institutional gaps. Many university libraries and information technology services offer general support for research data management, especially regarding grant data management compliance (Lee and Stvilia [2017](#)). However, familiarity with domain-specific vocabularies, gazetteers, and data modeling concerns, as well as professional engagement with ongoing archaeological informatics research, requires dedicated and specialized experts. As noted by



Yakel and colleagues (2019), dedicated, domain-specific data curation services promote greater data reuse. However, most university libraries and general-purpose repositories lack the staffing to provide specialized data curation services to best meet researcher needs (Faniel *et al.* 2018; Johnston *et al.* 2018).

Open Context has pioneered "data sharing as publication" in archaeology (Kansa 2012; see Muñoz 2013 regarding this model in the broader digital humanities). This approach seeks to motivate archaeological engagement with ethically appropriate FAIR data (Findable, Accessible, Interoperable, Reusable; Wilkinson *et al.* 2016) practices (summarized in Table 2). The archaeological community needs to invest in approaches that incentivize such engagement (Bruce & Cordewener 2018). If digital data remain tangential to the archaeological profession, the community will have little understanding of how to navigate the risks and opportunities afforded by data.

As introduced above, the Open Context team focuses on curation and access issues and depends upon other institutions to meet data preservation needs. We believe that this "separation of concerns" can be a useful strategy that other teams can and already do adopt. For example, the "Endangered Archaeology in the Middle East and North Africa" (EAMENA) project focuses efforts on data curation, modeling, and capacity building, but it also uses external digital repository services to meet longer term data citation and data preservation needs. Similarly, the Pleiades gazetteer team focuses on building and curating high quality data about ancient places, and it too uses external services for data preservation.

Thus, Open Context's separation of concerns *vis-à-vis* data archiving is practiced by multiple online information systems created and used by archaeologists. A small and modestly financed team can focus their efforts on specialized data curation while relying upon externally provisioned data archiving services to meet preservation needs. Overall, this strategy may help realize some of the advantages of specialized data and metadata curation provided by domain repositories while still benefiting from the cost-savings afforded by generalist repositories for basic preservation needs.

While Open Context and other information systems share this general strategy, the strategy is not without limitations and drawbacks. Simply put, data curation is never finished. Standards, expectations, and communities continually evolve. There will always be a need to continually revisit, revise and adapt metadata— all to keep pace with changing expectations. Without active metadata maintenance, data stored in a generalist repository may run the risk of curatorial neglect – even if data files remain safe from loss or corruption. Nevertheless, unless it becomes much more financially and organizationally feasible to set up a permanent digital repository, a strategy that combines small-team curated data with externally provisioned repository services will likely remain Open Context's best available option to meet data preservation needs.

4. A Comparative Look at Open Context's Approach to Data Publishing

Open Context provides specialized services to publish structured data and related media on the Web. As such, rather than hosting reports, articles, presentations, and other forms of research communication that are commonly found in many more conventional digital repositories, Open Context focuses on primary-source digital content (databases, photographs, field-notes, 3D models) that archaeologists use to build arguments and interpretations. Thus, Open Context does not replicate the role of book or journal publishers or repositories of digitized literature (JSTOR, the HathiTrust). Instead, Open Context offers services that enrich existing publishing by sharing vast bodies of primary, structured data (Kansa 2016).



Since data curation involves methodological, ethical and theoretical challenges as intellectually rich as any other archaeological research domain, it requires multiple approaches and broad, intellectually diverse engagement. As already discussed, several other online projects including [EAMENA](#), [Pleiades](#), the [Levantine Ceramics Project](#), the [Chaco Research Archive](#) and the [Digital Archaeological Archive of Comparative Slavery](#) (DAACS), among others, organize experts to curate archaeological (and related) data. Each of these other information systems organizes teams of experts to create and curate data into common databases. These databases have their own schemas, defining a common class of entities and their descriptive attributes and relations.

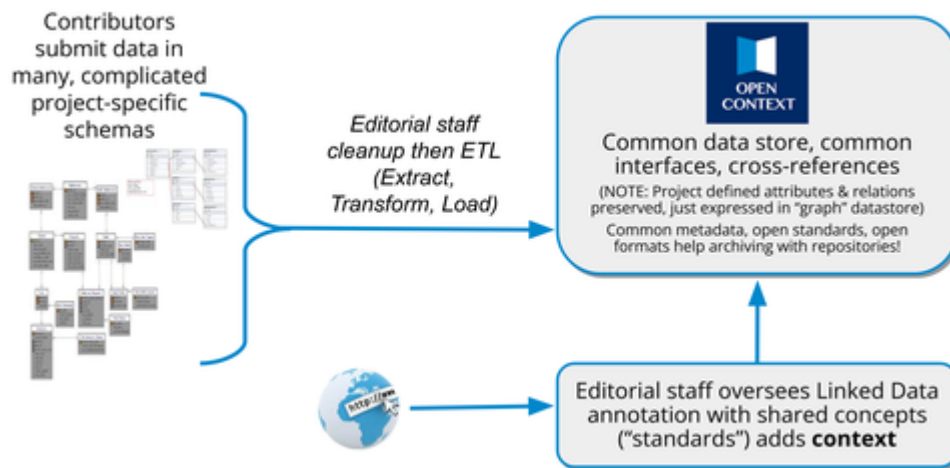


Figure 1: Diagram of Open Context ETL and Standards Annotation Publishing Workflows

4.1 Open Context's ETL Processes and Schema-Free(ish) Abstractions

Routine use of "Extract, Transform, Load" (ETL) processes for data acquisition is a major characteristic that sets Open Context apart from other archaeological data infrastructures. ETL is a commonly used phrase to describe procedures that migrate data from a legacy system for integration into another, larger database. In Open Context's case, data contributors do not upload their own data or add metadata themselves. Rather, data contributors submit their publication-ready datasets via file transfers to the Open Context team. Open Context's staff then maps each dataset (and every single record) to a common, very abstract schema and loads the data into a single integrated database. This database holds millions of individual records and their attributes all originally sourced from many hundreds of individual database tables and spreadsheets.

Open Context also differs from many other information systems in terms of database design. It uses a PostgreSQL relational database managed by a custom Python application built using the open-source Django framework. While these components are widely used, Open Context's implementation involves a high degree of abstraction. Essentially, Open Context implements some aspects of "schema-free" database design in its backend architecture. A schema-free database design allows new data structures (patterns of descriptions and relationships) to be appended to a datastore without requiring that datastore itself to be modified. Each dataset imported into Open Context (via ETL processes) can have its own unique set of descriptive attributes, vocabularies, linking relationships, etc. Open Context needs a highly abstracted schema-free style database structure to represent these unique and dataset-specific attributes. Project defined descriptions and relationships typically include stratigraphic relationships between different excavation units,



descriptions of animal bones, pottery types, daily field note entries, geospatial data, and much more. In this way, Open Context can ingest diverse data into a common datastore while still retaining each dataset's original terminology and other systems of description.

Certain ideals motivated this design decision. If one considers that datasets are intellectual and scholarly works that involve a host of interpretive, theoretical, and practical considerations (see Ryan [2004](#); Hacigüzeller *et al.* [2021](#)), then there can be value in sharing and preserving the diverse ways researchers grapple with these issues. In other words, we see significance and value in the bespoke, contingent, and idiosyncratic ways that different researchers organize and describe their data (see also Batist [2025](#)). While efficient interoperability across datasets can be very useful, *data can have significance beyond aggregation*. In our view, the various ways researchers organize and describe their data has intellectual value, even if maintaining this diversity can complicate interoperability. Open Context supports this diversity by ingesting full datasets without forcing projects to abandon their own terminologies and descriptive systems.

4.2 Published Content and User Interfaces

Open Context's abstract (schema free) design enables search, query, analysis and citation of *each individual record* of data (describing archaeological information on vastly different scales — from whole sites, to contexts, to individual objects). This granularity and integration into a common database allows for publication and public exhibition of archaeological research data dynamically on the Web. Users can search and browse linked records of sites, contexts, objects, ecofacts and media (images, digitized field notes, 3D models). Thus, Open Context provides a uniform interface to display search results and individual records from many different field projects and collections. A common API ("application program interface"; technologies that enable software to exchange data) to all this diverse content makes it easier for others to develop independent software for visualization and analysis. Instructors (see Graham *et al.* [2019](#)) and enthusiasts have used this API for teaching and their own projects (including [augmented reality](#), [virtual reality](#), [text mining](#), [map visualizations](#), and [machine-learning training](#)).

Open Context provides user interfaces for searching, browsing, and review of individual item records in a manner that is very similar to museum catalog websites or other specialized databases for archaeology. The overall user interface design has gone through several revisions, enabled by improvements to faceted search indexing software and evaluation, and user-centered design feedback from masters students at the University of Michigan School of Information. Unlike typical museum catalogs, Open Context often provides much more detailed information about archaeological context. For example, Open Context enables users to navigate from a record of a sherd (originally from a ceramics data table) to a record of a stratigraphic unit (originally from another excavation database) from which the sherd was recovered (see Figure 2).

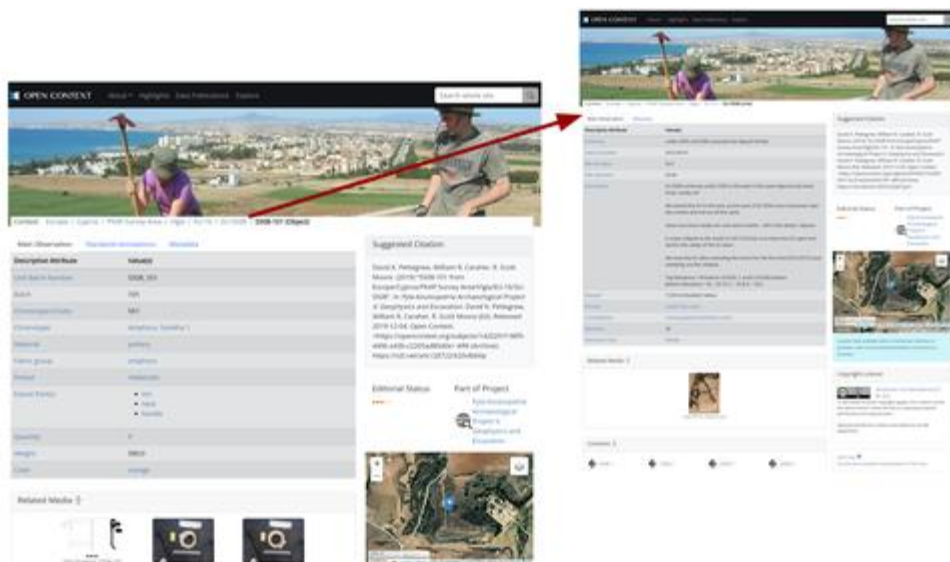


Figure 2: Related Artifact and Stratigraphic Unit Records

Open Context's main search functions combine full-text search (with type-ahead autocompletion) and hierarchical faceted search for querying structured data. Search results are summarized with dynamic map visualizations (see Figure 3a) and dynamic visualizations of the chronological distribution (Figure 3b) of search results. The same search interfaces support information retrieval and visualization at large continental or regional scales and more "micro" scales within individual archaeological sites (Figure 4), provided that the data sources included detailed geospatial data.

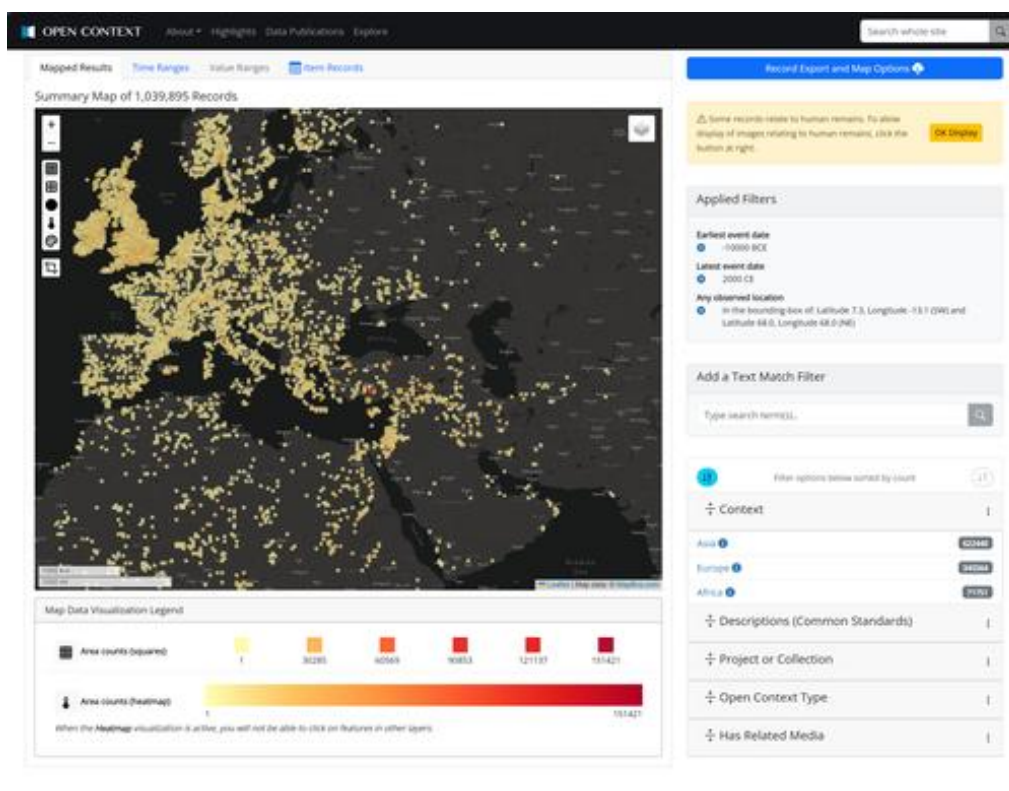


Figure 3a: Example Search Result Visualization at a Large Regional Scale (Map)

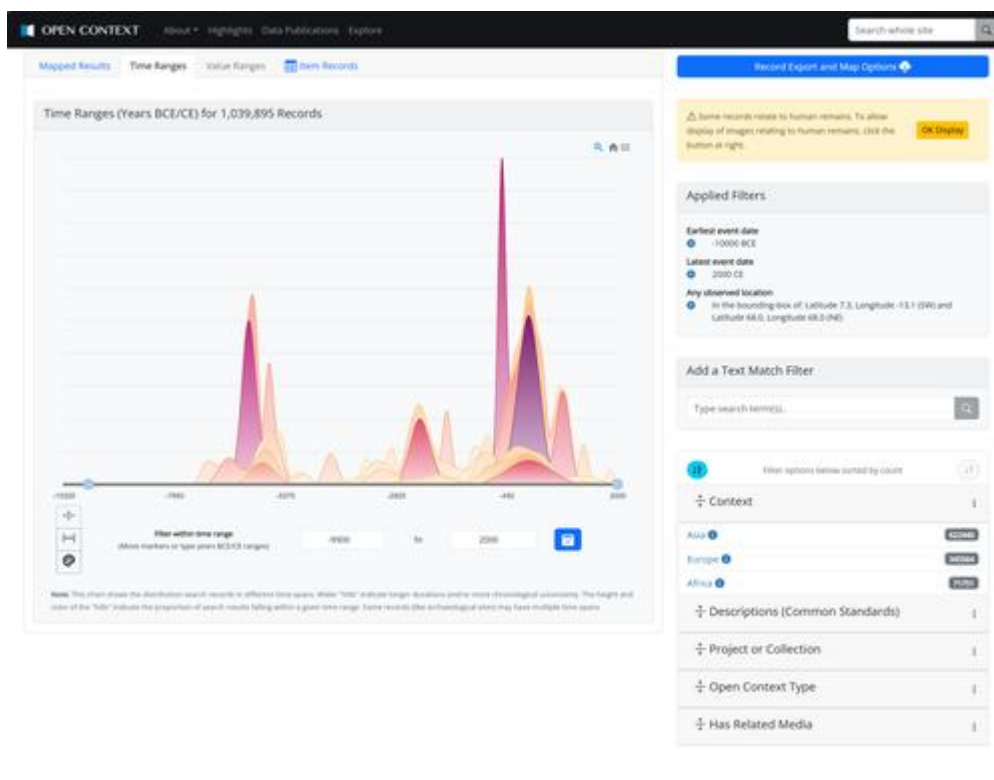


Figure 3b: Example Search Result Visualization at a Large Regional Scale (Chronology)

Open Context's schema-free abstraction provides a common platform for many hundreds of projects, hosting large quantities of convenient and directly accessible curated and integrated archaeological data. Comprehensive indexing, search and visualization features make these data easier for users to explore and retrieve. In reducing access barriers, Open Context eases the time and labor of data reuse. This promotes serendipitous discovery of relevant data, makes exploratory research easier, and can make experimental use of different data analysis techniques less laborious.

The Web interaction patterns and types of user interface features offered by Open Context are fairly routine for cultural heritage and research databases. Open Context's main area of distinctiveness centers on its diversity of content and data sources (see Table 1). Open Context manages a wide variety of cultural heritage data. At the largest (continental) scale, Open Context hosts the [Digital Index of North American Archaeology](#) (DINAA), a dataset of archaeological "site file" records aggregated from several state government offices across the United States. With an even larger geographic scope, Open Context also hosts *Cross-referenced p3k14c*, a global [database](#) of archaeological radiocarbon samples (Bird *et al.* 2024). But Open Context also hosts data at much smaller, archaeological survey and intra-site scales. These examples include *Murlo* (Poggio Civitate) (see Figure 4, Tuck 2012), *Kenan Tepe* (Parker and Cobb 2012), *Madaba Plains Project-`Umayri* (Clark and Herr [in prep](#)), and more. At these smaller scales, Open Context publishes detailed records of stratigraphic units, inventories of objects (artifacts), zooarchaeological and paleobotanical records, and associated digital images, mapping data, and occasionally 3D models.



Table 1: Examples of Different Types of Data Publications Hosted by Open Context

Publication Type	Example
Stand-Alone <i>Stand-alone projects use Open Context as the primary dissemination channel, rather than a supplement to a journal or monograph publication.</i>	Murlo / Poggio Civitate Excavation Project (Tuck 2012)
Informatics Research <i>These projects emphasize linking and integration of many disparate data sets. Such projects facilitate larger scale "meta-analyses" of aggregated data.</i>	Digital Index of North American Archaeology (DINAA) (Anderson <i>et al.</i> 2015); EOL Anatolian Zooarchaeology Linked Data Project (AAJ , Arbuckle <i>et al.</i> 2014 , Kansa <i>et al.</i> 2014)
Grant Data Management <i>Projects whose data publication fulfills plans described in a data management plan required by a granting body (such as the National Science Foundation).</i>	Architecture and Urbanism at Seyitömer Höyük (Harrison 2017); Oracle Bones in East Asia (Brunson <i>et al.</i> 2016); Northern Mesopotamian Pig Husbandry (Price 2016); The Amulets of the Kerma Culture (D'Itria 2022)
Archival <i>Open Context publishes museums and other institutionally managed archival collections.</i>	ARCE Sphinx Project 1979-1983 Archive (Lehner 2017); Badè Museum of Biblical Archaeology (Brody 2010); Fort Snelling (Hoffman 2016); Abu Huerya Digital Data Archive (Moore <i>et al.</i> 2023); 3D History for Ukraine (Radchenko <i>et al.</i> 2023)
Article Reproducibility <i>Data sets related to conventional journal publications, shared by researchers so that readers can access the primary data and replicate or build on the work.</i>	Differentiating Local from Nonlocal Ceramic Production at Iron Age Sardis (Kealhofer <i>et al.</i> 2017); Ceramics, Trade, Provenience and Geology: Cyprus in the Late Bronze Age (Stopic <i>et al.</i> 2013); Chogha Mish Fauna (Atici <i>et al.</i> 2010)
Supplement Monograph <i>Certain projects use Open Context to publish full data sets, image collections, and other media to supplement publication in conventional monographs, especially archaeological site reports.</i>	Petra Great Temple Excavations (Joukowsky 2007); Kenan Tepe (Parker and Cobb 2012); Gabii Project (Opitz <i>et al.</i> 2017); Visualizing Votive Practice (Counts <i>et al.</i> 2020a); Madaba Plains Project-`Umayri (Clarke and Herr in prep)

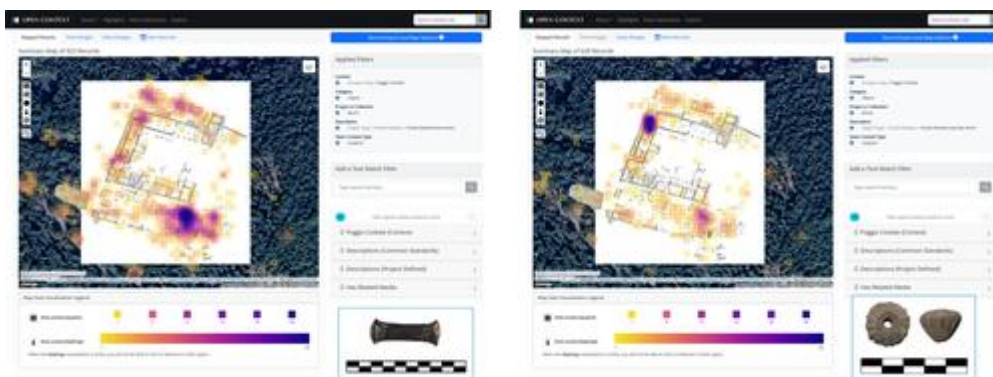


Figure 4: Open Context Query Results Mapping Intra-Site Distributions of Textile Implements

The different datasets described above come from different data sources (spreadsheets and relational databases) created by different people and organizations. These different data sources used different descriptive systems and terminologies to document a wide array of sites, contexts, artifacts and ecofacts. Nevertheless, Open Context provides common data hosting and services for these large and diverse collections.

4.3 Trade-offs of Schema-Free Abstractions

As described above, Open Context provides a common software and data infrastructure to manage archaeological data from hundreds of different sources. In principle, this enables certain "economies of scale". Each dataset, each with unique terminologies and descriptive systems, that Open Context publishes does not require its own custom database and software application. Nevertheless, Open Context's emphasis on schema-free design implies a different set of trade-offs in comparison with other more typical approaches to archaeological database design.

On the positive side, schema-free design makes it easier for Open Context to manage ETL workflows to ingest data. Mappings between a source dataset and Open Context's generalized and abstract structures tend to be simple and easy to configure. In contrast, the less abstract and more rigid database schema typical of other systems would require much more detailed (and laborious and complicated) mapping between specific data elements unique to a given dataset and the specific data elements expected by the larger target database. The simplicity of Open Context's schema-free abstraction reduces some of the effort required in ETL processes.

In addition, the schema-free abstraction of Open Context generally allows for more complete ingest of source data. For example, if Open Context had a more conventional and rigid structure, certain data elements from a source dataset may not map to *any* data elements of the larger target database. Without abstraction, these unmappable data elements would need to be ignored and discarded, or the target database would need potentially costly modifications and extensions. Therefore, the schema-free abstraction chosen by Open Context helps make it more cost-effective to more comprehensively import (via ETL) source data into a common data store.

While schema-free abstraction helps to simplify and speed ETL processes for Open Context, it makes other aspects of data management more complex and difficult. It is more challenging to create intelligible and appealing user interfaces for data that can vary widely in topic and description. Because some imported data can have complex hierarchic structures, it is also more difficult to efficiently index such complex data to enable search, discovery, visualization and analysis. The schema-free design also makes it more challenging to design user interfaces to edit database



records. Because of the high-level of abstraction, it is harder to make forms with validation logic and other usability features that would simplify and streamline delete, update, and creation operations for users. While Open Context has some of those features, they are somewhat difficult to use and mainly support work conducted by Open Context's experienced editorial team. These editing features are rarely used by non-staff users of Open Context.

These challenges ultimately lead to usability issues and sometimes performance problems. Although Open Context maintains common interfaces to publish millions of records from hundreds of different sources, not all of those records are easily comparable. Open Context has no standard, globally enforced way of describing potsherds, animal bones, archaeological sites, or anything else. Although accessible and queryable through a common interface, such records are typically described using different attributes and vocabularies. That diversity complicates search, visualization, and analysis. Open Context needs additional strategies to help make this wide-ranging diversity of content more intelligible and usable.

4.4 Annotations to Promote Interoperability

While data is not valuable only in terms of efficient interoperability and aggregation, interoperability is nevertheless still an obvious major goal. After all, interoperability is the "I" of the FAIR data principles (see above, and discussion about data granularity issues below). Open Context attempts to promote interoperability while it still represents the descriptive systems and vocabularies unique to each given dataset. In order to manage these competing needs, Open Context adopts a number of strategies.

First, while Open Context's internal datastore emphasizes schema-free design strategies, it is not wholly schema-free. Open Context does require certain general database attributes for all the records that it manages. Many of these attributes track data provenance, creation and update times, and certain generalized classification attributes to help organize records ingested through ETL workflows. Some provenance attributes track key metadata that associate data records with "projects" (the main unit of publication in Open Context). These attributes help inform how Open Context manages citation and attribution metadata (see discussion of persistent identifiers below). Provenance tracking also makes it easier to "undo" changes introduced to Open Context's database from a specific ETL process that may have contained errors. Other attributes describe if a given record describes a "media resource" (which will be expected to have associated media files at some location), a "document" (which will be expected to have HTML text), or "type" (which is a controlled vocabulary concept used in description), etc. These common data elements imposed on an otherwise largely schema-free database design provide the basis for many of Open Context's user-interface and search features.

A second approach to interoperability is facilitated by Open Context's ability (via schema-free abstraction) to represent individual records imported from source datasets. This means Open Context maintains a high-level of granularity representing information. This granularity, which identifies each individual site, context, object, and other records with a unique URI (Web address and identifier), enables Open Context to participate in Linked Open Data methods and standards. Stable Web URIs allow Open Context-published records of individual sites, bone specimens, artifacts, etc. to be "targets" of links from outside systems. For example, some Open Context records documenting late Roman coins are referenced in this way by the "Online Coins of the Roman Empire" (OCRE), a specialized numismatic database.



Similarly, records of "predicates" (descriptive attributes or linking relations) and "types" (controlled vocabulary concepts used in description) that Open Context imports from source datasets can also be referenced for Linked Data purposes. The Open Context editorial team links collections, ontologies, and vocabularies curated by other institutions and expert communities across the Web. The clearest example of how these links promote interoperability can be seen with zooarchaeological data. To help promote interoperability of animal bone records, Open Context's editorial team creates linking annotations between project specific terms and concepts curated by outside expert communities. For example, the "type" (controlled vocabulary concept) called "[astragalus](#)" from the *Neolithic and Bronze Age cattle data from Switzerland* project has a "close match" (a relationship assigned by an Open Context editor) with the UBERON defined concept "[talus](#)". Through such annotations that link multiple project-specific terms to UBERON, animal bone records in Open Context can be indexed and searched using more standardized terminology applied across multiple projects and multiple datasets. At the same time, Open Context still retains and represents each given dataset's original, and sometimes unique, set of descriptive terms.

As discussed, Open Context "record-by-record" granularity and the inclusion of "predicates" (attributes) and "types" (classification terms) makes it feasible for Open Context's editors to create cross-referencing annotations that relate dataset specific terminologies with more widely used community and domain standards. These cross-references enable searches and queries across multiple datasets submitted by different contributors.

In addition to aligning vocabularies, this granularity and schema-free emphasis also enables Open Context's editors to associate specific records to information resources managed in other repositories and databases across the Web. For example, Open Context publishes the Digital Index of North American Archaeology (DINAA), which includes data from over 800,000 site records from state and other US government agencies (at **limited geospatial precision** and with other sensitive information redacted to **protect site security**). These data give the single most comprehensive overview of human settlement of North America from the Pleistocene to the present. Open Context's granularity mobilizes these data so they can serve as a gazetteer for integrating museum and library collections (articles, books; Wells *et al.* [2017](#)).



Table 2: How Open Context Implements FAIR Data Practices

FAIR Characteristic	Open Context Implementation
F1. (Meta)data are assigned a globally unique and persistent identifier	<p>Each individual record (site, context, object, etc.) is assigned:</p> <ul style="list-style-type: none">• (Stable) Web URIs• Archival Resource Keys (ARKs)• Document Object Identifiers (DOIs) assigned to projects, data tables and other more "aggregate" resources <p>Metadata also expresses identifiers via Schema.org and Dublin Core standards.</p>
F2. Data are described with rich metadata	<p>Each record has Schema.org "Dataset" metadata. Projects have Dublin Core metadata, referencing:</p> <ul style="list-style-type: none">• Library Congress subject headings• PeriodO time periods• Gazetteer (Geonames, Wikidata, Pleiades) documented geographic places.
F3. Metadata clearly and explicitly include the identifier of the data they describe	<p>Open Context publishes very granular data, with persistent identifiers assigned to named entities (individual sites, stratigraphic units, artifacts, ecofacts). Each record is associated with specific "projects" (also identified with persistent identifiers) to provide information about the intellectual provenance of information. The Dublin Core and Schema.org standards express these metadata.</p>



F4. (Meta)data are registered or indexed in a searchable resource	Data is dynamically accessible and searchable on Open Context's Web-based platform, which features comprehensive indexing, full-text search, and faceted search. SiteMap XML and Schema.org metadata are provided for web-crawlers and commercial search engines.
A1. (Meta)data are retrievable by their identifier using a standardised communications protocol	Content is retrievable via a common API ("application program interface") and displayed through a uniform user interface on the publicly accessible website.
A1.1 The protocol is open, free, and universally implementable	Open Context makes all data available in a variety of representations (HTML, JSON-LD) over the Web using the HTTPS protocol. Bulk data exports are also available from Zenodo via the HTTPS protocol.
A1.2 The protocol allows for an authentication and authorisation procedure, where necessary	Because it lacks the administrative capacity to manage access permissions, Open Context publishes only open access data.
A2. Metadata are accessible, even when the data are no longer available	Open Context employs a "separation of concerns" strategy, focusing on curation and access while relying on external digital repositories (like California Digital Library - Merritt and Zenodo) to provide long-term data preservation.
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Structured data (including data and metadata) is communicated to outside systems using formal standards like GeoJSON (for geospatial information) and the JSON-LD format (for Linked Data annotations and Schema.org).
I2. (Meta)data use vocabularies that follow FAIR principles	Open Context's editors annotate project-specific terms with widely used community and domain standards, such as UBERON (for zooarchaeology) and concepts from the Getty AAT and GBIF.
I3. (Meta)data include qualified references to other (meta)data	Open Context uses Web URIs to identify and reference specific concepts and other named entities from other data providers. These include URIs to gazetteer described places, URIs to PeriodO published time periods, and URIs to controlled vocabulary concepts in the Getty AAT and other sources.



R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	<p>Open Context uses Dublin Core and Schema.org metadata standards to document:</p> <ul style="list-style-type: none">• Attribution / Data creators• Subjects / topics (using Library of Congress subject headings)• Chronological coverage (using PeriodO time periods)• Spatial Coverage (using gazetteer defined places)• Background information (free text abstract, descriptions)
R1.1. (Meta)data are released with a clear and accessible data usage license	<p>Dublin Core and Schema.org metadata includes a "copyright license" attribute. Open Context uses standard Creative Commons licenses, usually the Attribution License.</p>
R1.2. (Meta)data are associated with detailed provenance	<p>Each record is associated with specific "projects" (also identified with persistent identifiers) to provide information about the intellectual provenance of information. The Dublin Core and Schema.org standards express these metadata, including authorship information.</p>
R1.3. (Meta)data meet domain-relevant community standards	<p>While not fully harmonizing all content, Open Context's "high touch" approach to curation attempts to at least partially annotate data with domain standards (ontologies, controlled vocabularies, and related data curated in other platforms).</p>

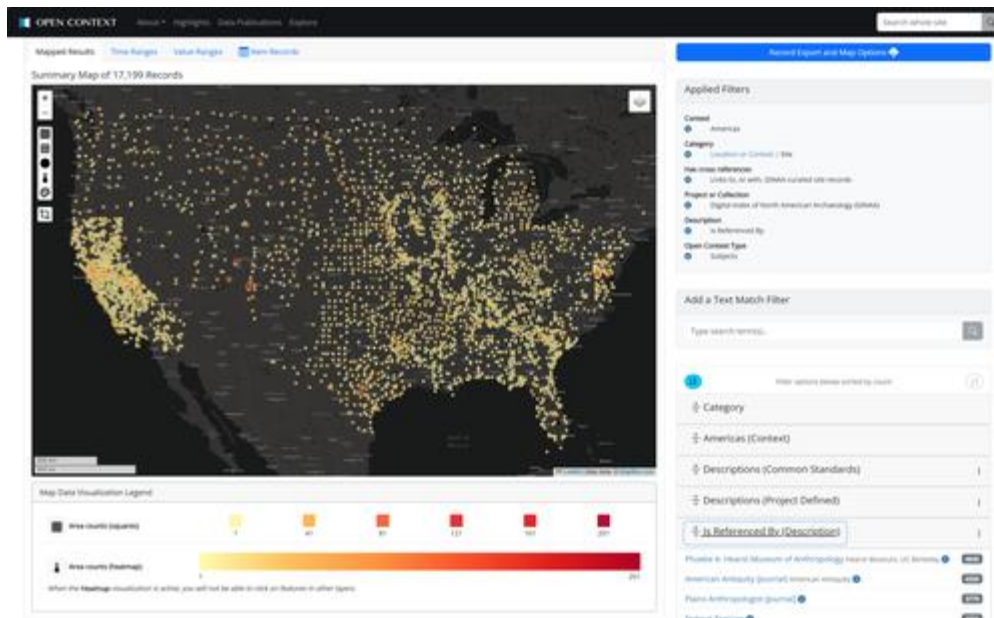


Figure 5: DINAA Records with Links to Resources in tDAR, JSTOR, the Federal Register, etc.

Open Context's approach toward interoperability does not fully harmonize with all data to make records from one project fully fungible with records from another project. Open Context currently only makes very limited use of domain ontologies like the CIDOC-CRM that would more fully semantically harmonize data. However, while use of such ontologies may advance interoperability, they would come at the cost of adding more labor and complexity (see Tudhope *et al.* 2011) to Open Context's ETL workflows. In addition, there's much more to harmonizing data than semantics. Methodological differences, sampling choices, and other factors that shape data creation pose significant methodological and theoretical challenges to the interpretation of aggregated datasets. Open Context does not attempt to solve these issues for data reusers. Instead, Open Context's (admittedly) somewhat simplified and purposefully vague semantics together with Linked Data cross-references across datasets mainly facilitates search and discovery of relevant (but still diverse) information (see a more general discussion of pragmatic use of "weak semantics" in Baker and Sutton 2015). Open Context does not intend to make this information "standardized" in every conceivable way.

4.5 Machine Audiences on the Web

The metadata and Linked Data annotations applied to Open Context published data has implications that go beyond the user interface. In reality, human user interactions represent only a small fraction of the overall traffic served by Open Context. As a publicly accessible website, Open Context also responds to a great deal of automated requests made by software agents (bots, web crawlers, and the like). Many of these requests come from commercial search engines, and recently commercial Artificial Intelligence (AI) developers have also begun to direct bot traffic to Open Context. Because software agents (bots) play such an important role on the Web, a large fraction of our software development and systems administration attention needs to be devoted to managing bot traffic.

In many ways, bots play a key role in "interoperability". Interactions mediated by software agents communicate data published by Open Context to other systems and services. As described below, commercial search engines (and AI-bots) make up the vast majority of these interactions. However, Open Context also communicates data with other scientific data infrastructures using a variety of standards with varying levels of semantic specificity (see Figure 6). For example, at a low level of



semantic specificity, most Open Context resources have geospatial information, and this information is communicated using the GeoJSON standard. More semantically specific information is communicated to outside systems such as iSamples.org, which aggregates information about individual physical specimens. Open Context annotations to concepts in the Getty Art and Architecture (AAT) and GBIF, expressed in a JSON-LD API, get read and incorporated into the iSamples index.

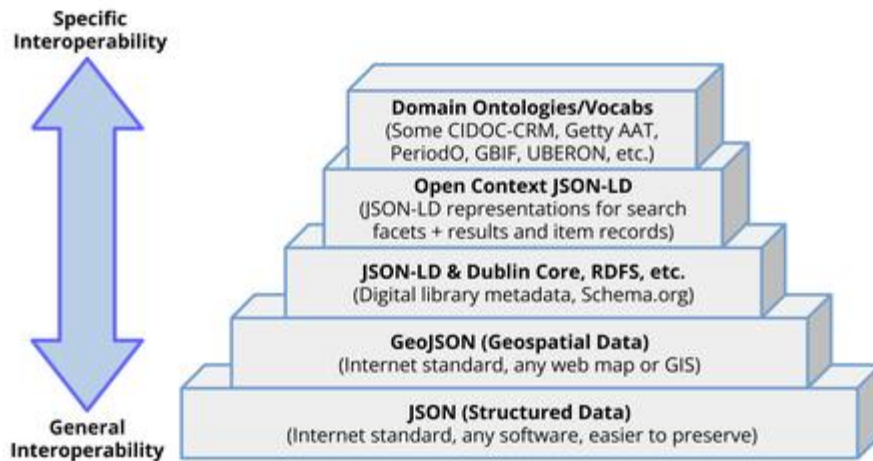


Figure 6: Different Standards and Levels of Interoperability

While Open Context makes structured data easily accessible via APIs (Figure 6), the types of structured data that see the most access and use are those that are recognized by commercial search engines. Search engines and now commercial AI providers play an important role in public discovery and access. These are key considerations for the "F" (Findable) and "A" (Accessible) aspects of the FAIR Principles. For this reason, Open Context implements a number of features aimed at web-crawlers. These features include:

- **Extensive Linking:** Open Context publishes archaeological data records as individual webpages that richly link to related context, media, and other webpages. Because these links are expressed in simple HTML, web-crawlers can easily follow links to discover new pages to index.
- **SiteMap XML:** Open Context also provides search engines with lists of webpages to index using the SiteMap XML standard. We use the SiteMap XML list to prioritize certain webpages for search engine indexing. We select content that is either richly described or has unique descriptive attributes for inclusion in the SiteMap XML list. This helps to make sure that search engines capture a more comprehensive index of Open Context content than would be feasible through unguided crawling.
- **Schema.org Metadata:** Open Context also includes Schema.org metadata (expressed in the JSON-LD format) in the web resources it publishes. The metadata describes Open Context content using the Schema.org "Dataset" schema, including bibliographic attributes such as publisher, creators, copyright license, identifiers, and access permissions. These forms of structured data help shape how commercial search engines represent Open Context content in their indices.



4.6 Social and Professional Limitations on Interoperability

While Open Context's annotations with Linked Data make a high degree of data integration feasible, Linked Data has coverage and specific challenges. The most important limitation centers on gaps in the availability of relevant "standard" vocabularies and ontologies. For example, while Roman numismatics benefits from excellent and highly detailed standard controlled vocabularies available as Linked Data, the same cannot be said for glyptic (seals and seal impression) for Neolithic Mesopotamia. While widely applicable, more general-purpose controlled vocabularies for the cultural heritage sector, such as the Getty Art and Architecture Thesaurus (AAT) lack the specificity researchers may require for specialized studies.

For the most part, our professional community has yet to develop and share computationally actionable controlled vocabularies needed for the specific and nuanced description of material culture in many parts of the world and time periods. Researcher energies and expertise have yet to be directed toward such efforts. Instead, most discussion of material culture classification and typology takes place in literature (articles and books) and very rarely does any of this discussion lead to the development and publication of classification systems as structured data, much less linked data. Moreover, many aspects of ancient material culture defy easy classification, especially if they originated from household or small-scale, *ad hoc*, or individualistic contexts of production. There are important archaeological method and theory questions concerning classification of such materials (Okumura and Araujo [2024](#)).

Thus, problems in archaeological interoperability closely relate to the manner in which archaeologists choose to publish and disseminate their research and their supporting typologies. Simply put, typologies are generally not considered a form of "data" and are not published in a manner that allows them to be easily integrated with other archaeological data. One approach may use Large Language Models (LLMs) and other AI technologies to help distill published archaeological typologies into more easily usable structured data. However, even if LLM powered approaches gained feasibility, LLMs may be problematic to deploy and use because much archaeological literature is proprietary (owned by commercial publishers) and guarded by licensing and other restrictions. In addition, the LLMs themselves are often opaque "black boxes" with little transparency in how they are trained. To further undermine trust, LLMs also typically poorly maintain and respect information provenance. Thus, for the time being, LLMs seem like a problematic way to address vocabulary alignment and other archaeological data integration needs.

While the lack of specialized typologies makes material culture data harder to integrate, data creation practices also play a role in complicating interoperability. We have also found that common data quality and consistency problems make raw data difficult to integrate. For that reason, Open Context includes both data cleaning and integration into our editorial workflows. Archaeological data are often created without much formal data validation or other techniques (such as drop-down lists or checklists) used to encourage consistent data entry. Spelling variations, abbreviations and typographic errors are all common in many archaeological datasets.

In addition, many archaeologists lack training in data modeling and others have practical and pragmatic reasons to make certain compromises in data modeling, even if those compromises adversely impact the analytic potential of data. This particularly complicates recording of attributes that can have multiple values. For example, a given archaeological potsherd can have multiple types of decoration (painting, slip, incisions, plastic application, etc.). However, if a researcher used only one column in a data table to record decoration, then instances where a sherd may have multiple decorative types may be awkward to record.



Such commonplace constraints on data modeling do not necessarily indicate sloppiness or incompetence. Data modeling decisions are driven by a number of factors, including habituation with certain software tools (such as Excel), expediency, and how a researcher may intend to use different elements of data (see Batist [2025](#)). For example, a researcher may not intend to use values in a multi-attribute column for quantification— thereby reducing the impact of inconsistent expression of values. In addition, certain data modeling limitations may be perfectly acceptable to an individual archaeologist working with a relatively small number of records. At small scales, a human being can use their judgement to account for such inconsistencies. However, if one attempts to aggregate such data into larger collections (such as Open Context), then such problems become too time-consuming to navigate on a record-by-record basis.

For these reasons, Open Context's editors devote significant efforts toward data cleaning prior to ETL import into Open Context. While tools like [OpenRefine](#) and some one-off custom scripts (using the Python programming language) can improve the efficiency of data cleaning, most datasets require some level of review and cleanup. These editorial workflows can be time consuming and can involve back-and-forth conversations with data contributors. Thus, Open Context typically needs to devote some time and technical expertise to rework a dataset created to meet the needs of an individual researcher (or small team) into a dataset capable of wider interoperability and aggregation.

5. Open Context in Comparison with the tDAR Repository

A direct comparison between Open Context and tDAR helps to illuminate how different systems with different architectures and service models can all negotiate challenges in archaeological data management (see Sheehan [2015](#) for an earlier comparison). [The Digital Archaeological Record](#) (tDAR) is the leading United States based disciplinary digital repository for archaeology (Nicholson *et al.* [2021](#)). tDAR has a history of large institutional support and investments, including promotion and governance from some of American archaeology's leading scholars and major financial support from the Andrew W. Mellon Foundation. tDAR also has a history of significant efforts to promote wider interoperability with partners internationally (particularly with the ADS) and with other US based scientific data infrastructures (especially the DataONE collaboration).

Like most digital repositories, tDAR primarily provides digital preservation and access services for digital files described by metadata. tDAR's emphasis on archiving data files has many advantages in terms of costs and ease of submission. Unlike Open Context, where data travel through staff-supervised editorial and ETL processes, researchers submitting data into tDAR can use "self-service" interfaces to upload files and provide metadata. tDAR's approach aims to reduce cost and complexity barriers to using basic data preservation services to meet certain compliance needs, including grant required data management plans.

At this time, tDAR's most numerous records are bibliographic references to offline content (c. 393,000 citation-only metadata records), followed by image files (c. 30,000), reports (c. 20,000), datasets (2,174), and ontologies. Open Context's holdings present a very different numerical picture. Open Context has published comparatively more images (c. 175,000) and supplementary field note documents (c. 14,000) to date than accessioned into tDAR but has no citation-only metadata records.

Because of their very different architectures and content, comparing Open Context and tDAR involves "apples and oranges" challenges (summarized in Table 3). It is therefore difficult to directly compare the amount of structured data stored in each of these systems. Open Context currently has



published more than two million citable records from over 200 different "projects", making it comparable to major museum online collections (but with much richer archaeological provenience data). Some of the larger projects in Open Context involved ETL of dozens of different data tables from multiple source spreadsheets and relational databases. Nevertheless, tDAR probably has many more records of structured data distributed across its archive of 2,174 data files than Open Context has in its central database. To explain further, each data file in tDAR might include thousands of records described within a single file, whereas Open Context represents each record individually, reflecting a vastly different levels of granularity.



Table 3: Granularity Comparison between Open Context and Digital Repositories



Digital Repository

Granularity	High ("item-level") identification and retrieval of individual records of specific sites, artifacts, contexts, etc.	Low ("collection-level") Information typically aggregated in larger chunks (data files)
Persistent Identifiers (PIDs) and Citation	PIDs (and Web URIs) assigned to specific archaeological entities (sites, coins, bones, etc.)	PIDs assigned to collections and / or digital files
Discovery, Querying	Common datastore and indexing for specific record-level content, not just collection metadata	Mainly indexes collection-level metadata. Content within data files is typically more opaque to search and discovery
Cost and Labor	Labor intensive data editing and ETL (for ingest) costs	Low-cost "self-service" models for deposit and metadata documentation



5.1 Item and Collection-Level Citation Granularity

The differences in granularity between Open Context and tDAR illustrate a fundamentally important but rarely discussed (however, see Katsianis *et al.* [2023](#)) aspect of information management in archaeology. tDAR, like many other digital repositories, mainly concerns itself with managing digital files. Each file can be a large (coarse) aggregate of many individual records of data. In contrast, Open Context and other database-oriented information systems (such as Pleiades, EAMENA, the Levantine Ceramics Project, etc.) typically organize information at a much finer degree of granularity.

A digital repository like tDAR's emphasis on relatively coarse "collection level" (or file level) metadata has implications for search and data reuse. When one searches and cites data in tDAR, one discovers and references a file, which itself can be a very large aggregate of data. To fully explore the contents of a data file, a user must discover that file by metadata description and then download and open it with appropriate software. A given data file in a repository like tDAR can describe hundreds or thousands of records of archaeological sites or artifacts. Individual sites, contexts and objects described in these files cannot be specifically (or directly) referenced or cited.

In contrast, Open Context's high level granularity enables retrieval, linking, and citation of much more specific "item-level" entities. Through integration of the EZID service provisioned by the California Digital Library (CDL), Open Context mints a variety of persistent identifiers (PIDs, see Koster [2020](#)). Open Context assigns Document Object Identifiers (DOIs), a type of persistent identifier widely used by digital repositories and journal publishers, to "collection-level" resources (mainly "projects" and static downloadable data tables. Open Context also assigns a different type of persistent identifier, ARKs for individual "specimen" (artifact, ecofact, and sample) records (Kansa and Kansa [2022](#)). Though less widely supported by indexing services (like CrossRef and DataCite), ARKs can be generated free-of-charge and assigned to millions of individual resources. These persistent identifiers can enable cross-disciplinary discovery, tracking, and citation of specific archaeological sites, ecofacts, artifacts, contexts, much as DOIs enable citation of publications and datasets.

Open Context issued PIDs are beginning to see citation in various scholarly publications. The edited volumes, *Visualizing Votive Practice* (Counts *et al.* [2020b](#)) and *Corinthian Countrysides: Linked Open Data and Analysis from the Eastern Korinthia Archaeological Survey* (Pettegrew [2024](#)) both extensively cite many Open Context ARK identifiers. These citations help cross-reference the scholarly interpretation and argumentation presented in books with specific records of primary data hosted by Open Context (Caraher and Pettegrew [2025](#)). As described elsewhere (Kansa and Kansa [2022](#)), use of PIDs, especially for the identification of item-level resources can play an important role in maintaining the contextual integrity of archaeological data.

In addition, PIDs enable tracking of data citations, potentially enabling quantification of citation metrics and other measures of scholarly impact. However, we should caution against over interpretation of such metrics. The impact and value of data curation should not be reduced to a popularity contest measured by citation counts. Archaeological data management helps to preserve irreplaceable documentation records of cultural heritage and should be valued even if few resources see much citation over short time horizons. Moreover, the infrastructure for tracking data citations is still in its infancy. Indexes of DataCite and CrossRef DOI citations are spotty and incomplete, and to our knowledge, no service aggregates counts of ARK citations.



5.2 Collection- and Item-Level Granularity Issues with Interoperability

The fundamental differences in granularity between Open Context and tDAR are reflected in many other archaeological information systems. Granularity differences go beyond data citation affordances and also help to shape patterns in interoperability. Katsianis and colleagues ([2023](#)) provide an excellent introduction to this topic, and how different levels of granularity impact interoperability. As they discuss, resource granularity fundamentally shapes interoperability programs:

However, in contrast to contexts in which such efforts succeeded, such as the granular integration of coin datasets (Felicetti et al. [2015](#)), the integration and retrieval of excavation data resources has remained limited and is mainly available in the ARIADNE portal at collection level and on the basis of subject keywords, rather than their item-level integration as individual objects.

In this sense, Open Context and tDAR further demonstrate how archaeological information systems can work at very different levels of granularity.

These granularity distinctions between Open Context and tDAR shape patterns in interoperability. While tDAR provides tools for data integration of archived datasets via ontology alignment, tDAR's main emphasis on interoperability has focused on information exchanges with other (peer) digital repositories such as the ADS (see example of interoperability efforts between European repositories in Richards [2023](#)). In contrast, examples of Open Context's interoperability with outside systems have mainly included exchanges with similar "fine-grained" databases, including OCRE (discussed above) and more recently iSamples. The [iSamples](#) project aggregates records and persistent identifiers to document individual physical specimens from several multidisciplinary sources, including SEASR (geology), the Smithsonian (biodiversity), and Open Context (archaeology). The Open Context team is also currently working in collaboration with the developers of [Neotoma](#), a large integrated database of records curated by paleoenvironment researchers.

This discussion highlights how different data infrastructures can choose very different paths toward implementing high-level principles like FAIR. To repositories, "interoperability" may mean the ability to exchange commonly understood metadata about archived digital files. To a more granular database like Open Context, interoperability means the ability to exchange mutually intelligible data about specific archaeological "entities" such as: sites, bones, ceramics, contexts, etc. While both valuable, these are very different kinds of interoperability.

5.3 Working with Multiple Levels of Granularity

Thus, granularity issues seem to play an important role in how interoperability efforts have evolved within archaeology and with other scientific information systems. While differing levels of granularity can complicate some forms of interoperability, data management needs extend beyond interoperability. A digital repository like tDAR can provide essential digital preservation services through less labor-intensive self-service models than Open Context. After all, digital repositories mainly manage metadata in larger "chunks" (data files) and typically devote less attention to the individual data records contained within those data files. For this reason, tDAR has archived larger quantities of structured data (in various data files) than Open Context has published through its editorially supervised ETL processes.

In addition, some forms of interoperability can still be achieved despite differences in information granularity. The Open Context-hosted DINAA project (described above) publishes data that already enriches collections metadata managed by tDAR. tDAR references DINAA site URIs (Web identifiers)



provided by Open Context to enhance its own metadata, and Open Context links its DINAA records to tDAR-archived resources. This allows Open Context users to discover tDAR-archived records through cross-referencing across these platforms. Essentially, tDAR uses some 1500 archaeological site records curated by Open Context to augment some of its own metadata records, thereby bridging some of the granularity differences between these two systems.

6. Equity, Openness, and Data Protections

The collaboration between tDAR and Open Context demonstrates the feasibility of developing a degree of interoperability across systems despite differences in information granularity. At the same time, this example also highlights areas of potential complementarity around different frameworks for regulating access and security.

As implied by its name, Open Context is an open access resource. While tDAR provides different levels of data access to its users, Open Context publishes open data with no login restrictions. *Nevertheless, we emphatically do not believe that all data should be open.* However, Open Context itself lacks the administrative capacity to responsibly manage access permissions and associated security and liability risks (Kansa [2016](#)). Simply put, managing more sensitive data that requires access controls and processes to appropriately grant or deny authorization requests would not be feasible given Open Context's available financial, technical, and administrative resources. Moreover, Open Context lacks the knowledge and the legitimacy required to understand specific data sensitivities and risks in many cultural contexts, particularly Indigenous cultural contexts. Instead, Open Context focuses curation efforts on datasets that (hopefully) offer communities more benefits and fewer risks through open access dissemination. For that subset of cultural heritage data where openness is appropriate, Open Context reduces barriers to access and reuse through various technical (data standards that promote interoperability) and legal (Creative Commons licenses that grant explicit reuse permissions) means (see Kansa [2012](#)).

While open data should not be universal, it should not be seen as something in opposition to more guarded models of data management. Open data can be a powerful tool that can enhance metadata documentation of restricted-access data. As discussed, some 1,500 (low sensitivity, open access) Open Context published site records cross-reference with access-restricted records in tDAR. Thus, a user can learn about the existence of restricted data that need additional permissions to fully access (such as that housed by tDAR). This example illustrates that open access data can act as a finding aid (an index) to more complete but restricted-access data governed by different authorities. Open data used as an index can direct interested users toward the systems, organizations and communities that manage restricted access information. They can then follow the protocols and processes those communities deem necessary when they seek authorization to access restricted data.

DINAA's role as an open access index (finding aid) first developed in collaboration with tDAR has been extended to other open access and restricted access information systems. The Open Context team downloaded public domain United States Federal Government notifications from the Federal Register to find references to archaeological sites present in the DINAA catalog. Because DINAA site records contain geospatial information, this effort allowed the Open Context team to develop a spatial index of archaeologically relevant US government notifications in the Federal Register. Similarly, the DINAA team obtained (with JSTOR's permission) archaeological journal articles to identify archaeological site references to index. This also powers open access map-based user interfaces to search restricted-access archaeological literature stored by [JSTOR](#).



6.1 The CARE Principles and the Limits of FAIR

The above examples illustrate how open and more guarded information systems can play complementary roles. In principle, open access information resources can play a similar role in advancing CARE (Collective benefit, Authority to control, Responsibility, and Ethics; Carroll *et al.* [2020](#), GIDA [2020](#)) data practices. The CARE Principles for Indigenous Data Governance are often discussed along with the FAIR Data Principles as the two leading high-level frameworks for guiding good data management practices.

While both sets of frameworks have great prominence, the CARE Principles differ from FAIR in that they promote Indigenous self-determination, sovereignty, and community benefits while FAIR mainly focus on more technical considerations, without explicit consideration of equity, risk, and community issues. This is not to say that the FAIR Principles lack ethical grounding. On the contrary, the FAIR Principles have a historical basis in the "Free Culture" (or Libre Culture) movements that advanced an information "commons" free from exclusive, proprietary control by large corporate or government interests. Free Culture activists worked to build alternative information systems free from the surveillance, monopolies, "rent-seeking", and draconian legal penalties common to systems of commoditized intellectual property (for archaeology, see Kansa *et al.* [2013](#)). "Open Government", "Open Science", "Open Source", "Open Educational Resources", "Open Access", and "Open Data" are all related manifestations of activist attempts to carve out spaces for the exchange of knowledge free from legal and financial encumbrances and risks.

The ubiquity and success of open-source software, Wikipedia, greater open access to research, and other developments of the past 30 years highlight some of the Free Culture movement's successes. Nevertheless, the Free Culture movement (and by extension, the FAIR Principles) has received criticism and has faced increasing challenges (Mirowski [2018](#)). Many of the concerns center on issues of equity. Who pays for, and who benefits from, information in the "commons"? Increasing wealth disparities and corporate power can lead to very uneven capacity in who can benefit from the resources of the commons. While open resources like Wikipedia and Common Crawl (an open dataset derived from the mass-crawling of websites) are non-proprietary and non-exclusive, the *capacity to make effective use* of these resources -especially at scale- is very unequally distributed. For example, using these large open datasets to train a Large Language Model requires large budgets, sophisticated technical expertise, and vast (expensive) computing resources. The most powerful LLMs are owned by well-financed corporations because they disproportionately have the capacity to capitalize on both open and proprietary information resources.

Openness (even when applied in ethically appropriate settings) may often be necessary, but it is clearly not sufficient to promote wider equity (Kansa [2014](#)). The impacts of removing access or intellectual property restrictions cannot be understood without also looking at wider patterns in the distribution of wealth, technical capacity, and demographic diversity in those that shape and benefit from Free Culture initiatives. Openness (by itself) cannot undo wider systems of structural inequality, especially entrenched legacies of colonialism. If misapplied, openness can lead to continued harms.

The CARE Principles provide a framework for navigating how information plays a role in advancing the sovereignty of Indigenous communities and their resistance to systems of structural inequality. While publication of the CARE Principles helped to introduce and popularize these issues within research data management circles, ethical and policy debates about intellectual property in cultural heritage saw decades of prior scholarship (examples: Nicholas and Bannister [2004](#); Chander and Sunder [2004](#); Christen [2012](#)). Thus, various research programs attempted to navigate cultural



property questions, including as they relate to scientific data sharing (see also Heitman [2022](#)), long before publication of the FAIR and the CARE Principles.

Unlike FAIR, which is more technically focused (see [Table 2](#)), Open Context's work to advance CARE practices centers more on relationship building and evolving conversations that are difficult to neatly summarize in a single table. The Open Context team has long participated in cross-cultural data ethics discussions, even before the initial technical development of the platform itself (see Kansa *et al.* [2005](#); and reviewed in Kansa [2016](#)). This prior work, particularly participation on the 2008-2016 *Intellectual Property Issues in Cultural Heritage project* (IPinCH Team [2014](#)), informed development of Open Context's data accession, editorial, and intellectual property policies as well as Open Context's terms and conditions for service. These policies attempt to encourage beneficial and not harmful uses of the platform. For example, Open Context's policies explicitly ask contributors to engage in "community archaeology" prior to submitting datasets so that data submissions reflect wider consensus and approval from different constituencies. We recently updated these policies to explicitly reference and endorse the CARE Principles.

6.2 Ongoing Data Ethics Challenges

While data ethics and Indigenous data sovereignty have been a key policy focus for Open Context from the very beginning, progress in implementation has remained painfully slow. Although [Local Contexts](#) is attempting to help fill this void (Anderson and Hudson [2020](#)), there are still no widely accepted metadata standards that could be used to help identify and better describe data of significance to Indigenous or other descendant communities. Open Context provides flagging and content warnings to "opt in" to view records associated with ancestral human remains, even though all such records are only published from social contexts where such documentation is not generally considered harmful. Open Context uses biological taxa and other metadata criteria to enable this feature, but this is the only example of metadata and interface features that relate to cultural sensitivities.

Without such standards, there is no mechanism to communicate expectations for respectful and ethical treatment of content to search engines or other research data infrastructure. Besides the need for more metadata support, the Open Context team faces ongoing challenges with respect to both the time-depth and geographic scope of data submissions. Many datasets submitted to Open Context derive from much older field investigations and collections studies. Those earlier studies may not necessarily fulfill contemporary definitions of "community archaeology", but the data may be currently managed by people and organizations that engage in high standards of ethical conduct. The Open Context team relies on the contextual knowledge and expertise of data contributors to guide good-faith judgements about data dissemination risks and benefits, especially for cases that involve legacy data.

Most models of good practice in community archaeology involve long term cultivation of relationships and trust across different constituencies. These good practices help make archaeology more inclusive and responsive to community needs. However, these community archaeology practices typically take place in very localized settings, involving small numbers of researchers and local community partners. How does one adapt models for local collaboration to work for data management and data governance at larger regional, continental, and global scales? After all, DINA aggregates legacy data from several government sources and spans much of continental North America. Similarly, the *Cross-referenced p3k14c* project aggregates publicly reported radiocarbon sample data from across the world. These datasets intersect with the ancestral territories of many hundreds of different Indigenous communities as well as many other descendant communities.



Open Context is obviously not alone in managing research data of large geographic scope (for CARE data in the Earth Sciences, see Jennings *et al.* [2025](#)). Other scientific data infrastructures similarly curate data from across the world including regions with different histories of conflict, colonialism, diasporas and migrations, and other sometimes traumatic experiences. Open Context and other platforms need governance models to ethically and responsibly manage data. It is not possible for a small team of researchers or a data publishing service like Open Context to develop deep and meaningful personal relationships of trust with representatives of many hundreds of different communities. Other mechanisms and processes, applicable to a global scope, need to be developed to make data curation more responsive and inclusive to community needs. As discussed, FAIR and CARE have emerged as two of the leading high-level frameworks for good data management practices. While these frameworks reflect some current consensus about abstract goals, it is still challenging to define specific methods and practices for their implementation (Bruce & Cordewener [2018](#); see also Kansa *et al.* [2025](#)).

7. Sustainability and Future Directions

The above discussion about governance highlights some of the ongoing professional and ethical challenges in archaeological data curation. Efforts to understand and respond to governance needs require similar levels of time, effort, and expertise as required to deploy appropriate technologies. Of course, funding, time, and technical resource constraints shape how we can respond to all of these challenges.

Limits on resources require certain strategies. Open Context fundamentally depends upon a well-developed foundation of open-source software. In addition, as already discussed, Open Context depends on digital library provisioned technical infrastructure, particularly persistent identifier services and digital repository services. But we also depend upon a network of wider capacity and expertise in library and information science, the digital humanities, scientific data infrastructures, and archaeology. This wider expertise, sometimes available through scientific information services, contributes information standards, controlled vocabularies, ontologies, and the expertise needed to provide peer-review services for our data publications. It also goes without saying that we also depend upon the wider archaeological community to conduct the field work, build the community partnerships, and create the datasets we seek to curate.

Because so much of Open Context's work depends upon contributions from the wider community, much of our recent work has focused on investing in broader community partnerships and capacity building. Our team launched multiple initiatives relating to instruction, professional development, and other forms of capacity building. We also broadened collaboration with other data infrastructures inside and outside of archaeology. These collaborations include:

- iSamples: With NSF support, the iSamples project (Davies *et al.* [2021](#)) integrates Open Context curated archaeological data with data from SESAR, GEOME, and the Smithsonian (representing earth sciences, biological collections, ethnography and archaeology).
- "Ethical Open Science": The FAIROS Research Coordination Network, funded by the NSF, integrates Open Context data with quaternary biodiversity data curated by Neotoma. In addition to interoperability efforts, this project also explores models of improved data governance, especially in response to the CARE Principles.
- FAIR+CARE Cultural Heritage Network: This project, funded by the Institute of Museum and Library Services (IMLS, a US government granting agency), seeks to promote models of better ethical practice in cultural heritage data management. This project was launched in



2023 in partnership with Digital Antiquity (tDAR). In April 2025, the Trump Administration abruptly terminated this grant, but a court order later reinstated it.

While the contributions of the wider community make it possible for Open Context to function, operating and maintaining Open Context still requires more direct investment. To stay online, cloud computing service fees need to be paid and the in-house software maintenance, data editing and data curation labor conducted by the Open Context staff needs to be financed. Most of the financing for Open Context's ongoing core operations and shorter-term projects ultimately comes from grants. Overhead charged on grants for short term projects (including the ones listed above) helps to finance core operations such as cloud computing service fees. It is only in 2024 that the Open Context team secured grant funding (from the NSF Archaeometry program) to directly offset core operational costs.

Obviously, this emphasis on contingent "soft money" support in the form of grants involves great risk and uncertainty. The Trump Administration's abrupt termination of grants and elimination of granting agencies compounds this financial risk. Unfortunately, there are very few alternative sources of revenue. Archaeology is not awash in discretionary funding that can be directed toward data curation. Data publication fees (a one-time fee charged to data contributors) cover only about 10% - 20% of our operational expenses. It is not feasible (nor equitable) to charge more for our services because most archaeologists lack access to more funding.

Nevertheless, Open Context has achieved some notable longevity. The service has operated continuously for nearly 20 years and has done so outside the umbrella of any university or government agency. To maintain continuity in operations in the face of uncertain and intermittent grant support, the Open Context staff needs to adopt pragmatic and expedient employment strategies to maintain salaries. Over the years, our team has taken on additional employment as faculty, software engineers, and university administrators to maintain stability in their personal income. Thus, Open Context's staff often works on a part time basis to continue critical research and development while relying on stable income from other sources. While sometimes stressful and often distracting from work on Open Context's mission, engaging in outside employment has had rewards. For example, Eric Kansa worked as a software engineer for a public health nonprofit. In doing so, he worked with a Google alumnus and received additional training and mentoring in industry standards for software engineering. This professional development experience helped to inform significant improvements to Open Context without the need to hire an outside software engineer.

7.1 Changing Technologies

As discussed, over the long-term Open Context has relied upon a variable mix of income sources to sustain its operations and its staff. Financial sustainability has involved a great deal of pragmatism and compromises and continuous adaptation to a changing landscape. Similarly, the technology picture has seen constant change.

Since its initial launch in 2006, Open Context has integrated a changing mix of open-source software technologies. We have significantly overhauled the software components used by Open Context on several occasions. In 2012, we implemented Solr search component to add faceted search and performance. In 2014, we migrated away from a bespoke MySQL and PHP application to a more maintainable application using Python, Django, and PostgreSQL components. Between 2020 and 2022, we made a significant upgrade to the Django based application to further improve maintainability, performance, and user-interface features.



Even though Open Context uses conventional and widely adopted and supported "off-the-shelf" open-source software components, integrating these components still demands significant effort. Open Context's emphasis on "schema-free" abstraction, though a pragmatic approach for ETL-based data in-take (see above), makes software development more challenging. Therefore, software maintenance and updates make significant time and resources demands on the Open Context staff.

It may be possible to reduce some resource demands required by software. The Open Context team is currently working to integrate the open-source Arches project as a core data management component. In many respects, [Arches](#) uses very similar technologies and design strategies as Open Context's backend software. Arches also uses Python, Django, and PostgreSQL components to manage a "schema-free" style data store. Because Arches meets Open Context's central need for "schema-free" abstraction, and it uses familiar technology components, it is an attractive choice for use with Open Context. Moreover, Arches has a much larger and better financed software development team. A dedicated endowment, led by the Getty Foundation, provides unparalleled financial guarantees for the continued maintenance of the software. Finally, Arches is also used by many more organizations, including most notably Historic England. This wider network of implementations will help further maintain and enhance the software, enhancing its prospects for sustainability. For these reasons, adopting Arches can help reduce the long-term burden of maintaining and updating Open Context's software.

There are several operational and technical questions that need to be solved to use Arches for Open Context. For one thing, although Arches has "schema-free" capabilities, most Arches implementations use these capabilities to enable non-technical users to design their own database schemas for use in Arches. Therefore, in practice, most Arches implementations define custom specified schemas. The specific Open Context need for schema-free abstraction to support ETL of diversely described and organized new data has no direct parallel in current uses of Arches. A core capability of Arches is centered on its deep support for ontologies to structure data modeling within its graph-style database. While powerful, it will require time and experimentation to understand how to take advantage of Arches' impressive support for rich semantic modeling while still retaining Open Context's capability to efficiently ingest and represent various "bespoke" systems of description unique to individual datasets. Nevertheless, despite this initial challenge, in the long term, Arches should help reduce software maintenance and development costs for Open Context.

8. Conclusion

"It's tough to make predictions, especially about the future."

Yogi Berra

The future is obviously uncertain, and the future for the digital record of the past is even more so. The Trump Administration's sudden (and likely illegal) elimination of funding for both the IMLS and NEH in 2025 highlights the precarity of archaeology in the United States. Direct public support for archaeology through grants will likely be severely curtailed. In addition, the current Administration will likely weaken or ignore enforcement of historical preservation laws and regulations that currently sustain the cultural resource management (CRM) sector. Similarly, economic shocks as well as threats of politically motivated legal and tax-status attacks on universities and nonprofit foundations make it difficult for archaeology to find sources of nongovernmental funding.

US based archaeologists need to find ways of navigating and surviving these increasingly harrowing circumstances. As discussed, Open Context developed from peculiar organizational circumstances to make significant contributions to archaeological data dissemination. However, Open Context is very



much embedded within larger systems. It depends upon externally provisioned digital library services to maintain persistent identifier infrastructure and data preservation services. It also depends upon an international and multidisciplinary community that curates ontologies, vocabularies, and the data Open Context needs to enhance interoperability. Moreover, Open Context also depends upon continued grant funding, either directly or indirectly through the collection of data publishing fees financed through grant supported data management plans.

These supporting institutions, services, funding streams, and community efforts need to thrive to open viable pathways toward longer term sustainability. Open Context, like all other similar services will need to navigate a new period of political instability that threatens public support for research and possibly much more. Use of internationally distributed infrastructure (such as Zenodo) may help build greater resilience, at least for the data already published by Open Context. By continuing to participate in and support the wider data curation community, Open Context hopes to weather looming institutional and political risks.

Appendix: Open Context FAQs

What kind of data is in Open Context?

Open Context publishes structured data (typically from tabular sources like spreadsheets and relational databases), maps, media, and field notes from archaeological research, excavations, and surveys.

Is Open Context a data repository?

No, Open Context is a data publisher, but we regularly accession published data into digital repositories. Open Context receives repository services for data archiving from the California Digital Library, Zenodo, and the Internet Archive.

Do I have to pay to use data published in Open Context?

No, Open Context is free of charge for all to use. We expect users to abide by Open Context's Terms of Use and Intellectual Property policies.

Do I have to pay to publish my data?

Most data contributors pay to publish their data. We have a sliding scale of fees, depending on the size and complexity of each project. To keep costs down, we subsidize our data publishing services with grant funding.

Can I publish my data for free?

In some cases, we will publish data sets at no charge, depending on our current grant funded project needs or alignment with our strategic goals.

Do I need to log in to use Open Context?

No, Open Context is open and free to all users. To avoid risks to user privacy, we do not track or monitor user activity.

Can I publish my data myself (is there a self-service option)?



No, we have no self-service option to upload files. All data publications go through an editorial process that involves editing and data migration (ingest). There is frequent back and forth between data authors and Open Context's editors to prepare data for publication.

How do I publish data with Open Context?

For every data publication, we work one-on-one with data contributors. Contact us to start a conversation about publishing your data.

What kinds of files do you accept?

We prefer that data contributors send us tabular data (CSV, Excel, or equivalent). We also accept images, maps, 3D models, and occasionally PDFs. We accept data in open formats or that can be viewed using open-source software.

What files do you not accept?

- We are not a repository for print publications, so we do not accept PDFs of published papers. Rather, we link to them from your data publication.
- We do not accept data that requires proprietary software for viewing. All data must be viewable using open-source software options.
- We do not accept data that the data contributor does not have permission to publish. Data authors must secure all permissions and agree to Open Context's Terms of Use and Intellectual Property policies before submitting data to Open Context.
- We do not accept data that is sensitive or that requires it to be behind an access barrier.

Can I download data from Open Context?

Yes, you can download full data sets or portions of data sets by clicking on a record export button next to the search results.

How do I cite or attribute data I use from Open Context?

Every page in Open Context has a "suggested citation" at the top right of the screen. Always be sure to include the persistent unique identifier (PID) in your citation. This may be a DOI (<https://doi.org/10.6078/...>) or an ARK (<https://n2t.net/ark...>). Open Context PIDs are co-owned by the University of California. If Open Context ceases operations, the University of California can update resolution of PIDs to locations where the data are archived in repositories.

What is a PID?

A PID (persistent unique identifier) is a string of letters and numbers used to distinguish between and locate different objects, people, or concepts on the Internet. PIDs allow things to be specifically and uniquely identified and located. Organizations that issue PIDs commit to the long term and reliable maintenance of these identifiers and their associated content.

How long will my data be available on Open Context?

Open Context archives all data publications with digital repositories. If Open Context ceases operation, all data (including PIDs used to identify specific data records) will still be accessible in bulk from these repositories. However, the repository copies of Open Context data will be bulk-



access only. While the data would be preserved, one would need to load these datasets into appropriate software to retrieve specific data records.

Bibliography

ACRA (American Cultural Resources Association) 2019 'Communication Strategy for "Promoting Synergy Between the Academy and the CRM Industry"', adopted by the Board of Directors, American Cultural Resources Association, January 14. https://www.digitalantiquity.org/wp-uploads/2019/03/2019-ACRA-Comm-Strategy_Promote-synergy-CRM-Academy-approved-ACRA-BoD_01-14-2019.pdf [Last accessed: 18 April 2025]

Altschul, J.H. and Klein, T.H. 2022 'Forecast for the US CRM Industry and Job Market, 2022–2031', *Advances in Archaeological Practice* **10**(4), 355–370. <https://doi.org/10.1017/aap.2022.18>

Anderson, D. G., Wells, J., Yerka, S., Noack Myers, K., DeMuth, R.C and Bissett, T.G. 2015 'Digital Index of North American Archaeology (DINAA)', Sarah Witcher Kansa, Eric C. Kansa (Ed). Released: 2015-10-21. Open Context. <https://doi.org/10.6078/M7N877Q0>

Anderson D.G., Bissett, T.G., Yerka, S.J., Wells, J.J., Kansa, E.C., Kansa, S.W, Myers, K.N., DeMuth, R.C. and White, D.A. 2017 'Sea-level rise and archaeological site destruction: An example from the southeastern United States using DINAA (Digital Index of North American Archaeology)' *PLOS ONE* **12**(11), e0188142. <https://doi.org/10.1371/journal.pone.0188142>

Anderson, J. and Hudson, M. 2020 'The Biocultural Labels Initiative: Supporting Indigenous rights in data derived from genetic resources', *Biodiversity Information Science and Standards* **4**: e59230. <https://doi.org/10.3897/biss.4.59230>

Arbuckle, B. S., Kansa, S.W., Kansa, E.C., Orton, D., Çakırlar, C., Gourichon, L., Atici, L. *et al.* 2014 'Multiple Routes and Husbandry Strategies Characterize the Westward Spread of Early Domestic Livestock across Neolithic Turkey', *PLOS ONE* **9**(6), e99845. <https://doi.org/10.1371/journal.pone.0099845>

Atici, L., Kansa, S.W. and Lev-Tov, J.S.E. 2010 'Chogha Mish Fauna'. Released: 2010-08-24. Open Context. <https://doi.org/10.6078/M7M906K5> ARK (Archive): <https://n2t.net/ark:/28722/k26m3c58g>

Baker, T. and Sutton, S.A. 2015 'Linked data and the charm of weak semantics: Introduction: The strengths of weak semantics', *Bulletin of the Association for Information Science and Technology* **41**, 10–12. <https://doi.org/10.1002/bult.2015.1720410406>

Batist, Z. 2025 'Locating creative agency in archaeological data work', Knowledge Commons, ver.2 *PCI Archaeology* (for preprint peer-review and recommendation). <https://doi.org/10.17613/8eqy4-n1m82>

Bird, D., Miranda, L. , Vander Linden, M., Robinson, E., Bocinsky, R.K., Nicholson, C., Capriles, J. *et al.* 2024 'Cross-referenced p3k14c'. Released: 2024-07-22. Open Context. <https://opencontext.org/projects/cdd78c10-e6da-42ef-9829-e792ce55bdd6> <https://doi.org/10.6078/M7G73BT5>

Brody, A. 2010 'Bade Museum'. Released: 2010-02-16. Open Context. <https://doi.org/10.6078/M7959FQ3> ARK (Archive): <https://n2t.net/ark:/28722/k2gm8qj5g>



- Bruce, R. and Cordewener, B. 2018 'Open Science Is All Very Well but How Do You Make It FAIR in Practice?', *JISC Blog*, 12 July 2018. <https://web.archive.org/web/20200723054643/https://www.jisc.ac.uk/blog/open-science-is-all-very-well-but-how-do-you-make-it-fair-in-practice-12-jul-2018> [Last accessed: 18 December 2025]
- Brunson, K., Li, Z. and Flad, R. 2016 'Oracle Bones in East Asia'. Released: 2016-04-04. Open Context. <https://doi.org/10.6078/M74B2Z7J>
- Caraher, W. and Pettegrew, D.K. 2025 'Mobilizing the Archaeological Report for a Future of Reuse: Linked Open Data and the Scholar-Led Publication', *Journal of Field Archaeology* **50**(8), 782–795. <https://doi.org/10.1080/00934690.2025.2528471>
- Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. 2020 'The CARE Principles for Indigenous Data Governance', *Data Science Journal* **19**(1), 43. <https://doi.org/10.5334/dsj-2020-043>
- Chander, A. and Sunder, M. 2004 'The Romance of the Public Domain', *California Law Review* **92**(5), 1331-1373. <https://doi.org/10.2307/3481419>
- Christen, K.A. 2012 'Does Information Really Want to be Free? Indigenous Knowledge Systems and the Question of Openness', *International Journal of Communication* **6**. <https://ijoc.org/index.php/ijoc/article/view/1618> [Last accessed: 16 March 2026]
- Clark, D.R. and Herr, L.G. in prep "Madaba Plains Project-`Umayri". Open Context. <https://doi.org/10.6078/M7W37TGB>
- Counts, D.B., Walcek Averett, E., Garstki, K. and Toumazou, M. 2020a 'Visualizing Votive Practice: Exploring Limestone and Terracotta Sculpture from Athienou-Malloura through 3D Models'. Released: 2020-07-28. Open Context. <https://doi.org/10.6078/M7X928D9>
- Counts, D.B., Walcek Averett, E., Garstki, K. and Toumazou, M. 2020b *Visualizing Votive Practice: Exploring Limestone and Terracotta Sculpture from Athienou-Malloura through 3D Models*, Digital Press Books **11**. <https://doi.org/10.31356/dpb011>
- Davies, N., Deck, J. Kansa, E.C., Kansa, S.W., Kunze, J., Meyer, C., Orrell, T., Ramdeen, S., Snyder, R., Vieglais, D., Walls, R.L. and Lehnert, K. 2021 'Internet of samples (iSamples): Toward an interdisciplinary cyberinfrastructure for material samples', *GigaScience* **10**(5), p.giab028. <https://doi.org/10.1093/gigascience/giab028>
- D'Itria, E. 2022 'The Amulets of the Kerma Culture'. Released: 2022-09-12. Open Context. <https://doi.org/10.6078/M74747Z0>
- Faniel, I.M., Austin, A., Kansa, E.C., Kansa, S.W., France, P., Jacobs, J. and Yakel, E. 2018 'Beyond the Archive: Bridging Data Creation and Reuse in Archaeology', *Advances in Archaeological Practice* **6**(2), 105–116. <https://doi.org/10.1017/aap.2018.2>
- Felicetti, A., Gerth, P., Meghini, C. and Theodoridou, M. 2015 'Integrating heterogeneous coin datasets in the context of archaeological research' in P. Ronzino and F. Niccolucci (eds) *Extending, Mapping and Focusing the CIDOC CRM (CRMEX 2015) Workshop, 19th International Conference on Theory and Practice of Digital Libraries (TPDL 2015)*, Poznan, Poland, September 17, 2015. <https://ceur-ws.org/Vol-1656/paper2.pdf> [Last accessed: 19 June 2025]



- G8 Open Data Charter 2013 *G8 Open Data Charter and Technical Annex*. <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex> [Last accessed: 21 April 2025]
- Geser, G., Richards, J.D., Massara, F. and Wright, H. 2022 'Data Management Policies and Practices of Digital Archaeological Repositories', *Internet Archaeology* **59**. <https://doi.org/10.11141/ia.59.2>
- GIDA (Global Indigenous Data Alliance) 2020 *CARE Principles for Indigenous Data Governance*, The CARE Principles Research Data Alliance International Indigenous Data Sovereignty Interest Group. <https://www.gida-global.org/care> [Last accessed: March 14, 2023]
- Graham, S., Gupta, N., Smith, J., Angourakis, A., Reinhard, A., Ellenberger, K., Batist, Z., Rivard, J., Marwick, B., Carter, M. and Compton, B. 2019 *The Open Digital Archaeology Textbook* [online] <https://web.archive.org/web/20250426040458/https://o-date.github.io/draft/book/using-application-programming-interfaces-apis-to-retrieve-data.html>
- Hacigüzeller, P., Taylor, J.S. and Perry, S. 2021 'On the emerging supremacy of structured digital data in archaeology: A preliminary assessment of information, knowledge and wisdom left behind', *Open Archaeology* **7**(1), 1709-1730. <https://doi.org/10.1515/opar-2023-0099>
- Harrison, L.K. 2017 'Architecture and Urbanism at Seyitömer Höyük, Turkey'. Released: 2017-05-05. Open Context. <https://doi.org/10.6078/M76W980X>
- Heitman, C.C. 2022 'Theorizing the Archive and Ethics of Open Access Archaeology' in E. Watrall and L. Goldstein (eds) *Digital Heritage and Archaeology in Practice: Data, Ethics, and Professionalism*, Gainesville: University Press of Florida. 220-40.
- Hemphill, L., Pienta, A., Lafia, S. Akmon, D. and Bleckley, D.A. 2022 'How do properties of data, their curation, and their funding relate to reuse?', *Journal of the Association for Information Science and Technology* **73**(10), 1432-1444. <https://doi.org/10.1002/asi.24646>
- Hoffman, N. 2016 'Historic Fort Snelling'. Released: 2016-06-23. Open Context. <https://doi.org/10.6078/M7G44NF9>
- IPinCH Project Team 2014 'Intellectual Property Issues in Cultural Heritage (IPinCH) Project' in C. Smith (ed) *Encyclopedia of Global Archaeology*, Springer, New York, NY. 3921-2924. https://doi.org/10.1007/978-1-4419-0465-2_36
- Jennings, L., Jones, K. Taitingfong, R. *et al.* 2025 'Governance of Indigenous data in open earth systems science', *Nature Communications* **16**, 572. <https://doi.org/10.1038/s41467-024-53480-2>
- Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C. *et al.* 2018 'Data curation network: A cross-institutional staffing model for curating research data', *International Journal of Digital Curation* **13**(1), 125–140. <https://doi.org/10.2218/ijdc.v13i1.616>
- Joukowsky, M.S. 2007 'Petra Great Temple Excavations'. Released: 2007-11-11. Open Context. <https://doi.org/10.6078/M7TB1515> ARK (Archive): <https://n2t.net/ark:/28722/k20r9w91w>
- Kansa, E.C. 2009 'Indigenous Heritage and the Digital Commons' in C. Antons (ed) *Traditional Knowledge, Traditional Cultural Expressions and Intellectual Property Law in the Asia-Pacific Region*, Kluwer Law International: Alphen aan den Rijn. 219–244.



Kansa, E.C. 2012 'Openness and archaeology's information ecosystem', *World Archaeology* **44**(4), 498–520. <https://doi.org/10.1080/00438243.2012.737575>

Kansa, E.C. 2014 'The Need to Humanize Open Science' in S. Moore (ed) *Issues in Open Research Data*, London: Ubiquity Press. <https://doi.org/10.5334/ban.c>

Kansa, E.C. 2016 'Click Here to Save the Past' in E. Walcek Averett, J. Michael Gordon and D.B. Counts (eds) *Mobilizing the Past for a Digital Future: The Potential of Digital Archaeology*, Grand Forks, ND: The Digital Press at the University of North Dakota. 443–472. https://dc.uwm.edu/artist_mobilizingthepast/19/

Kansa, E.C. and S.W. Kansa. 2022 'Promoting data quality and reuse in archaeology through collaborative identifier practices', *Proceedings of the National Academy of Sciences* **119**(43), p.e2109313118. <https://doi.org/10.1073/pnas.2109313118>

Kansa, E.C., Kansa, S.W. and Arbuckle, B. 2014 'Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology', *International Journal of Digital Curation* **9**(1), 57–70. <https://doi.org/10.2218/ijdc.v9i1.301>

Kansa, E.C., Schultz, J. and Bissell, A. 2005 'Protecting Traditional Knowledge and Expanding Access to Scientific Data: Juxtaposing Intellectual Property Agendas via a "Some Rights Reserved" Model', *International Journal of Cultural Property* **12**(3), 285–314. <https://doi.org/10.1017/S0940739105050204>

Kansa, E.C., Kansa, S.W. and Goldstein, L. 2013 'On Ethics, Sustainability, and Open Access in Archaeology', *The SAA Archaeological Record* **13**, 15–22. <https://zenodo.org/records/3373945>

Kansa, E.C., Kansa, S.W., Wells, J.J., Yerka, S.J., Myers, K.N., DeMuth, R.C., Bissett, T.G. and Anderson, D.G. 2018 'The Digital Index of North American Archaeology: Networking Government Data to Navigate an Uncertain Future for the Past', *Antiquity* **92**(362), 490–506. <https://doi.org/10.15184/aqy.2018.32>.

Kansa, S.W., Gupta, N., Martinez, D., Nicholson, C. and the FAIR+CARE Cultural Heritage Network 2025 'The FAIR+CARE Cultural Heritage Network: Advancing Practices to Support Ethical Data Management in Archaeology', *SAA Archaeological Record* **25**(4), 20–26. https://onlinedigeditions.com/publication/?i=852385&article_id=5041123&view=articleBrowser

Katsianis, M., Bruseker, G., Nenova, D., Marlet, O., Hivert, F., Hiebel, G., Ore, C.-E., Derudas, P., Opitz, R. and Uleberg, E. 2023 'Semantic Modelling of Archaeological Excavation Data. A review of the current state of the art and a roadmap of activities', *Internet Archaeology* **64**. <https://doi.org/10.11141/ia.64.12>

Kealhofer, L., Grave, P. and Marsh, B. 2017 'Differentiating local from nonlocal ceramic production at Iron Age Sardis using NAA'. Released: 2017-09-04. Open Context. <https://doi.org/10.6078/M74Q7S2X>

Kintigh, K.W. 2006 'The Promise and Challenge of Archaeological Data Integration', *American Antiquity* **71**(3), 567–578. <https://doi.org/10.2307/40035365>.

Koster, L. 2020 'Persistent identifiers for heritage objects', *Code4Lib Journal* **47**. <https://journal.code4lib.org/articles/14978> [Last accessed: 16 March 2026]



- Lee, D.J. and Stvilia, B. 2017 'Practices of research data curation in institutional repositories: A qualitative view from repository staff', *PLoS One* **12**(3), e0173987. <https://doi.org/10.1371/journal.pone.0173987>
- Lehner, M. 2017 'ARCE Sphinx Project 1979-1983 Archive'. Megan Flowers, Rebekah Miracle (ed). Released: 2017-11-14. Open Context. <https://doi.org/10.6078/M7CZ356B>
- Marx, A., Salas Rossenbach, K. and Bryas, E. 2021 'Digital Archiving and Data Stewardship in French Archaeology', *Internet Archaeology* **58**. <https://doi.org/10.11141/ia.58.26>
- Marsolek, W., Wright, S.J., Luong, H., Braxton, S.M., Carlson, J. and Lafferty-Hess, S. 2023 'Understanding the value of curation: A survey of researcher perspectives of data curation services from six US institutions', *PLoS One* **18**(11), e0293534. <https://doi.org/10.1371/journal.pone.0293534>
- McManamon, F.P. and Kintigh, K.W. 2010 'Digital Antiquity: Transforming Archaeological Data into Knowledge', *SAA Archaeological Record* **10**(2), 37–40. <https://www.tdar.org/images/SAA-McM-Kintigh.pdf> [Last accessed: 16 March 2026]
- Mirowski, P. 2018 'The future(s) of open science', *Social Studies of Science* **48**(2), 171-203. <https://doi.org/10.1177/0306312718772086>
- Moore, A., Hillman, G. and Legge, A. 2025 'Abu Hureyra Digital Data Archive'. Released: In prep. Open Context. <https://doi.org/10.6078/M7862DMQ>
- Muñoz, T. 2013 'Data Curation as Publishing for the Digital Humanities', *Journal of Digital Humanities* **2**(3), 14–22. <https://journalofdigitalhumanities.org/2-3/data-curation-as-publishing-for-the-digital-humanities/> [Last accessed: 16 March 2026]
- Nature 2013 'Announcement: Reducing our Irreproducibility', *Nature* **496**(398). <https://doi.org/10.1038/496398a>
- Nicholas, G. and Bannister, K. 2004 'Copyrighting the past? Emerging intellectual property rights issues in archaeology' *Current Anthropology* **45**(3), 327-350. <https://doi.org/10.1086/382251>
- Nicholson, C., Fernandez, R. and Irwin, J. 2021 'Digital Archaeological Data in the Wild West: the challenge of practising responsible digital data archiving and access in the United States', *Internet Archaeology* **58**. <https://doi.org/10.11141/ia.58.22>
- Novák, D., Oniszczuk, A. and Gumbert, B. 2023 'Digital Archaeological Archiving Policies and Practice in Europe: the EAC call for action', *Internet Archaeology* **63**. <https://doi.org/10.11141/ia.63.7>
- OMB (Office of Management and Budget) 2015 *Project Open Data*. <https://project-open-data.cio.gov/> [Last accessed: 20 December 2015] / <http://project-open-data.github.io/> [Last accessed: 6 January 2024]
- Okumura, M. and Araujo, A. 2024 'The Physics and Metaphysics of Classification in Archaeology' in *Between variability and singularity: crossing theoretical, qualitative and computer-based approaches to types and typologies in archaeology*, Zenodo. <https://doi.org/10.5281/zenodo.14559070>
- Opitz, R., Mogetta, M. and Terrenato, N. 2017 '2843 from Europe/Italy/Gabii/Area B/Unit 1468'. In *The Gabii Project*. Rachel Opitz, Marcello Mogetta, Nicola Terrenato (Ed). Released: 2017-06-04. Open Context. ARK (Archive): <https://n2t.net/ark:/28722/k2rn3cz83>



Ortman, S. and Altschul, J. 2023 'What North American Archaeology Needs to Take Advantage of the Digital Data Revolution', *Advances in Archaeological Practice* **11**(1), 90-103. <https://doi.org/10.1017/aap.2022.42>

OSTP (Office of Science and Technology Policy) 2014 *Improving the Management of and Access to Scientific Collections, memorandum, March 20, 2014. Washington, DC.* https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_memo_scientific_collections_march_2014.pdf [PDF][Last accessed: Jan. 4 2024]

Parker, B. and Cobb, P. 2012 'Kenan Tepe'. Released: 2012-02-27. Open Context. <https://doi.org/10.6078/M7H41PBJ> ARK (Archive): <https://n2t.net/ark:/28722/k2xw47z1d>

Pettegrew, D.K. 2024 *Corinthian Countrysides: Linked Open Data and Analysis from the Eastern Korinthia Archaeological Survey*, The Digital Press at the University of North Dakota. <https://thedigitalpress.org/corinthiancountrysides/> [Last accessed: 16 March 2026]

Price, M. 2016 'Northern Mesopotamian Pig Husbandry (Late Neolithic-EBA)'. Released: 2016-03-01. Open Context. <https://doi.org/10.6078/M71J97N1>

Radchenko, S., Nykonenko, D., Konovalov, D., Chernata, Y., Hordiienko, T., Prokopenko, S., and Yashchenko, P. 2023 '3D History for Ukraine: Archaeological Museum of the Taras Shevchenko National University of Kyiv'. Released: 2023-09-28. Open Context. <https://doi.org/10.6078/M7K935NG>

Richards, J.D. 2023 'Joined up Thinking: Aggregating archaeological datasets at an international scale', *Internet Archaeology* **64**. <https://doi.org/10.11141/ia.64.3>

Ryan, N. 2004 'Databases', *Internet Archaeology* **15**. <https://doi.org/10.11141/ia.15.8>

Sheehan, B. 2015 'Comparing Digital Archaeological Repositories: tDAR Versus Open Context', *Behavioral & Social Sciences Librarian* **34**(4), 173–213. <https://doi.org/10.1080/01639269.2015.1096155>

Stopic, A., Marsh, B., Bennett, J.W., Seeher, J., Kealhofer, L., Grave, P. and Schoop, U-D. 2013 'Ceramics, Trade, Provenience and Geology: Cyprus in the Late Bronze Age'. Released: 2013-09-21. Open Context. <https://doi.org/10.6078/M70V89RM>

Tuck, A. 2012 'Murlo'. Released: 2012-07-06. Open Context. <https://doi.org/10.6078/M77P8W98> ARK (Archive): <https://n2t.net/ark:/28722/r2p24>

Tudhope, D., May, K., Binding, C. and Vlachidis, A. 2011 'Connecting Archaeological Data and Grey Literature via Semantic Cross Search', *Internet Archaeology* **30**. <https://doi.org/10.11141/ia.30.5>

Vines, T.H., Andrew, R.L., Bock, D.G., Franklin, M.T., Gilbert, K.J., Kane, N.C., Moore, J-S., Moyers, B.T., Renaut, S., Rennison, D.J., Veen, T. and Yeaman, S. 2013 'Mandated data archiving greatly improves access to research data', *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* **27**(4), 1304–1308. <https://doi.org/10.1096/fj.12-218164>

Wells, J., Wiley, T., Finnegan, P., Chamorro, V., Edmonds M. and Kansa, E.C. 2017 "Digital Index of North American Archaeology, Linking Sites and Literature". In *Digital Index of North American Archaeology (DINAA)*. Open Context. <https://doi.org/10.6078/M7542KRZ>

The White House 2013 *Executive Order --Making Open and Machine Readable the New Default for Government Information*, May 9, 2013. <https://obamawhitehouse.archives.gov/the-press->



[office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-](#) [Last accessed: 6 January 2024]

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J. *et al.* 2016 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* **3**, 160018. <https://doi.org/10.1038/sdata.2016.18>

Yakel, E., Faniel, I.M., and Maiorana, Z.J. 2019 'Virtuous and vicious circles in the data life-cycle', *Information Research* **24**(2), paper 821. <http://InformationR.net/ir/24-2/paper821.html>

Yoon, A. and Lee, Y.Y. 2019 'Factors of trust in data reuse', *Online Information Review* **43**(7), 1245-1262. <https://doi.org/10.1108/OIR-01-2019-0014>