



PROJECT MUSE®

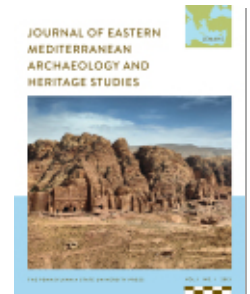
---

## We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication

Eric C. Kansa, Sarah Witcher Kansa

Journal of Eastern Mediterranean Archaeology and Heritage Studies,  
Volume 1, Number 1, 2013, pp. 88-97 (Article)

Published by Penn State University Press



➔ For additional information about this article

<https://muse.jhu.edu/article/501744>



## FORUM

### OPEN ARCHAEOLOGY

# We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication

**Eric C. Kansa**

University of California,  
Berkeley and Open Context,  
Berkeley, CA 94720,  
[ekansa@ischool.berkeley.edu](mailto:ekansa@ischool.berkeley.edu)

**Sarah Whitcher  
Kansa**

The Alexandria Archive Institute,  
San Francisco, CA 94127,  
[skansa@alexandriaarchive.org](mailto:skansa@alexandriaarchive.org)

#### ABSTRACT

Archaeologists create vast amounts of data, but very little sees formal dissemination. This failure points to several dysfunctions in the current structures of archaeological communication. The discipline urgently requires better data professionalism. Current technologies can help ameliorate this, but scholars generally lack the time and technological know-how to disseminate data. We put forth a model of “data sharing as publishing” as a means to address the concerns around data dissemination.

Online digital humanities collections have grown rapidly in scope and significance over the past decade. Museums, archives, and libraries increasingly make their collections available on the Web. In spite of these advances, data produced by individual researchers sees little dissemination. In archaeology, a discipline that relies upon destructive research methods, lack of information sharing not only inhibits scholarship, but also represents a tragic loss of irreplaceable cultural and historical knowledge. The discipline urgently requires a more professional approach if researchers are to make credible and replicable knowledge claims and act as better stewards of cultural heritage. Furthermore, more responsible

data-handling will make archaeological research more efficient, accessible to a wider audience, and more likely to support innovation and new research opportunities. Current technologies can help archaeologists to achieve these ends, but scholars generally lack the time and technological know-how to disseminate data in a meaningful and lasting way.

That dissemination and archiving are not yet common practice largely stems from institutional inertia. There are few professional channels of communication capable of supporting massive data sharing, largely because of a perceived lack of demand. Professionals work under a mandate of “publish or perish,” and publication is narrowly defined as conventional books and articles published by mainstream scholarly presses. If knowledge sharing practices are to become widespread, researchers must see clear evidence that data sharing is worth their time and effort. They need to be convinced that data dissemination will help further their career goals and contribute substantively toward knowledge creation (Harley et al. 2010). In fact, these two challenges are deeply interrelated; synthesizing disparate collections can open new research opportunities that cross disciplinary boundaries (Onsrud and Campbell 2007). In other words, with more data shared and available, research opportunities should expand.

An increasing number of initiatives are fostering transparency in scholarship. The National Science Foundation (NSF) and the National Endowment for the Humanities (NEH) have invested in developing technologies, standards, and other means to enable researchers to

access the relevant work of others in their fields. Creative Commons has led the debate over intellectual property legal requirements for free and open reuse of data.<sup>1</sup> Increasingly, government agencies are being lobbied by groups that want to see more openness and sharing of federally funded research. The NSF and the NEH have already adopted stricter data management requirements to encourage greater data sharing.

We argue here for a new model of “data sharing as publication” in order to address the technological, ethical, and professional concerns surrounding archaeological data distribution today. This model works best where existing workflows and norms of scholarly communication can be applied to the dissemination of structured data. To reach the point where researcher data can be used by a wider community, datasets must have sufficient quality and documentation. To give context, data also need to be related and linked with shared concepts and with other datasets available on the Web. Finally, appropriate workflows can enhance datasets by applying linked open data (see below) dissemination methods.

Research outcomes that apply linked open data will demonstrate to the professional community how data sharing as publication can both recognize individual scholarly contributions and create information resources with far greater capacity to enhance understanding than possible when research remains in isolation. Doing this requires effort, new skills, professional roles, and the creation of scholarly communication channels to meet the specific requirements of meaningful data dissemination. In the end, these efforts promise to increase professional acceptance of data sharing, thus ensuring that the research results are available today and preserved for use by future scholars.

### **Moving Beyond Informal Knowledge Sharing to Formal Publication**

Many researchers use informal channels (especially email) to share data with colleagues. The circulation of data within personal networks suggests that data is used to express and reinforce personal ties between researchers. This method of sharing also reflects concerns over

quality and trust, as researchers exchange data with trusted colleagues believed to be creators of reliable data. However, data shared in this way tend to see little documentation or investment in “clean-up” because they are not intended to be shared beyond a single colleague. Thus, to understand someone else’s dataset enough to use it, one would probably have to contact its original creator.

The Dissemination Information Packages for Information Reuse (DIPIR) project,<sup>2</sup> sponsored by the Institute of Museum and Library Services, has engaged several university and foundation digital repositories in the task of investigating the reuse of digital data in three disciplines—quantitative social sciences, archaeology, and zoology. DIPIR recently interviewed archaeologists about their data sharing needs and behaviors. The interviews indicate that scholars perceive data as very important and are concerned about data preservation. They also reveal that archaeologists tend to manage data somewhat casually, as illustrated by one response:

“I use an Excel spreadsheet . . . which I . . . inherited from my research advisers. . . . my dissertation advisor was still recording data for each specimen on paper when I was in graduate school so that’s what I started . . . then quickly, I was like, ‘This is ridiculous’ . . . I just started using an Excel spreadsheet that has sort of slowly gotten bigger and bigger over time with more variables or columns . . . I’ve added . . . color-coding . . . I also use . . . a very sort of primitive numerical coding system, again, that I inherited from my research advisers. . . . So, this little book that goes with me of codes which is sort of odd, but . . . we all know that a 14 is a sheep.” (CCU 13)

This kind of informal archaeological data management represents a challenge to dissemination and preservation efforts. Researchers often are reluctant to expose their raw data to outside scrutiny. Hence, it is more likely to be circulated only with trusted colleagues. Not only is sharing in this way insufficient to ensure the preservation of data, it is also not enabling new forms of large-scale data analysis that will carry the discipline forward. Few datasets see more formal and public modes of dissemination. In a casual setting, the person who emailed the data

may privately explain their methodology to the person “borrowing” it. Without such explanation, outsiders would have no way of knowing that “14 is a sheep.”

Such one-to-one transfers inherent in informal “back-channel” data sharing are also time-consuming and inefficient, since the next person asking for the same dataset would need to be given the same explanation all over again. Publicly releasing well-documented data to the entire research community enables an individual researcher to communicate more efficiently. Formally publishing data also promotes the preservation of data, since datasets exposed and documented through open access publication can also be made available to professionally-maintained digital repositories.

The Case for Open Data Publishing

Open data publishing promises to improve the efficiency and quality of data-sharing in much the same way that conventional publication improves the dissemination of research findings. Data sharing advocates agree that data sharing requires more than “dumps” of raw and undocumented data on the Web. Datasets typically need to have adequate documentation and consistency to be widely usable. Furthermore, researchers need to see clear signals of the quality of the data, to both find information likely to be reliable and communicate to their colleagues that shared data represents substantive scholarly contributions.

HOW ARCHAEOLOGISTS SHARE INFORMATION

To illustrate our model of data publishing, table 1 (below) describes approaches to data management and dissemination, specifically private and public modes of “data sharing” and “data archiving” (see also Atici et al. 2012). Table 1 should not be considered an exhaustive list, since other models of data dissemination are surely possible. Data publishing, sharing, and archiving are not mutually exclusive, and none of these models should be considered superior in all cases. Rather, the different models can play complementary roles in archaeological communications. By comparing these models and their relative strengths and weaknesses we can get a better sense of how archaeological communications need to evolve.

TABLE 1 MODES OF ARCHAEOLOGICAL DIGITAL DATA DISSEMINATION

Type	Channels	Advantages	Disadvantages
Private Sharing	Email, private file transfers between individuals	Informal, easy, low barrier to entry. Facilitates collaboration among colleagues, especially for works in progress. Personal reputation communicates quality and trust.	Explanation/documentation of data only communicated informally. Incentives to share only with friendly colleagues. No structures for formal attribution. Back-channel communications can be abused, resulting in data “theft” if a colleague takes undue advantage of a dataset. No institutional archiving. Data not available for outside use. No independent verification of claims based on data.
Public Sharing	Website	Informal, easy, low barrier to entry. Inexpensive global (if short-term) distribution. Researchers and institutions can showcase their data inexpensively on their own sites with their own branding. Quality and reliability usually signaled by institutional branding and researcher reputations.	Data may not see adequate documentation or measures for interoperability. Data may not have clear attribution or reliable/persistent citation. No clear process for ensuring and promoting data quality. May promote “fragmentation,” with many datasets scattered across many websites with little interoperability. Data may see no formal archiving and preservation.

(Continued)

TABLE 1 MODES OF ARCHAEOLOGICAL DIGITAL DATA DISSEMINATION (Continued)

Type	Channels	Advantages	Disadvantages
Archiving	Data repository (accessible via the Web) Examples: Archaeology Data Service, tDAR	Professionally managed data curation and preservation. Focused and centralized investment in computing infrastructure. Promotes better dissemination services, sometimes with powerful search and analysis tools (tDAR), than feasible for most websites. Focused and centralized investment may enable rapid development of significant collections of data. Formal requirements for metadata and data documentation.	No clear process for promoting data quality. Emphasizes preservation ethics and potential benefits to future researchers, but short-term benefits to contributing researchers may be less clear. Preservation emphasis focuses on managing “files” with content and structures that may be poorly suited to communication, use, and interoperability. Centralization of computing infrastructure may create “lock-in” effects and crowd out wider participation and innovation in informatics research.
Integrated Data Publishing	Website (plus data repository) Examples: Open Context, Pleiades	Aligns to professional expectations and rewards for publication and citation. Professional review of data. Outsiders review datasets and documentation to promote more data quality and adequate documentation. Data publisher works with digital libraries and repositories for data curation and preservation. These services enable reliable and persistent citation of data. Datasets may be aligned to a global schema or ontology to promote more semantic interoperability. Editorial review may promote standards alignment and formalized reference of relevant linked data, improving interoperability and semantic understanding. Moderate-scale investment, supporting some technical services beyond the capabilities of most single websites. May achieve sufficient scale of data to focus on themes (regional, temporal).	Data editing and review processes can be costly and time consuming. Usually not fully generalized for all archaeological data needs (i.e., one publishing system cannot accommodate all types of archaeological data). The need for multiple systems makes this a more expensive approach. Only a few members of the research community may have technical knowledge needed to fully use interoperability measures.

Note: Table 1 outlines different models of data dissemination, roughly from less to more informal approaches. The main exception to this generalization lies in digital archives; the leading digital data repositories are highly professional organizations with very elaborate and formalized archival processes. These processes, however, are typically focused on issues of metadata description and data preservation. Digital archives generally do not have formal processes for promoting data quality.

As already noted, datasets need to have adequate documentation and consistency to be widely usable. Data archiving and data publishing models focus attention on data documentation. Rich metadata description can be used to support powerful search and discovery tools and compelling user interfaces that help visualize and map collections of data. However, supplying such documentation requires time and effort. Without clear rewards for researchers to provide rich metadata, archives may face pressures to accept more limited and incomplete metadata descriptions, even if this means less potential for discovery.

Most research publishing models try to align the immediate incentives of researchers with the “public

good” that comes from the dissemination of high-quality data. Researchers invest effort in publication because they align the vast majority of *their own* research goals toward publication ends. The effort and back-and-forth between authors, editors, and reviewers all aim to hone argumentation and promote quality. Publishing in venues where editorial and review processes are highly formalized enhances the prestige of scholarly communications, and helps motivate researchers to submit papers to highly selective journals. Prestige and professional recognition motivate researchers to participate in labor-intensive communication systems.

These motivations can be extended to promote the dissemination of digital data. Datasets have rich potential for

wide community impact. They can be either reused and reanalyzed individually or analyzed in aggregate. As the pool of publicly available data grows, so does the potential for datasets to be combined and recombined with other data. Researchers participating in data publishing therefore can see continued use and impact of their data contributions, and in turn, earn rewards coming from enhanced prestige and recognition. In other words, data publication models can align professional and career interests with the research interests of the larger community (see Costello 2009; Griffiths 2009; Piwowar, Day, and Fridsma 2007).

## Data Publishing Workflows

Scholars are familiar with editorial workflows that transform manuscripts into completed publications. Researchers submit text files to journal editors, who then circulate manuscripts for review. When a paper is accepted, a researcher works with a journal editor through multiple revisions (many suggested by peer-review evaluations) before the manuscript is ready for publication. Email, versioning, and edit-tracking help coordinate the work. The final product is a work of collaborative “co-production” between authors, editors, reviewers, and type-setters.

Similarly, appropriate workflows and technology can facilitate data publishing. Datasets, however, have several important qualities that differ from manuscripts. Datasets can be quite large and full of complex interrelationships between various tables and multimedia files (images, videos, GIS, etc.). This is especially the case in archaeology where projects often involve large teams, including specialists who create their own datasets. Archaeological documentation is also highly multimedia and can generate tens of thousands of images and other media files (3D scans, GIS, remote sensing, etc.) that need association with other documentation. Our experience shows that it is common to see complex dependencies between various parts of an archaeological project. For example, diverse specialist datasets (zooarchaeology, ceramics, lithics, ground stone) at an excavation typically need to be related through reference to archaeological contexts.

In contrast to texts, human beings typically do not read data. Rather, they use data mediated through software that summarizes and visualizes datasets. Humans interpret texts via pattern recognition, heavily aided by background knowledge and expectations. The transactional nature of data introduces a different set of issues impacting the quality and usability of data. Whereas small errors in a text often go unnoticed, such errors can have dramatic impacts on the use and interpretation of a dataset. For instance, a misplaced decimal point in a numeric field can cause problems for even basic statistical calculations. Such errors can also break visualizations. Integrating, cleaning, and adequately documenting such large and complex datasets requires a great deal of effort and experience with data.

## Data Publishing With Open Context

The authors are involved in developing a system of formalized open data publication that addresses these concerns. Open Context (<http://opencontext.org>), in development since late 2006, publishes data and related media (images, maps, and narrative documentation) primarily in archaeology and related fields. Open Context has a team of editors and an editorial board comprising experts in various archaeological domains and specializations. Editorial boards can perform important signaling roles in academia by elevating the prestige of data sharing. Editorial oversight, coupled with clear and trustworthy citation practices, can make data dissemination a recognized and professionally valued form of publication. Table 2 details the publication process for Open Context.

To date, Open Context has published some 297,000 records and 65,000 media files from 24 different projects, some of which are very large and complex. This makes Open Context roughly the scale of a large museum collection and, as is increasingly the case with leading museums, Open Context data can flow into other systems via sophisticated Web services and APIs for analysis and visualizations. However, building this type of resource can be a slow process—it takes time and effort to create the rich output of a data publication. The time and effort required of data publication, though not onerous (see below), still



TABLE 2 DATA PUBLISHING PROCESS WITH OPEN CONTEXT

Step	Activities	Outcomes
Initial Proposal	Open Context editorial staff invites submission of dataset or researcher contacts Open Context with proposed data publication.	Editorial staff discusses with researcher the suitability of the project for publication with Open Context (if not suitable, editorial staff recommends alternative venues/archives).
Data and Metadata Submission	Open Context editors provide researcher with a list of required project information (metadata) needing completion. Researcher submits data to Open Context editors via email, ftp, DropBox.	Datasets are ready for editing and clean-up (using the Data Refine <sup>3</sup> tool). Open Context editors create initial draft of project.
Data Review	Open Context editorial staff reviews data for consistency and plausibility. Editors fix minor issues. More significant problems identified by editors are sent back to contributing researcher for resolution. Process continues until editors and contributor agree that the dataset is sufficiently cleaned.	Dataset is ready for metadata documentation and standards alignment.
Metadata Documentation and Standards Alignment	Open Context editorial staff add required project information/metadata provided by contributor. Editorial staff works with contributor to document data fields and values. If applicable, editorial staff references shared vocabularies and ontologies from elsewhere on the Web (to enable Linked Open Data applications).	Dataset is mapped into Open Context's data structure and is available for preview by contributor and editors. Dataset is ready for Editorial Board review and, if desired, external peer review. Dataset carries a star rating, based on how far along it stands in the review process. A project earns more stars as it sees additional editorial scrutiny.
Editorial Board Review and Optional Peer Review	Open Context Editorial Board members review documentation and data for methodological soundness, adequacy of documentation, and potential for reuse. Additional external peer review (using the above criteria) can be requested by the contributor.	Datasets approved for publication are released publicly on Open Context.
Post-Publication	Datasets are accessioned into the California Digital Library (CDL) for archiving. Datasets are added to GitHub (for public version control). Open Context collects metrics on use of datasets.	Contributor's datasets are preserved through permanent archiving with a university library (CDL). Contributor's datasets are citable (with persistent identifiers from the CDL). Subsequent changes/updates are easy to track (GitHub version control). Contributor gains information on use of published datasets (valuable for CVs and promotion).

needs to be justified in terms of downstream research efficiencies and impacts as well as new understanding about the past.

In the case of Open Context, data can be cited and retrieved at the highly granular level of “one URL per potsherd.” The ability to reference and retrieve highly granular data facilitates a great deal of flexibility and room for innovation in how archaeological data get integrated into other forms of scholarly communication. In practice, this means that an individual find or group of finds, identified by a stable Web address, can be linked and referenced in everything from a book or article, to a blog post, or even

a tweet. This makes records of finds, contexts, excavation logs, and other archaeological observations more directly integrated into research discourse. Visible Past, a publishing platform for the spatially-oriented humanities, recently illustrated such linking with a series of articles<sup>4</sup> referencing Nicholas Rauh's (2012) *Rough Cilicia* data published in Open Context.

Linked Open Data is still in its infancy in terms of research applications for archaeology. Nevertheless, it is not difficult to imagine how archaeologists can benefit from it. For instance, many researchers make typological parallels with finds they observe in their own excavations

and finds described by other researchers. Linked Open Data methods would allow researchers to explicitly identify such typological parallels, enabling software to use this data to help visualize networks of stylistic similarities in a study region. Linked Open Data applications go beyond speculation, and see increasing application and implementations (Isaksen et al. 2009; 2012).

Archaeologists need to see more direct research applications in order to better justify the added cost and effort required to publish Linked Open Data. In 2012, we received funding<sup>5</sup> to develop three Linked Open Data demonstration projects with Open Context. One project focuses on a collaborative and comparative analysis of several zooarchaeological datasets documenting early agricultural communities in Anatolia. The datasets will be published with Open Context and made comparable by linking and annotating them according to taxonomic concepts published by the Encyclopedia of Life (<http://eol.org>) and to phenotypic ontologies used to make morphological data interoperable. Another project, funded by the NEH, will involve the linking and annotation of data documenting archaeometric studies of ceramic and metal objects from the Late Bronze Age through Classical periods in the eastern Mediterranean. The project will link and annotate these data using the Pleiades Gazetteer (<http://pleiades.stoa.org>) and the Concordia Vocabulary, a simple ontology well-suited for studying trade and exchange relations. The NSF recently funded a third Linked Open Data publication effort, this time focused on the integration and dissemination of site file records (stripped of sensitive data, particularly precise geographic coordinates) maintained by State Historical Preservation Offices (SHPOs) in 11 states. This project, led by David G. Anderson and Joshua Wells, will enable search, discovery, and analysis of site file data now fragmented across state lines and locked in inaccessible databases.

Finally, Open Context is also following the example of Nomisma.org and the American Numismatic Society, which publish key reference collections of ancient coins to facilitate research using Linked Open Data. Scholars working in specific research areas routinely use reference collections to guide the identification and analysis of objects collected in field work. Similarly, Open Context will publish rich typologies and associated archaeometric data relating to East Asian ceramics that circulated

in Pacific trade routes over the past several hundred years. Linked Open Data can play an invaluable role in making reference collections a powerful tool in data integration.

## Qui Solvit? Sustainability and Open Data—Who Pays?

In addition to linking, openness is essential to making data work well for the research community. Briefly, open data is defined by three primary characteristics:

- *Technical Openness*: Data must be available in widely used, nonproprietary file formats that can work across multiple computing and software platforms.
- *Legal Openness*: Data must be free of encumbering intellectual property restrictions (copyright or contractual obligations).
- *Access*: Datasets must be made available freely and, unless there are overriding privacy or security needs, data releases need to be both comprehensive and sufficiently documented to enable reuse.

Archaeological data archives and data publishers typically try to ensure that the data they manage meet these criteria for openness. Many archaeologists have bitter personal experience trying to recover information from files in obsolete, proprietary data files. From a data preservation perspective, “technical openness” makes archiving easier, since data encoded in open file formats can be more easily migrated to more current file types. Similarly, archaeological data archives and publishers also often try to promote legal openness.

Data that cannot be reused and recombined with other data because of copyright or contractual encumbrances have less value to the research community. To achieve legal open interoperability and avoid conflicting licensing and contractual conditions, datasets need standard licenses. For most scientific datasets, best practice usually means using the Creative Commons Attribution License or the Creative Commons Zero (public domain) dedication. Of course, these are generalizations that are not applicable in all cases, especially where the ethical landscape of managing data requires consideration of different needs such as the needs of indigenous peoples whose values may require some data to come under



different legal and access regimes (see Kansa 2012; Kansa 2009; Christen 2009; Chander and Sunder 2004).

In general, data reuse is the fundamental point of both publishing and archiving efforts. Datasets need to be legally open and free from most intellectual property restrictions. All of this begs the question about financially sustaining Open Data, especially for a discipline under severe and protracted financial strain. Critics of Open Access publishing (of data or more conventional articles and books) claim that publication costs, especially costs of maintaining quality (through editorial and peer review), require subscriptions and other fee-based access charges. A belief in the necessity of access charges to maintain quality underlies the American Institute of Archaeology's recent and highly controversial position against Open Access (Bartman 2012).

In our view, the critics of Open Access miss the point of how scholarly communications fit into the larger picture of public support of research. Archaeology as a discipline is manifestly not financially sustainable. Archaeological research activities make no profit. Unlike the destructive antiquities trade, archaeologists theoretically work in the public interest, creating new knowledge about the past. The pursuit of the public good justifies continued public investment in archaeological research. If one considers the communication and preservation of research data and findings as an integral aspect of research, then scholarly data and other outputs should be aligned toward the public good.

Dysfunctions emerge when institutional and personal incentive structures collide with the public-good mission of archaeology. On an individual level, overly narrow definitions of what constitutes a recognized scholarly contribution can lead to inadequate treatment of data. If not valued by tenure committees and university "bean counters," datasets will continue to languish on individual hard drives and remain vulnerable to loss. Similarly, conventional publishing practices in archaeology largely lead to dysfunctional outcomes. Creating archaeological knowledge is very expensive. It requires special training, equipment, and access to often remote and hazardous locations, insurance, storage, conservation, lab analyses, and many other costly inputs. All of these inputs are largely financed through public sources. When disseminating the results of this costly research, archaeologists

usually author, edit, and review each other's manuscripts without any financial compensation. It is only at the last stage of all of this publicly supported effort that commercial and semi-commercial publication starts to "add value" by organizing (uncompensated) review processes, copy-editing, layout, and design. We do not wish to diminish the value publishers add to the communication of research. Layout, marketing, and design help promote effective communications. However, the costs and effort required at this last stage are a small fraction of the costs already invested, by the public, in the research process (see also Suber 2012).

Public investment already subsidizes archaeological research, and can subsidize the dissemination and preservation of research outputs. In other words, open access (and open data) can have the same sustainability strategy as the rest of archaeological research. This shift would better align archaeological publication (which currently typically results in closed access, copyright restricted, private intellectual property) back to the overall mission of archaeology to create public goods in the form of accessible and usable knowledge about the past. In our experience with Open Context, the costs of open access data publishing and archiving, including editorial review and data cleanup, stand as a small fraction of the overall costs of conducting archaeological field work. For instance, we recently published data from Kenan Tepe (Parker and Cobb 2012), a large and complex multi-year excavation of a Neolithic through Iron Age site in eastern Turkey. We estimate that our total publication costs (mainly labor) amounted to roughly \$10,000 to \$15,000. These costs are a small fraction of the roughly \$800,000 of direct costs needed to finance the actual excavations. The \$15,000 spent publishing with Open Context (and archiving with the California Digital Library) will help insure that the public's \$800,000 investment can be used and reused by the broadest community, now and into the future.<sup>6</sup>

## Conclusion and Future Vision

Archaeology is at a crossroads on the question of how to finance the dissemination of high-quality knowledge. The rise of open access and open data models challenges

the status quo in our understanding of how and why we conduct and communicate our research.

We acknowledge that some scholars are likely to push back against open data publishing—many simply because it requires time and the rewards are still uncertain. However, it is instructive to remember that though we may get frustrated by editorial oversight in conventional publication, we all can benefit from the results of that collaborative exercise. Though the current systems of peer review and publication are under pressure to evolve, the basic need for editorial oversight and review processes remains. Our advocacy of open data publication highlights how the collaboration between editors, reviewers, and contributing researchers can be expanded and extended to include primary data.

Open Context's "data sharing as publication" approach can better meet critical professional incentive needs by providing a citable, professionally edited, publication venue backed by a leading digital repository. Our goal is not to develop Open Context into a centralized "one repository to rule them all" system. Rather, it is to enable Open Context to participate in a distributed ecosystem. Just as multiple print journals exist, so can multiple data publishers. Publishing high-quality data aligned to standards requires effort and expertise. To distribute this effort, this model can and should be replicated and adapted by other research teams.<sup>7</sup> Archaeology's growing data challenges can only be surmounted through constant innovation and collaboration across the widest possible community.

## Notes

1. <http://creativecommons.org>
2. <http://dipir.org/>
3. Data Refine integrates Google Refine (<http://code.google.com/p/google-refine/>) and the Mantis Bug Tracker (<http://www.mantisbt.org/>) to facilitate collaborative editing and data documentation.
4. See for example, <http://visiblepast.net/see/cilicia/rough-cilicia-funerary-features-report-demo/>
5. Funding for various aspects of this project comes from the NEH (HK-50037), the American Council of Learned Societies, the NSF (BCS-1217240), and the Encyclopedia of Life.
6. For further discussion of the challenges and opportunities of open access and open data, see Kansa's (2012) contribution to a special issue of *World Archaeology* dedicated to this topic.

Kansa's paper and others explore questions of sustainability and how the open access and open data movements offer trenchant critiques of the current status quo.

7. Fortunately, Open Context is not the only effort exploring data publication models. The Journal of Open Archaeological Data, launched in 2011, has similar aims, but is a for-profit commercial effort. In spite of being commercial, it is fully open access and has adopted the most permissive of Creative Commons' license options (Attribution). The entry of a commercial journal in this niche is a welcome development, highlighting new possibilities for profitable (and hopefully sustainable) business practices that align with the public interest in open and reusable data.

## References

- Atici, L., S. W. Kansa, J. Lev-Tov, and E. C. Kansa. 2012. Other People's Data: A Demonstration of the Imperative of Publishing Primary Data. *Journal of Archaeological Method and Theory* 1(3): 1–19.
- Bartman, E. 2012. From the President: Open Access. *Archaeology* 65(3): 6.
- Chander, A., and M. Sunder. 2004. The Romance of the Public Domain. *California Law Review* 92(13). [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=562301](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=562301) (accessed April 6, 2012).
- Christen, K. 2009. Access and Accountability: The Ecology of Information Sharing in the Digital Age. *Anthropology News* 50(4): 4–5.
- Costello, M. J. 2009. Motivating Online Publication of Data. *BioScience* 59: 418–27.
- Griffiths, A. 2009. The Publication of Research Data: Researcher Attitudes and Behaviour. *International Journal of Digital Curation* 4(1): 46–56.
- Harley, D., S. K. Acord, S. Earl-Novell, S. Lawrence, and C. J. King. 2010. *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Berkeley, CA: Center for Studies in Higher Education.
- Isaksen, L., E. Barker, E. C. Kansa, and K. Byrne. 2012. GAP: A NeoGeo Approach to Classical Resources. *Leonardo* 45: 82–83.
- Isaksen, L., K. Martinez, N. Gibbins, G. Earl, and S. Keay. 2009. Linking Archaeological Data. In *CAA 2009: 37th Annual International Conference on Computer Applications and Quantitative Methods in Archaeology, Williamsburg, US*, 22–26 Mar 2009. <http://eprints.ecs.soton.ac.uk/18240/> (accessed January 31, 2010).
- Kansa, E. C. 2009. Indigenous Heritage and the Digital Commons. In *Traditional Knowledge, Traditional Cultural Expressions, and Intellectual Property Law in the Asia-Pacific Region*, ed. C. Antons, 219–44. Max Planck Series on Asian Intellectual Property Law 14. Alphen Aan Den Rijn: Kluwer Law International.
- Kansa, E. C. 2012. Openness and Archaeology's Information Ecosystem. *World Archaeology* 44: 498–520.

- Onsrud, H., and J. Campbell. 2007. Big Opportunities in Access to "Small Science" Data. *CODATA Data Science Journal* 6: 58–66.
- Parker, B., and P. Cobb. 2012. Kenan Tepe: (Overview). *Open Context*. Accessed November 26, 2012, doi: 10.6078/M7H41PBJ#.
- Piowar, H. A., R. S. Day, and D. B. Fridsma. 2007. Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. Accessed November 26, 2012, doi:10.1371/journal.pone.0000308.
- Rauh, N. 2012. Rough Cilicia. *Open Context*. <http://opencontext.org/projects/295B5BF4-0F44-4698-80CD-7A39CB6F133D> (accessed on November 26, 2012).
- Suber, P. 2012. *Open Access*. MIT Press Essential Knowledge Series. Cambridge, MA: MIT Press.

## RESPONSE

### Thoughts about Open Access Publishing in a Humanities Context

PATRICK H. ALEXANDER

Director, The Pennsylvania State University Press

820 North University Drive, USB1, Suite C, University Park, PA 16802, [pha3@psu.edu](mailto:pha3@psu.edu)

Eric Kansa and Sarah Whitcher Kansa hit the nail on the head: "The discipline [archaeology and heritage studies] urgently needs better data professionalism if researchers are to make more credible and replicable knowledge claims. . . . but scholars generally lack the time and technological know-how to disseminate [and archive] data in a meaningful and lasting way" (p. 88). Their urgent message holds true across the arts and humanities, and while differences exist when it comes to sharing v. protecting and long argument v. database, common issues remain. Moreover, their proposal that data sharing should be viewed as publishing addresses growing concerns that when data is not shared scholarship suffers and knowledge is lost. Picking up on Diane Harley's insights, Kansa and Whitcher Kansa reiterate that for data sharing to be viewed as publishing "researchers must see clear evidence that data sharing is worth their time and effort" (p. 88; see also pp. 90 and 92).

Key to Kansa and Whitcher Kansa's proposal's success is open data that are technically, legally, and financially accessible (free) (p. 94). They invited me to engage their article on this point. Philosophically, I wholeheartedly endorse their diagnosis of the problems.<sup>1</sup> Practically, however, the questions of sustainability and "Who Pays?" may not be answered as easily as we would all like. Applying a uniform solution across the humanities poses immense challenges. They write, "In our view, the critics of open access miss the point of how scholarly communications fits into the larger picture of public support of research. Archaeology as a discipline is manifestly not financially sustainable" (p. 95). Here's where we might disagree. No matter where funding originates and no matter to what extent data are open and free, research and publication, scholarly communications, must be sustainable, just as archaeological excavations must be sustainable. "Sustainable," by the way, means more than eking by; it also involves planning for future resources, adapting to new technologies (data curation, archiving, including new forms of scholarship), and maintaining an infrastructure. Kansa and Whitcher Kansa recognize this when they point to making the data available beyond the excavation. But in reality, "sustainable" means "profitable," because planning, adapting, and maintaining an infrastructure require funding beyond an initial project. Kansa and Whitcher Kansa's example from the excavations at Kenan Tepe of how public investment (here I understand them to mean taxpayer-funded research) "can subsidize the dissemination and preservation of research outputs" (p. 95) is extremely helpful. They calculate that "The \$15,000 spent publishing with Open Context (and archiving with the California Digital Library) will help insure that the public's \$800,000 investment can be used and reused by the broadest community, now and into the future" (p. 95). From a financial/sustainability perspective, it would be helpful to know what other results the \$800,000 publicly funded investment yielded besides the Open Context publication and archiving with CDL. It would likewise be vital to know whether the publication expenses cited included overhead costs at Open Context and CDL (salaries, etc.), and at Kenan Tepe. With funds available, to publish in an Open Access format the publicly funded results of the excavations at Kenan Tepe makes perfect sense. The "larger picture of public support"