

UC Berkeley

UC Berkeley Previously Published Works

Title

Archaeological Analysis in the Information Age: Guidelines for Maximizing the Reach, Comprehensiveness, and Longevity of Data

Permalink

<https://escholarship.org/uc/item/3m35127j>

Journal

Advances in Archaeological Practice, 8(1)

Authors

Kansa, Sarah W

Atici, Levent

Kansa, Eric C

et al.

Publication Date

2020

Peer reviewed

Archaeological Analysis in the Information Age: Guidelines for Maximizing the Reach, Comprehensiveness, and Longevity of Data

Sarah W. Kansa , Levent Atici, Eric C. Kansa, and Richard H. Meadow

ABSTRACT

With the advent of the Web, increased emphasis on “research data management,” and innovations in reproducible research practices, scholars have more incentives and opportunities to document and disseminate their primary data. This article seeks to guide archaeologists in data sharing by highlighting recurring challenges in reusing archived data gleaned from observations on workflows and reanalysis efforts involving datasets published over the past 15 years by Open Context. Based on our findings, we propose specific guidelines to improve data management, documentation, and publishing practices so that primary data can be more efficiently discovered, understood, aggregated, and synthesized by wider research communities.

Keywords: zooarchaeology, data management, reproducible research, data documentation, guidelines, data reuse

Con el advenimiento de la Web, el mayor énfasis en el “manejo de datos de investigación” y las innovaciones en las prácticas de investigación reproducibles, los investigadores tienen más incentivos y oportunidades para documentar y divulgar sus datos primarios. Este trabajo busca ofrecer a los arqueólogos una guía sobre cómo compartir información, señalando los desafíos más recurrentes en la reutilización de datos archivados que recopilamos a partir de observaciones sobre flujos de trabajo y del reanálisis de bases de datos publicadas a lo largo de los últimos 15 años en Open Context. Sobre la base de nuestros resultados, proponemos algunos lineamientos específicos para mejorar el manejo de datos, la documentación y las prácticas de publicación, de modo que los datos primarios puedan ser descubiertos, comprendidos, agrupados y sintetizados de manera más eficiente y por comunidades de investigadores más amplias.

Palabras clave: arqueozoología, investigación reproducibles, manejo de datos, documentación de datos, lineamientos, reutilización de datos

Archaeologists, particularly those specializing in zooarchaeology and other domains that emphasize quantitative methods, have long recognized that scientific rigor requires access to full and well-documented data (Driver 1992; Grigson 1978; Meadow and Zeder 1978). Only recently, however, have technologies caught up to the vision of full data description and dissemination. In addition to the decreasing cost of data storage, the Web offers new opportunities to efficiently and comprehensively document and disseminate our work. Data dissemination is becoming an expected part of scholarly communication, such that open access to publicly funded research results is now encouraged and even mandated by many governments and institutions (Kimbrough and Gasaway 2015; ScienceEurope 2019; White House 2013). Even so, while policies calling for data dissemination have gained traction, actual data sharing practices remain largely piecemeal and inconsistent. For data sharing to facilitate reproducibility and to open new research opportunities that involve reusing shared data,

we must overcome significant challenges in archiving, curating, disseminating, and integrating diverse datasets.

The challenges and opportunities afforded by computers and the Web require us to reexamine normative data management practices and consider both new approaches and refinements of extant practices. In this article, informed by years of data management experience and integrative data analyses, we describe some of the problems inherent in current, normative practices of data collection and documentation in archaeology. These problems can inhibit future data reuse by imposing costs in terms of the time and effort needed to clean, translate, and understand other researchers’ datasets. We advocate integrating conventional publishing with new forms of data publication and version control to provide open access to primary datasets. We then offer a set of guidelines for data collection and dissemination in the “digital age.” These are practical steps that practitioners can take to

improve their data collection and documentation to better leverage these new forms of communication and access.

Our primary aim is to provide guidelines relevant to the new forms of data documentation and dissemination that have been enabled by technological advances. While we draw on examples from zooarchaeology, these guidelines are relevant to anyone working with archaeological data. Recognizing that changing workflows in response to these guidelines may be difficult and time-consuming, we conclude with a list of incremental steps that archaeologists can take toward creating “5 Star” (best-practice) data. We must note that this work is not intended to supplant previous guidelines but to supplement them in order to consider new technologies and opportunities for data dissemination and reuse. For example, other aspects of the practice of zooarchaeology, such as preservation, sampling bias, and recovery methods, have been covered in depth by others (see, among others, Driver 1992; Grigson 1978; Peres 2010; Reitz and Wing 2008). The guidelines presented here build on the existing body of work on good practices while adding new insights based on our own experiences working with archaeological and zooarchaeological data integration.

THE PROBLEM: DISCONNECTS BETWEEN DATA COLLECTION AND REUSE

Archaeology often relies on destructive methods. Furthermore, archaeological projects frequently face funding limitations and physical storage constraints that can lead, for example, to the discard of partial or complete faunal assemblages recovered during fieldwork. Because physical specimens and contexts may not be afforded long-term conservation, the discipline urgently needs strategies to preserve the digital documentation of physical objects and features that may no longer exist. In addition to addressing ethical demands for better stewardship, many argue that greater data sharing can open new research opportunities to address larger-scale “big picture” research questions (Altschul et al. 2018; Kansa 2011; Kansa and Kansa 2013; Kintigh et al. 2014a, 2014b, 2015). If multiple kinds of data can be efficiently and meaningfully aggregated, it will become possible to discern and explore new spatial and temporal patterns using analytically rigorous methods. Finally, data sharing can promote greater scientific reproducibility by making the evidence that underlies interpretive claims more open to inspection (Marwick 2017). This goal could also naturally foster adherence to more consistent and comparable approaches to the recording and reuse of data. Indeed, such arguments to promote data sharing have played a profound role in shaping public policy, and many private and public funders of archaeology now require grant seekers to supply data management plans as part of their proposals.

While data sharing policies and initiatives have strong justification, we have yet to see many significant research outcomes from such efforts (but see Anderson et al. 2017; Hammer and Ur 2019; Richards 2017; Styles and Colburn 2019). Entrenched reporting practices, together with intense competition for scarce funding and employment, all work to inhibit data dissemination. Moreover, simply making data accessible may not be enough to encourage reuse. We need to develop better methods to model, create,

clean, and document more usable data. We also need broader consensus among researchers about what constitutes good data sharing practices, including development of common vocabularies as advocated by Beebe (2017). That consensus needs to be accompanied by more professional recognition and reward for excellence in data usability and sharing.

MATERIALS AND METHODS

Archaeologists collect a complex array of descriptive data, which can be organized and modeled in database schemas in idiosyncratic ways by different researchers and research projects. While there is no “one size fits all” solution for data management, there are common cross-cutting needs and requirements to organize analytically useful data. We identify these common needs and outline a baseline set of good data management practices. These guidelines bring together the authors’ experiential observations working with two sources of qualitative data over the past decade. Aspects of both have been published in various venues over the years, but this is the first time that lessons from these sources have been integrated into a set of practical guidelines.

Data Source 1: Open Context Data Publications

Since 2006, Open Context has provided open data publishing services for archaeology. Its 120 data publications contain more than 1.5 million individual data items. These are integrated into a common unified database and archived with the University of California’s California Digital Library (CDL) and the Zenodo repository (CERN). The data publishing process follows a workflow that emphasizes editorial services for cleanup, documentation, and curation of data. Open Context editors work closely with data authors to decode data, document data with appropriate description and metadata, and facilitate cross-referencing and linking to relevant data published across the Web. This data publishing work over the past 15 years has illuminated the great diversity of data collection and documentation practices in archaeology.

Data Source 2: Observations on Zooarchaeological Data Integration Efforts

In 2014, with funding from the Encyclopedia of Life and the National Endowment for the Humanities, the authors participated in a project, led by Benjamin Arbuckle, to integrate archaeological data for the study of the origin and spread of domestic animals in Anatolia. This group, the Central and Western Anatolia Neolithic Working Group (hereafter, the Anatolia Zooarchaeology Project), was the first of its kind to bring together dozens of individuals to share and reuse data from many sites and to publish the final datasets in Open Context. The Anatolia Zooarchaeology Project integrated more than 500,000 specimens from 42 chronological phases and 17 archaeological sites in Turkey. Arbuckle played a key role in recruiting participants for the study and overseeing the data analysis (Arbuckle et al. 2014), while the Open Context team worked with the data authors on data cleanup and annotation in order to make the datasets interoperable (for more on this process, see Kansa et al. 2014; Yakel et al. 2019). Our efforts to clean, align, and integrate these datasets over the course of several

months, together with lessons from publishing 120 datasets in Open Context, inform the discussion in the Results section.

RESULTS: OBSERVATIONS ON DATA FRICTIONS IN THE “DIGITAL AGE”

These two data curation and integration efforts provide clear evidence for the many entrenched and normative data creation and sharing practices that inhibit and undermine future analyses. Here we specify the pervasive and persistent practices that negatively impact data reuse.

Use of Spreadsheets for Data Collection Negatively Affects Future Data Reuse

Database technologies, especially relational databases, can be powerful tools for organizing and modeling the kinds of complex observations made by archaeologists. However, in our experience, most researchers prefer to record data on spreadsheets. The specific layout of a researcher’s spreadsheet is typically tailored to the project at hand, often based on how it was collected in notebooks decades ago (Faniel et al. 2013). In cultural resource management (CRM), data practices may be dictated by the CRM company or may not be structured at all, varying from project to project. Because archaeologists often lack training in database design and management, they may create idiosyncratic datasets or enter data in inconsistent ways that may change over time as methods are added or refined. Lack of training in databases and the “home grown” nature of resulting datasets is also evident in the way that scholars model (layout and structure) their data.

In a surprising number of cases, data creators do not provide unique identifiers for individual specimens in their datasets (ex. “Specimen #PC-1094”). This creates ambiguity that can complicate quantification. In zooarchaeology, for example, without unique specimen identifiers, it is difficult to know if two identical records from a single context (that both read, for example, “sheep/goat radius, shaft fragment”) refer to two different specimens, or if these records were duplicated accidentally. More troubling, without unique identification of specific specimens, it becomes impossible to “join” together multiple descriptions of specimens. For example, one table may describe taxonomic classifications made for bone specimens, and another, perhaps created by a different researcher, may provide taphonomic observations for the same specimens. If these two tables do not use a common set of unique identifiers, the different sets of descriptions cannot be brought together. While this may seem obvious, many zooarchaeological datasets lack unique identifiers, perhaps because data creators assume that no such “joins” of additional information will ever occur.

Other aspects of data modeling can also impede analysis and reuse. We commonly see poor organization of observations. Observations (or “descriptive attributes”) often need to be broken into more modular components. In the Anatolia Zooarchaeology Project, although all participants described the fusion status of bone elements, the way that they entered them into their spreadsheets varied greatly. Some entered data descriptions into a single cell while others split them into multiple cells. When the time comes to share those data, these variations make it very

TABLE 1. Incremental Data Management Practices for Archaeology.

Star Rating	Criteria
(No Stars)	<ul style="list-style-type: none"> • Specimen-level data are not publicly available • Data are shared only in summary tables in publications • Data are shared via e-mail
★	<ul style="list-style-type: none"> • Data are decoded • Structured data are publicly accessible on the Web • Data have a standard copyright license permitting reuse (e.g., Creative Commons attribution)
★★	<ul style="list-style-type: none"> • Structured data are in a digital repository with a persistent identifier (DOI) • Data have appropriate discovery and attribution metadata
★★★	<ul style="list-style-type: none"> • Data records include all expected (common) fields (such as context, specimen ID, taxon, element, age, sex, measurements, etc.)
★★★★	<ul style="list-style-type: none"> • Data integrity, internal consistency, structural coherence, and adherence to best practices are established (numeric values with numbers only, controlled vocabulary terms consistent) • Data author has created a paradata document and linked it to the dataset (clearly describing data collection and sampling methods, analytical procedures, reference materials, confidence in identifications)
★★★★★	<ul style="list-style-type: none"> • Data use stable Web identifiers to reference relevant community-curated recording attributes, controlled vocabularies, and ontologies • Data use stable Web identifiers to document archaeological context • Data use stable Web identifiers to document paradata • Data use stable Web identifiers to document reference specimens used • Data and data documentation have been checked by another subject matter expert and tested for interoperability

difficult to reuse and combine different datasets. As an example of better practice (described in Table 1) “*Ovis aries* humerus” should not occur in only a single cell but should be split into multiple fields. Making descriptive attributes more modular also makes it much easier to create more consistent data with common data validation techniques such as picklists and drop-downs.

Many of the problems we observed in how zooarchaeologists model data result directly from the severe constraints that spreadsheets impose on data structures. Spreadsheets typically keep data as self-contained “flat” tables. However, zooarchaeologists often need to use a different set of observations for different types of specimens. For example, each skeletal element can be described with a different set of measurements. Tables can become very large and unwieldy with different columns dedicated to recording attributes applicable to only a small number of records. Similarly, flat tabular data structures lead to awkward modeling of “many-to-many” relationships between specimens and parts of

specimens. For example, articulations between different specimens, or the relationships between individual teeth and mandibles, are all difficult to describe if modeled in a single spreadsheet. In such cases, zooarchaeologists often use free-form text comment fields to express these relationships, even though such fields are difficult to present consistently and quantitatively.

Good research practices require appropriate data modeling. For zooarchaeology, a relational database approach typically provides better support for the field's data modeling requirements (Jones and Hurley 2011). However, we do not want to imply that good practice *requires* use of a relational database. After all, database and other software technologies vary and continually evolve, and spreadsheets can be an effective solution if used properly. Rather than focusing on the details of software implementations, the discipline needs to apply greater attention, professionalism, and rigor to data modeling so that the appropriate data management software is used appropriately.

Status Quo Data Sharing Practices Impede or Complicate Reuse

Even in the “digital age,” archaeologists continue to share synthetic and/or secondary data in summary tables in the published literature. Although such summary tables may be sufficiently detailed to support the claims made in an article, sharing only summarized results may constrain future reuse. More granular sharing of research data offers more analytic flexibility. Granular data also can be more easily merged, compared with other data, and aggregated in different ways using different combinations of attributes.

Another common and harmful data sharing practice is the one-to-one exchange of information between two researchers, usually via e-mail. With no public version of record, datasets can be altered or modified in undocumented ways. Similarly, the lack of a public version of record may increase risks of “scooping” because informally shared data lack publicly maintained attribution and citation information (for a list of advantages and disadvantages of different modes of data sharing, see Kansa and Kansa 2013:Table 1). Finally, informally shared data can increase the risk of favoritism, whereby a researcher may grant preferential data access to allies but not to perceived competitors. Thus, access to well-documented, public (published) data is fundamental to both reproducibility and ethical practice.

Use of Codes Leads to Information Loss

In the early days of data entry, coding offered an efficient way of recording large amounts of information, speeding data entry, and reducing storage requirements when computer memory was low (e.g., Meadow 1978; Redding et al. 1978; Uerpmann 1978). Furthermore, adoption of codes encourages more careful consideration of data recording methods because all possible entries must have a code (Driver 2011). Thus, using codes can speed data entry and produce data that is more consistent and easier to manipulate. However, it also carries the risk of data loss, should the code sheet become disassociated from the data. Furthermore, researchers use a diversity of alphanumeric codes, often tailor-made for their projects, and these codes may be extended or updated over time, resulting in a dataset that can diverge from a code sheet.

In our work with the Anatolia Zooarchaeology Project datasets, and indeed, throughout our years of work publishing archaeological data in Open Context, we have found that although decoding can be straightforward with one-to-one pairing of code to term, the decoding process can be complicated by interdependencies between codes or by custom amendments to the codebook. In one case, a researcher adopted an existing coding system, but added some new codes over the course of the recording. These amendments, which had been written directly on the paper codebook, were not incorporated into the digital version of the canonical codebook that was cited. In addition, the decoding process is time-consuming for reusers, who may only be seeking to quickly assess the dataset for its suitability in their research (see Kansa et al. 2014). Thus, to maintain the data's integrity and facilitate its reuse, decoding should be done by the original analyst whenever possible.

Methodological “Standards” See Uneven Implementation

Because data sharing is still not routine, researchers lack consensus on how to align the organization of their databases with the recording practices they may consider to be “standard.” Zooarchaeologists routinely describe specimens according to a number of common and shared recording attributes for measurements, tooth-wear, eruption patterns, etc. (see, among others, Behrensmeyer 1978; Boessneck 1969; Driesch 1976; Payne 1969; Uerpmann 1978). However, the specific implementation and modeling of these often-complex attributes varies from dataset to dataset, thereby complicating interoperability. Figure 1 illustrates how this challenge plays out in measurement data. In this example, although many zooarchaeologists use the measurement system developed by Angela von den Driesch (1976), individual analysts record measurement attributes in various ways to facilitate their own data entry. Some enter all measurements in one cell while others allocate each measurement a separate cell and still others make each measurement name a field header.

From a data reuse perspective, idiosyncratic and sometimes problematic data modeling can make large-scale data integration difficult, requiring translation and time-consuming reorganization of datasets before reuse becomes possible. Datasets published with Open Context undergo an editorial process during which the data are checked for such inconsistencies and are then organized and documented to facilitate future understanding and reuse. Ideally, as the discipline becomes more aware of how data modeling affects data reuse, this will cultivate a “virtuous cycle” (see Yakei et al. 2019)—a process whereby increased data sharing leads to better sharing of good data modeling practices and better opportunities for reuse, thereby motivating more researchers to share more and better-modeled data.

DISCUSSION: IMPROVED DATA COLLECTION, DOCUMENTATION, AND REUSE

New research based on the aggregated analysis of large-scale data requires reusable data. To work toward this goal, we need strategies to incrementally bootstrap archaeology toward greater professionalism and deeper engagement with data. Unless the

Data Model 1:

Specimen ID	Identification	Measurements
1	<i>Ovis aries</i> humerus	Bd: 32.6; SD: 15.4
2	<i>Sus</i> astragalus	GLm: 42.1

Data Model 2:

ID	Genus	Species	Element	Measurement 1	Measurement 2
1	<i>Ovis</i>	<i>aries</i>	Humerus	Bd: 32.6	SD: 15.4
2	<i>Sus</i>		Astragalus	GLm: 42.1	

Data Model 3:

Bone #	Taxon	Bone	Bd	SD	GLm
Bone 1	Sheep	Humerus	32.6	15.4	
Bone 2	Pig	Talus			42.1

Data Model 4:

ID	Species	Element	Meas. 1 Name	Meas. 1 Value	Meas. 2 Name	Meas. 2 Value
1	<i>Ovis aries</i>	humerus	Bd	32.6	SD	15.4
2	<i>Sus</i>	astragalus	GLm	42.1		

Data Model 5:

Bone #	Taxon	Element	Measurement Name	Measurement Value
1	<i>Ovis aries</i>	humerus	Bd	32.6
1	<i>Ovis aries</i>	humerus	SD	15.4
2	<i>Sus</i>	astragalus	GLm	42.1

Data Model 6:

ID	Taxon	Element
1	<i>Ovis aries</i>	Humerus
2	<i>Sus</i>	Astragalus

Primary Key	Foreign Key Bone	Type	Value
1	1	Bd	32.6
2	1	SD	15.4
3	2	GLm	42.1

FIGURE 1. Although analysts may use the same standards to record their data, variations in the way that they model their data (the layout of their datasets and the terms they use) can make the data difficult to use and compare. This example shows six different ways that simple observations on two specimens might be recorded. Although all models record the same set of attributes, the specific way that the data are organized, as well as the different terms that are used, lead to trade-offs in data entry and data analysis. Notice that in Data Model 5, the Specimen ID key indicates that *the same specimen* has different measurement attributes recorded on different rows. For Data Model 6, the specimens are modeled in a relational database, where the taxon and element are linked to their measurements in a different table (the green highlighted columns show the link across database tables).

reuse of data in a wider community itself becomes a professional goal, simply tacking on a few tentative data sharing and archiving measures will not necessarily promote new and better research. A host of norms and practices need to change to make data dissemination a more vibrant aspect of research.

We propose the following general guidelines for good practices that, if considered widely, should help facilitate data sharing, reuse, and preservation. They are intentionally broad to facilitate application to most projects, and they should be seen as building

upon—and in addition to—archaeological data sharing guidance provided by others (such as Atici et al. 2013:Tables 1–3; Marwick and Birch 2018:Table 8). These guidelines will certainly see modifications in the future as new technologies and data management systems emerge and as practitioners expand on them.

General Guidelines

1. *Provide data in open, structured formats.* Do not limit your data sharing to modes optimized for presentation (such as PDFs or

images of figures). Even full data tables provided on paper (or the digital analogue of paper) are of limited use because reusers often must transcribe them (particularly in the case of printed material), wasting time and making errors along the way (Dibble 2015). Provide data in open and common formats convenient to use with spreadsheets, databases, and other software.

2. *Provide archaeological context.* The archaeological context must always accompany every item in a dataset. To maximize the reuse potential of the dataset, contextual information should include temporal information and spatial coordinates (specific or with reduced precision, if specific location data involve security risks). These data should be for the site itself, as well as for the individual contexts and how those contexts relate to each other spatially and temporally. Sites and contexts need to be identified explicitly (see point 5) in ways that facilitate cross-referencing with other information sources that further document related sites and contexts.
3. *Develop a data management plan for every project.* Project directors and specialists should develop data management plans at the outset of every project. The plan can serve as a memorandum of understanding, detailing the mutual needs and intentions of the project and the analyst, including, where relevant, input from descendant communities or other stakeholders. These plans can also be part of contract agreements in public or commercial projects. Plans for data management should include all stages of the data lifecycle from excavation to publication and/or other forms of digital dissemination. Revisit the plan frequently, update it, and share it with others to build community consensus on what characterizes a good data management plan.
4. *Provide analytical context.* Contextual information also pertains to the context of the data creation and analysis. This “paradata” (or what Huggett 2014 terms “data provenance”) includes detailed information about the excavation of the remains, the analyst’s training and expertise, where analysis took place, which methods and reference materials were used, how the dataset was modified, etc. Include and clearly link to your dataset a document containing this paradata, as such high-quality documentation can serve as a “trust marker” for data reuse (Faniel and Yakel 2017). This may be published and cited as a separate peer-reviewed paper describing the detailed methods (the “systematic paleontology” described by Wolverton [2013]) or shared online and linked to the dataset (as advocated by Lau and Kansa [2018] and as exemplified in the “data cleaning protocol” of Atici [2013]). A key component of paradata is reporting sampling biases, which helps reusers understand missing data. For example, if your dataset has no fish bones, is it because none were present in the excavated contexts or because somebody bagged them separately or sent them to another analyst? Are some records redacted because of data sensitivity or security concerns? In his 1992 paper, and in his additional comments included in the paper’s republication two decades later, Driver (2011) gives examples of how different approaches to data collection and analysis can affect research results. The issue of the context of data creation becomes critical when the dataset becomes the main object being studied—that is, when it is used for new research beyond the original researcher’s focus. *It is essential for the data reuser to understand more about the data than just the data themselves.* Indeed, a recent study involving interviews with archaeologists about data sharing practices concluded

that analytical context is the foremost concern of data reusers (Faniel et al. 2013). There is general agreement that for many research questions, the absence of well-documented contextual information renders a dataset nearly useless. However, this does not mean that data with minimal contextual information should not be shared. Sometimes even limited analysis on a dataset can be informative (Atici et al. 2013; Jones and Gabe 2015), particularly for less extensively studied regions or chronological periods. Furthermore, future analysts may find effective uses for datasets with minimal contextual documentation, so at the very least they should be archived. In all cases, however, it is critical that the person accessing such datasets is aware of their limitations.

5. *Use identifiers to promote interoperability.* “Sharing data” is not simply a matter of uploading a spreadsheet onto a website, creating a “supplementary materials” section of a publication, or even depositing data in an archive. Researchers seeking to aggregate data need evidence that such data can be considered commensurate and comparable. Widespread use of common standards can help improve the interoperability of data. Standards should be globally unambiguous, persistent, and easy to de-reference (find and retrieve). In current best practice, this means participating in Linked Open Data, where data, concepts, and other “entities” are identified by stable Web URLs. That is, the Web address does not just point to a location on the Web (such as <https://google.com>) but also identifies a specific thing, such as an individual database record or a term in a classification system (Figure 2). Linked Open Data is generally discussed in highly specialized digital library and software engineering contexts, remote from most practicing archaeologists. While full implementation of Linked Open Data (and Semantic Web) technologies requires a great deal of technical expertise and investment, one essential aspect of Linked Open Data practice—namely, the use of stable Web URLs as identifiers—is easily within the reach of archaeologists. Preparing a Linked Open Data set can be as simple as adding stable Web URLs to new columns in a spreadsheet so that related data can be found and classification terms can be understood (Figure 3; also see examples provided in Kansa 2015; LeFebvre et al. 2019). Ideally, stable Web URLs are institutionally backed by libraries, publishers, and other research organizations. Such organizations support persistent URL schemes that enable creation of stable Web URLs (such as ARKs or DOIs). Stable Web URLs can be used to link the paradata document that describes methods used in the creation of an archaeological dataset to the journal publication that discusses those data, and vice versa. Furthermore, any subsequent publications that draw on the paradata can be linked to it, thus building a robust body of methodologically connected research. If different data creators share common elements of paradata, and if they document their paradata with common identifiers, then they will explicitly signal that their data are more readily comparable, thus facilitating discovery and reuse. Finally, reference to common standards does not preclude researchers from innovating or custom-tailoring recording protocols for more nuanced needs specific to a given project. Local systems of meaning and recording should exist alongside more global approaches that have achieved community-wide consensus. However, ideally both global and local systems of recording should use stable Web URLs for explicit identification.

Uberon multi-species anatomy ontology

Class: scapula

Term IRI: http://purl.obolibrary.org/obo/UBERON_0006849

Definition: Endochondral bone that is dorsoventrally compressed and provides attachment site for muscles of the pectoral appendage. [database_cross_reference: <https://github.com/obophenotype/uberontypes/issues/339>][database_cross_reference: <https://github.com/obophenotype/uberontypes/issues/339>][database_cross_reference: <https://github.com/obophenotype/uberontypes/issues/339>][database_cross_reference: <https://github.com/obophenotype/uberontypes/issues/339>]

Annotations



- has_alternative_id: UBERON:0001124; UBERON:0010004
- has_exact_synonym: scapulae; hypercoracoides@fr; scapula@fr; hypercoracoidesum@fr
- has_narrow_synonym: shoulder blade
- has_obo_namespace: uberon
- has_related_synonym: scapula bone; scapulae
- has_relational_adjective: scapular
- <http://www.geneontology.org/formats/obo/owl#id:UBERON:0006849>
- in_subset: http://purl.obolibrary.org/obo/uberon/core#uberon_slim; http://purl.obolibrary.org/obo/uberon/core#uberon_slim; http://purl.obolibrary.org/obo/uberon/core#uberon_slim
- taxon_notes: present in all tetrapods with even vestiges of anterior limbs, e.g., turtles & birds & mammals.

Dama dama (Linnaeus 1758)



OPEN CONTEXT About Explore Data Record Contact Donate Search whole site

Context: Turkey / Öküzini Cave / VIII / Square 18d / Bone 18d22-61 (Animal Bone)

Main Observation Standards Annotations

Annotations (6)

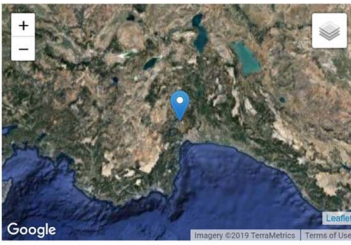
Property or Relation	Value(s)
Has anatomical identification [Standard: Open Context Zooarchaeology Annotations]	scapula Q [Standard: UBERON (Uber Anatomy Ontology)]
Has Biological Taxonomy [Standard: Biological Taxonomy Vocabulary]	Dama dama (Linnaeus, 1758) Q [Standard: Encyclopedia of Life]
Has physiological sex determination [Standard: Open Context Zooarchaeology Annotations]	Nonidentified Q [Standard: Open Context :: Zooarchaeology of Öküzini Cave]
Has fusion character [Standard: Open Context Zooarchaeology Annotations]	Distal epiphysis fused Q [Standard: Open Context Zooarchaeology Annotations]
BG [Standard: Open Context Zooarchaeology Annotations]	24.5
GLP [Standard: Open Context Zooarchaeology Annotations]	37.2
Editorial Note	Open Context editors work with data contributors to annotate datasets to shared vocabularies, ontologies, and other standards using 'Linked Open Data' (LOD) methods. The annotations presented above approximate some of the meaning in this contributed data record to concepts defined in shared standards. These annotations are provided to help make datasets easier to understand and use with other datasets.

Suggested Citation

Levent Atici. "Bone 18d22-61 from Turkey / Öküzini Cave / VIII / Square 18d". (2013) In *Zooarchaeology of Öküzini Cave*. Levent Atici (Ed.). Released: 2013-02-26. Open Context. <<http://opencontext.org/subjects/AD54136C-E747-40B8-7191-73929AF0349B>> ARK (Archive): <https://n2t.net/ark:/28722/k2445nr96>

Editorial Status: Peer-reviewed (4 stars)

Part of Project: Zooarchaeology of Öküzini Cave

Mapping Data: 

Open Context editors work with data contributors to annotate datasets to shared vocabularies, ontologies, and other standards using **Linked Open Data (LOD)** methods. The annotations presented above approximate some of the meaning in this contributed data record to concepts defined in shared standards. These annotations are provided to help make datasets easier to understand and use with other datasets.

FIGURE 2. A zooarchaeological specimen published in Open Context, showing annotations that precisely identify how descriptive attributes link elsewhere on the Web to concepts defined by expert communities. Datasets that use Web identifiers to cross-reference to a common set of concepts will be more discoverable, easier to understand, and more interoperable.

6. *Be explicit with copyright.* Copyright licenses may seem esoteric, but without them, current intellectual property law in many international jurisdictions would preclude reuse of certain datasets without negotiating specific permissions with their owners (Kansa 2012). The lack of copyright licenses has a

similar impact as the statement “data will be made available upon request.” Essentially, the lack of a clear and standard license means data languishes in a “you can look but not touch” state. Many options exist for archaeological data, including the suite of licenses offered by Creative Commons

(a)


	A	B	C	D	E	F	G
1	Site	Period	Trench	Locus	Bone Cat #	Taxon	Element
2	Poggio Civitate	Archaic (580 BCE - 482 BCE)	T32	11	PC-00001	Bos taurus	Humerus
3	Poggio Civitate	Archaic (580 BCE - 482 BCE)	T32	11	PC-00002	Ovis aries	Horn core
4	Poggio Civitate	Archaic (580 BCE - 482 BCE)	T32	11	PC-00003	Ovis/Capra	Mandible
5	Poggio Civitate	Archaic (580 BCE - 482 BCE)	T32	11	PC-00004	Sus scrofa	Scapula
6	Poggio Civitate	Archaic (580 BCE - 482 BCE)	T32	11	PC-00005	Capra hircus	Phalanx 1



FIGURE 3a. An example of adding Linked Open Data to a dataset.

(b)

	A	B	C	D	E	F	G	H	I	J	K
1	Site	Pleiates URI	Period	PeriodO URI	Trench	Locus	Bone Cat #	Taxon	EOL URI	Element	UBERON URI
2	Poggio Civitate	https://pleiades.org/Archaic	Archaic (580 BCE - 482 BCE)	http://n2t.net/Archaic	T32	11	PC-00001	Bos taurus	http://eol.org/p/Bos_taurus	Humerus	http://purl.obolibrary.org/obo/uberon:0000001
3	Poggio Civitate	https://pleiades.org/Archaic	Archaic (580 BCE - 482 BCE)	http://n2t.net/Archaic	T32	11	PC-00002	Ovis aries	http://eol.org/p/Ovis_aries	Horn core	http://purl.obolibrary.org/obo/uberon:0000002
4	Poggio Civitate	https://pleiades.org/Archaic	Archaic (580 BCE - 482 BCE)	http://n2t.net/Archaic	T32	11	PC-00003	Ovis/capra	http://eol.org/p/Ovis_capra	Mandible	http://purl.obolibrary.org/obo/uberon:0000003
5	Poggio Civitate	https://pleiades.org/Archaic	Archaic (580 BCE - 482 BCE)	http://n2t.net/Archaic	T32	11	PC-00004	Sus scrofa	http://eol.org/p/Sus_scrofa	Scapula	http://purl.obolibrary.org/obo/uberon:0000004
6	Poggio Civitate	https://pleiades.org/Archaic	Archaic (580 BCE - 482 BCE)	http://n2t.net/Archaic	T32	11	PC-00005	Capra hircus	http://eol.org/p/Capra_hircus	Phalanx 1	http://purl.obolibrary.org/obo/uberon:0000005

 OPEN CONTEXT

Annotation Vocabularies

- Biological Taxonomy Vocabulary
- British Museum Thesaurus
- CIDOC-CRM
- Concordia
- Digital Index of North American Archaeology Vocabulary
- Dublin Core Terms
- Eastern Woodlands Household Archaeology Data Project
- ★ Encyclopedia of Life
- GeoNames
- Getty Art and Architecture Thesaurus
- Information Artifact Ontology
- Kerameikos.org
- Levantine Ceramics Project
- Library of Congress Subject Headings
- Open Context (General)

- Open Context Zooarchaeology Annotations
- ORCID
- Paleoindian Database of the Americas (PIDBA)
- ★ PeriodO
- ★ Pleiades
- RDF-Schema
- Simple Knowledge Organization System (SKOS)
- tDAR Site Name Keywords
- ★ UBERON (Uber Anatomy Ontology)
- Units of Measurement Ontology
- Wikidata
- WikiMapia
- Wikipedia
- WorldCat

FIGURE 3b. A list of some of the annotation vocabularies used by Open Context to link data across the Web, indicating those used in the example dataset in Figure 3a.

(<https://creativecommons.org/licenses/>) and the strategies and Traditional Knowledge Labels developed by LocalContexts (<http://localcontexts.org/tk-labels/>; see also Kansa et al. 2005).

7. *Integrate data ethics from the beginning.* Improving the transparency of data improves the overall transparency of the entire research process. This makes analyses more reproducible and credible. Greater data transparency can also lead to

greater public recognition and attribution for many different collaborators, including student researchers, who may participate in a given project. At the outset, expectations for proper credit and attribution should be negotiated, and data should be modeled appropriately to provide the provenance of different contributions. Such issues of recognition must be broadly considered. Archaeologists often work with communities that have survived brutal colonial histories. Efficient interoperability, while often desirable, may not always be the ethical priority. Some animal remains, as is often the case for human remains, can have special significance among members of descendant communities. Archaeologists need to collaborate with community members to understand the specific cultural context of data collection in order to make more inclusive decisions regarding data management, including dissemination and archiving practices. Nevertheless, although decisions about access need to be made inclusively, other concerns about data quality, documentation, and interoperability still apply. Data with access restrictions should still be managed to high standards.

Zooarchaeology-Specific Guidelines

In addition to the more general guidelines we have provided, we have developed a set of zooarchaeology-specific guidelines for better aligning data creation with the needs of longer-term reuse. We group these guidelines into three stages of data production: (1) planning and recording, (2) dissemination and archiving, and (3) reuse and training. Professional communities should promote even more detailed guidelines and curate them in publicly accessible venues so that discussions and changes to the guidelines are clearly documented. For example, the International Council for Archaeozoology might maintain one or more sets of “living” guidelines in a version-controlled forum like GitHub.

We must reiterate that these guidelines are meant to build on the guidance previously published in zooarchaeology by providing additional considerations for large-scale aggregation. Although these guidelines can be applied to new projects, projects currently underway, and old (legacy) datasets, the extent to which they can be applied depends on a number of factors including access to the original analyst, availability of contextual information, and the amount of time and effort needed to clean up a dataset.

PLANNING AND RECORDING

1. *Validate data from the start.* Develop and implement better recording, collection, and documentation strategies from a project’s inception. Tools such as drop-down menus limit recording errors and improve data validation (ensuring the accuracy and quality of data). Equally important, explore how ongoing projects can retrospectively resolve some of the ambiguities in their data. This may involve tactics such as breaking up cells with multiple items in them (e.g., if multiple analysts are listed in one cell separated by commas, move them into separate cells). Data surveying to ensure data integrity and cleanup can be performed at every stage of a project, not just at its inception. In the future, all the changes can be incorporated to accommodate these new fields.
2. *Consider reuse from the start.* Data collection and recording protocols should facilitate later downstream reuse by the wider research community. We refer zooarchaeologists to the report by Grigson (1978) and criteria listed by Atici and colleagues (2013). The former introduces a blueprint with information reported in three “must have” sections: (1) introduction with notes on the site, excavation, conservation, and recording of bone specimens; (2) presentation of the basic data; and (3) interpretation of the data (Grigson 1978:121). The latter offers a similar three-tiered approach to reporting data with (1) criteria for evaluating and improving data integrity, (2) archaeological data sharing criteria, and (3) common variables for zooarchaeology-specific data sharing (Atici et al. 2013:Tables 1–3). Accordingly, zooarchaeologists should associate the information they deem basic and essential with their spreadsheet or database either as a separate linked database or in a worksheet in the same database.
3. *Integrate data from the start.* Data recording protocols and database management design should be prioritized from the beginning of the project in consultation with project leadership and other specialists. Too often, zooarchaeological and other “specialist” datasets end up fragmented and isolated because they do not easily relate to data in the project’s master database (Faniel et al. 2018). Archaeological context and other identifiers must exactly align to promote greater “referential integrity,” meaning greater reliability in the links between records of zooarchaeological and other project data.
4. *Other good recording practices:*
 - Provide unique identifiers that are easy to look up and retrieve for specimens in a comparative collection (if used), including online comparands. Museums and other institutions hosting reference collections should assign globally unique and persistent identifiers to specimens in order to facilitate such cross-referencing.
 - Do not put multiple values for an attribute in the same cell. For example, when recording osteometric data, do not cluster all measurements for one specimen in a single cell; rather, enter each measurement into its own cell.
 - If you plan to include analytic uses of references to other data records, avoid notations that describe a range of values (“Bones 1 to 4”). These kinds of text descriptions may be easy to write but are very difficult to parse reliably and consistently for computational analyses.
 - To facilitate computational operations and permit the use of pivot tables and other worksheet features, each row in a worksheet should describe a single specimen (with columns housing the designated variables such as taxon, element, age, etc.).
 - Take a conservative approach to identifications by identifying specimens to the degree *actually* possible, not *ideally* possible. For instance, if not confident about a taxonomic identification at the species level, analysts should refer to a more general taxonomic category such as genus, family, order, or class. That is, they should feel confident in their assignment, even if it is simply to a general size category such as “large animal.” Zooarchaeologists should either identify a specimen to any level when confident or denote the specimen as “nonidentified.” They should avoid adding question marks such as “Bos?”. Rather, if uncertainty cannot be eliminated for some of the nonessential zooarchaeological attributes (see Atici et al. 2013:Table 3), use a “Comments” field for noting uncertainties.

Zooarchaeologists should practice this conservative approach when recording other variables, such as element, and refrain from guesswork and over-identifications to avoid compromising the integrity of the analysis.

DISSEMINATION AND ARCHIVING

1. *Decode all data*, but keep a version of the original coded data, as well as the code sheet. Although translating coded data requires a great deal of effort, a decoded dataset never has to be decoded again. Thus, over the long term, decoding early in the lifecycle of a dataset saves a great deal of time and effort. Importantly, it also helps resolve questions or ambiguity in the dataset, ideally in consultation with the original data author.
2. *Provide primary data for age observations*. All tooth-wear stages should be left in their original form, and an additional column added if the analyst wishes to add an interpretive level (such as age in months or years).
3. *Document project metadata*. Define field headings, avoid abbreviations and acronyms, indicate which unit of measurement you used, and cite standards.
4. *Share full, structured data (not just summary tables)*. “Structured data” have a logical organization, are clearly defined, and can be easily accessed and analyzed with a spreadsheet, database, or other software. More granular and specific structured data enable greater freedom and flexibility for analysis using different combinations of attributes describing individual records.
5. *Publish the details of your work* that cannot be accommodated by conventional publications as a paradata paper in a journal or online, which you can cite and link to both your raw datasets and your published analyses. Ideally, this paradata paper is a “living” document (using a version control service such as GitHub) that describes the methods and the context of the data collection. Create a field(s) in your database where you can link to the paradata document for each specimen. If you change or update your methods, update the paradata document and create a new version.
6. *Plan for data dissemination before* the conventional publication comes out so you can incorporate the stable citable link to your data into your publication.

TRAINING AND REUSE

1. *Promote and share good methods and practices* and train others. Scholars in positions of teaching zooarchaeology need to practice and teach appropriate forms of data modeling. Shared data will provide examples of good approaches to data modeling that others can emulate and refine.
2. *Be an informed reuser*. Use your knowledge of good data management to assess a dataset’s integrity before you begin to reuse it. Surveying the dataset to identify potential errors/problems/limitations followed by data cleaning and standardization can significantly increase the potential of reuse by others. Read the dataset’s documentation and search the published literature to see how the dataset has been used in research.
3. *Promote better norms in data citation*. Cite the data you reuse as you would any other publication, ideally using globally unique persistent identifiers (such as DOIs; Marwick and Birch 2018).

4. *Relate your data to other data*. Consider using Linked Open Data to relate specimens in your dataset to the wider world of data available on the Web. This does not mean that you no longer will have the freedom to define new ways to describe data. Rather, it requires adding field(s) where you can add links to common classifications, such as those provided by Global Biodiversity Information Facility (www.gbif.org) or UBERON (<http://uberon.github.io/>), in order to link classifications across projects. Kansa (2015) describes this process for zooarchaeology. Using linked data not only helps cross-reference data to common classification systems, it can also help communicate spatial and geographic context.
5. *Request data access*. Current professional evaluation often centers on publication in peer-reviewed journals. If these journals, editors, and peer reviewers required authors (where ethically appropriate) to share the data behind their analyses and offered guidance on how to do so, then many more archaeologists would participate in data sharing. Such measures, which have been advocated by Marwick and Birch (2018) would promote greater quality, credibility, and trust in the science presented in the literature. Such measures would also help increase the amount of useful data available for reanalysis. Likewise, data outputs, where relevant, should be assessed as part of tenure and promotion reviews. While the individual being assessed could make a case for their data outputs, committee members should update requirements to include data. Recent guidelines such as those endorsed by the SAA (Driver et al. 2018) and the Archaeological Institute of America (AIA 2018) provide key information to help prepare for such assessments. Finally, archaeologists working in CRM should become attentive to company or project data collection methods, as increasing professional and institutional expectations for data access may affect future updates to data collection in this sector.

CONCLUSIONS: TOWARD “5 STAR” DATA IN ARCHAEOLOGY

This article aims to promote elements of good practice to improve the reproducibility of archaeological research, particularly zooarchaeology, and to facilitate analytically robust research that involves the aggregation, integration, and synthesis of data. While outlining a vision for more ideal data management practices, we recognize that established norms, workflows, incentives, technical knowledge, and habits cannot change overnight. Tim Berners-Lee, one of the key early architects of the World Wide Web, similarly realized that many researchers would need to incrementally adopt certain Linked Open Data practices. He promoted a simple and widely referenced “5 Star” checklist of criteria that describe key steps to “encourage people—especially government data owners—along the road to good linked data” practices (Berners-Lee 2006). Adapting this “5 Star” approach for archaeological data can similarly help better communicate expectations. Rather than setting unreasonably high expectations that make the “perfect the enemy of the good,” we suggest a simple path toward incrementally adopting better data management practices (Table 1).

The practices we recommend in this article can help ensure that key information is associated with the data so that any form of reuse will be better informed. Essential elements of good practice include

- sharing data at the most basic interpretive level; that is, as specific and granular as possible, in anticipation of its reuse;
- linking published data to a conventional publication of your interpretive analysis in order to give it some context; and
- budgeting time in your research to developing a solid data management plan, cleaning and decoding your data before dissemination, linking your data to terms and ontologies to disambiguate its meaning, and documenting your work in a detailed paradata document.

Finally, we strongly emphasize the importance of viewing data as a first-class research outcome that is as important as, if not more important than, the interpretive publications that result from their analysis. Guidelines like these should help guide peer-review processes in the academic sector and, hopefully, contracting requirements in the commercial sector, so that good data management practices can become a more normal and expected aspect of archaeological practice. Expertise about good practice exists, but in highly specialized settings. We need community-wide fluency with the basics of good data practices to make data management meaningful. Wider data literacy also means more researchers need to have the skills and expertise to better use available data. Advancing community-developed good practices for the documentation, dissemination, and reuse of primary datasets will lead to improved understanding and increased trust in data created by others. In turn, this will lead to better integration of datasets and thus more opportunities to address broader and more significant research questions. We encourage the global archaeology community to consider these guidelines and develop data documentation and preservation policies for our discipline.

Acknowledgments

This research has been supported in part by a grant from the National Endowment for the Humanities (PK-50072-08). Many of the topics discussed in this paper are further explored in ongoing research, funded by another National Endowment for the Humanities grant (PR-234235-16). Any views, findings, conclusions, or recommendations expressed in this article do not necessarily represent those of the National Endowment for the Humanities. We thank the reviewers for their helpful comments and suggestions.

Data Availability Statement

There is no primary data related directly to this study; however, we do cite datasets used in previously published works. These are referenced in the bibliography.

REFERENCES CITED

- Altschul, Jeffrey H., Keith W. Kintigh, Terry W. Klein, William H. Doelle, Kelley A. Hays-Gilpin, Sarah A. Herr, Timothy A. Kohler, Barbara J. Mills, Lindsay M. Montgomery, Margaret C. Nelson, Scott G. Ortman, John N. Parker, Matthew A. Peeples, and Jeremy A. Sabloff
2018 Fostering Collaborative Synthetic Research in Archaeology. *Advances in Archaeological Practice* 6:19–29.
- Anderson, David G., Thaddeus G. Bissett, Stephen J. Yerka, Joshua J. Wells, Eric C. Kansa, Sarah W. Kansa, Kelsey Noack Myers, R. Carl DeMuth, and Devin A. White
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. DOI:10.1371/journal.pone.0188142.
- Arbuckle, Benjamin S., Sarah Witcher Kansa, Eric C. Kansa, David Orton, Canan Çakırlar, Lionel Gourichon, Levent Atici, Alfred Galik, Arkadiusz Marciniak, Jacqui Mulville, Hijkje Buitenhuis, Denise Carruthers, Bea De Cupere, Arzu Demirergi, Sheelagh Frame, Daniel Helmer, Louise Martin, Joris Peters, Nadja Pöllath, Kamilla Pawłowska, Nerissa Russell, Katheryn Twiss, and Doris Würtenberger
2014 Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey. *PLoS ONE* 9(6):e99845. DOI:10.1371/journal.pone.0099845.
- Archaeological Institute of America (AIA)
2018 *Considerations Regarding the Tenure and Promotion of Classical Archaeologists Employed in Colleges and Universities, with Addendum, Guidelines for the Evaluation of Digital Technology and Scholarship in Archaeology*. Electronic document, <https://www.archaeological.org/programs/professionals/careers/tenure/>, accessed July 8, 2019.
- Atici, Levent
2013 Data Cleaning Protocol from Iran / Chogha Mish. In *Chogha Mish Fauna*, edited by Levent Atici, Sarah W. Kansa, and Justin Lev-Tov. Open Context. Electronic document, <http://opencontext.org/media/1B937408-370B-4221-FA7AA2787B6C4BF1>, accessed September 23, 2019; ARK: <https://n2t.net/ark:/28722/k2dj59g74>, accessed September 23, 2019.
- Atici, Levent, Sarah Kansa, Justin Lev-Tov, and Eric Kansa
2013 Other People's Data: A Demonstration of the Imperative of Publishing Primary Data. *Journal of Archaeological Method and Theory* 20:663–681.
- Beebe, Caroline
2017 Standard Descriptive Vocabulary and Archaeology Digital Data Collection. *Advances in Archaeological Practice* 5:250–264.
- Behrensmeyer, Anna K.
1978 Taphonomic and Ecologic Information from Bone Weathering. *Paleobiology* 4:150–162.
- Berners-Lee, Tim
2006 Linked Data. Electronic document, <https://www.w3.org/DesignIssues/LinkedData.html>, accessed June 8, 2019.
- Boessneck, Joachim
1969 Osteological Differences between Sheep (*Ovis aries* Lineé) and Goat (*Capra hircus* Lineé). In *Science in Archaeology*, 2nd ed., edited by Don R. Brothwell and Eric S. Higgs, pp. 331–358. Thames & Hudson, London.
- Dibble, W. Flint
2015 Data Collection in Zooarchaeology: Incorporating Touch-Screen, Speech-Recognition, Barcodes, and GIS. *Ethnobiology Letters* 6:249–257. DOI:10.14237/eb1.6.2.2015.393.
- Driesch, Angela von den
1976 *A Guide to the Measurement of Animal Bones from Archaeological Sites*. Bulletin 1. Peabody Museum of Archaeology and Ethnology, Harvard University, Cambridge, Massachusetts.
- Driver, Jonathan C.
1992 Identification, Classification and Zooarchaeology. *Circaea* 9:35–47.
2011 Identification, Classification and Zooarchaeology. *Ethnobiology Letters* 2:19–39.
- Driver, Jon, Ted Boebel, Lynne Goldstein, P. Nick Kardulias, Fred Limp, Heather Richards-Rissetto, and LuAnn Wandsnider
2018 Report of the SAA Task Force on Guidelines for Promotion and Tenure for Archaeologists in Diverse Academic Roles. White paper, released January 31, 2018. Society for American Archaeology. Electronic document, https://documents.saa.org/container/docs/default-source/doc-careerpractice/promotion-and-tenure-white-paper-02-12-19-web.pdf?sfvrsn=8a5815cd_4, accessed September 23, 2019.
- Faniel, Ixchel, Anne Austin, Eric Kansa, Sarah Kansa, Phoebe France, Jennifer Jacobs, Ran Boytner, and Elizabeth Yakei
2018 Beyond the Archive: Bridging Data Creation and Reuse in Archaeology. *Advances in Archaeological Practice* 6:105–116. DOI:10.1017/aap.2018.2.

- Faniel, Ixchel, Eric Kansa, Sarah W. Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel
2013 The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. In *JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 295–304. ACM, New York. DOI:10.1145/2467696.2467712.
- Faniel, Ixchel M., and Elizabeth Yakel
2017 Practices Do Not Make Perfect: Disciplinary Data Sharing and Reuse Practices and Their Implications for Repository Data Curation. In *Curating Research Data, Volume One: Practical Strategies for Your Digital Repository*, edited by Lisa R. Johnston, pp. 103–126. Association of College and Research Libraries, Chicago. Electronic document, http://www.ala.org/acrl/sites/ala.org/acrl/files/content/publications/booksand-digitalresources/digital/9780838988596_crd_v1_OA.pdf, accessed September 23, 2019.
- Grigson, Caroline
1978 Towards a Blueprint for Animal Bone Reports in Archaeology. In *Research Problems in Zooarchaeology*, edited by D. R. Brothwell, Kenneth D. Thomas, and Juliet Clutton-Brock, pp. 121–128. University College London Institute of Archaeology, London.
- Hammer, Emily, and Jason Ur
2019 Near Eastern Landscapes and Declassified U2 Aerial Imagery. *Advances in Archaeological Practice* 7:107–126. DOI:10.1017/aap.2018.38.
- Huggett, Jeremy
2014 Promise and Paradox: Accessing Open Data in Archaeology. In *Proceedings of the Digital Humanities Congress 2012: Studies in the Digital Humanities*, edited by Clare Mills, Michael Pidd, and Esther Ward. Digital Humanities Institute, Sheffield, UK. Electronic document, <https://www.dhi.ac.uk/openbook/chapter/dhc2012-huggett>, accessed September 23, 2019.
- Jones, Emily Lena, and Caroline Gabe
2015 The Promise and Peril of Older Collections: Meta-Analyses and the Zooarchaeology of Late Prehistoric/Early Historic New Mexico. *Open Quaternary* 1(1):Art. 6. DOI:10.5334/oq.ag.
- Jones, Emily Lena, and David A. Hurlay
2011 Relational Databases and Zooarchaeology Education. *SAA Archaeological Record* 11(1):19–21.
- Kansa, Eric
2011 New Directions for the Digital Past. In *Archaeology 2.0: New Approaches to Communication and Collaboration*, edited by Eric C. Kansa, Sarah Whitcheer Kansa, and Ethan Watrall, pp. 1–26. Cotsen Institute of Archaeology Press, Los Angeles.
2012 Openness and Archaeology's Information Ecosystem. *World Archaeology* 44:498–520.
- Kansa, Eric C., and Sarah W. Kansa
2013 We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication. *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 1:88–98.
- Kansa, Eric C., Sarah W. Kansa, and Benjamin Arbuckle
2014 Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology. *International Journal of Digital Curation* 9(1):57–70. DOI:10.2218/ijdc.v9i1.301.
- Kansa, Eric C., Jason Schultz, and Ahrash N. Bissell
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* 12:285–314. DOI:10.1017/S0940739105050204.
- Kansa, Sarah W.
2015 Using Linked Open Data to Improve Data Reuse in Zooarchaeology. *Ethnobiology Letters* 6:224–231. DOI:10.14237/ebi.6.2.2015.
- Kimbrough, Julie L., and Laura N. Gasaway
2015 Publication of Government-Funded Research, Open Access, and the Public Interest. *Vanderbilt Journal of Entertainment and Technology Law* 18:267–302.
- Kintigh, Keith W., Jeffrey H. Altschul, Mary C. Beaudry, Robert D. Drennan, Ann P. Kinzig, Timothy A. Kohler, W. Fredrick Limp, Herbert D. G. Maschner, William K. Michener, Timothy R. Pauketat, Peter Peregrine, Jeremy A. Sabloff, Tony J. Wilkinson, Henry T. Wright, and Melinda A. Zeder
2014a Grand Challenges for Archaeology. *American Antiquity* 79:5–24.
2014b Grand Challenges for Archaeology. *Proceedings of the National Academy of Sciences* 111(3):879–880. DOI:10.1073/pnas.1324000111.
- Kintigh, Keith W., Jeffrey H. Altschul, Ann P. Kinzig, W. Fredrick Limp, William K. Michener, Jeremy A. Sabloff, Edward J. Hackett, Timothy A. Kohler, Bertram Ludäscher, and Clifford A. Lynch
2015 Cultural Dynamics, Deep Time, and Data: Planning Cyberinfrastructure Investments for Archaeology. *Advances in Archaeological Practice* 3:1–15.
- Lau, Hannah K., and Sarah W. Kansa
2018 Zooarchaeology in the Era of Big Data: Contending with Interanalyst Variation and Best Practices for Contextualizing Data for Informed Reuse. *Journal of Archaeological Science* 95:33–39.
- Marwick, Ben
2017 Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory* 24:424–450.
- Marwick, Ben, and Suzanne E. Pilaar Birch
2018 A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing. *Advances in Archaeological Practice* 6:125–143. DOI:10.1017/aap.2018.3.
- Meadow, Richard H.
1978 “Bonecode”—A System of Numerical Coding for Faunal Data from Middle Eastern Sites. In *Approaches to Faunal Analysis in the Middle East*, edited by Richard H. Meadow and Melinda A. Zeder, pp. 169–186. Peabody Museum, Harvard University, Cambridge, Massachusetts.
- Meadow, Richard H., and Melinda A. Zeder
1978 *Approaches to Faunal Analysis in the Middle East*. Peabody Museum, Harvard University, Cambridge, Massachusetts.
- Michelle J., LeFebvre, Laura Brenskelle, John Wiczorek, Sarah W. Kansa, Eric C. Kansa, Neill J. Wallis, Jessica N. King, Kitty F. Emery, and Robert Guralnick
2019 ZooArchNet: Connecting Zooarchaeological Specimens to the Biodiversity and Archaeology Data Networks. *PLoS ONE* 14(4):e0215369. DOI:10.1371/journal.pone.0215369.
- Payne, Sebastian
1969 A Metrical Distinction between Sheep and Goat. In *The Domestication and Exploitation of Plants and Animals*, edited by Peter J. Ucko and G. W. Dimbleby, pp. 295–305. Duckworth, London.
- Peres, Tanya M.
2010 Methodological Issues in Zooarchaeology. In *Integrating Zooarchaeology and Paleoethnobotany: A Consideration of Issues, Methods, and Cases*, edited by Amber M. VanDerwarker and Tanya M. Peres, pp. 15–36. Springer, New York.
- Redding, Richard H., Melinda A. Zeder, and John McArdle
1978 “BONESORT II”—A System for the Computer Processing of Identifiable Faunal Material. In *Approaches to Faunal Analysis in the Middle East*, edited by Richard H. Meadow and Melinda A. Zeder, pp. 135–147. Peabody Museum, Harvard University, Cambridge, Massachusetts.
- Reitz, Elizabeth J., and Elizabeth S. Wing
2008 *Zooarchaeology*. 2nd ed. Cambridge University Press, Cambridge.
- Richards, Julian
2017 Twenty Years Preserving Data: A View from the United Kingdom. *Advances in Archaeological Practice* 5:227–237. DOI:10.1017/aap.2017.11.
- ScienceEurope
2019 Plan S: Making Full and Immediate Open Access a Reality. Electronic document, <https://www.coalition-s.org/>, accessed July 23, 2019.
- Styles, Bonnie W., and Mona Colburn
2019 Taphonomic, Environmental, and Cultural Influences on Archaic Faunal Assemblages at Modoc Rock Shelter, Illinois, USA. *Quaternary International*, in press. DOI:10.1016/j.quaint.2019.03.015.
- Uerpmann, Hans-Peter
1978 The KNOCOD System for Processing Data on Animal Bones from Archaeological Sites. *Approaches to Faunal Analysis in the Middle East*, edited by Richard H. Meadow and Melinda A. Zeder, pp. 149–167. Peabody Museum, Harvard University, Cambridge, Massachusetts.

White House

2013 Executive Order: Making Open and Machine Readable the New Default for Government Information. Electronic document, <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>, accessed April 18, 2019.

Wolverton, Steve

2013 Data Quality in Zooarchaeological Faunal Identification. *Journal of Archaeological Method and Theory* 20:381–396. DOI:10.1007/s10816-012-9161-4.

Yakel, Elizabeth, Ixchel M. Faniel, and Zachary J. Maiorana

2019 Virtuous and Vicious Circles in the Data Life-Cycle. *Information Research* 24(2). Electronic document, <http://www.informationr.net/ir/24-2/paper821.html>, accessed September 23, 2019.

AUTHOR INFORMATION

Sarah W. Kansa (sarahkansa@gmail.com, corresponding author) and **Eric C. Kansa** ■ Open Context, 125 El Verano Way, San Francisco, CA 94127, USA

Levent Atici ■ Department of Anthropology, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy., Las Vegas, NV 89154, USA

Richard H. Meadow ■ Department of Anthropology, Harvard University, 11 Divinity Avenue, Cambridge, MA 02138, USA