Digital Data and Data Literacy in Archaeology Now and in the New Decade

Eric Kansa and Sarah Whitcher Kansa

OVERVIEW

Digital data play an increasingly important role in how we understand the present and the past. The challenges inherent in understanding and using digital data are as intellectually demanding as any other archaeological research endeavor. For these reasons, data management cannot be regarded as a simple compliance or technical issue. For data to be meaningfully preserved and used in intellectually rigorous ways, they need to be integrated fully into all aspects of archaeological practice, including ethics, teaching, and publishing. In this review, we highlight some of the significant and multifaceted challenges involved in managing data, including documentation, training, methodology, data modeling, trust, and ethical concerns. We then focus on the importance of building data literacy broadly among archaeologists so that we can manage and communicate the data our discipline creates. This involves more than learning to use a new tool or finding a data manager for one's excavation or survey. Long-term, responsible stewardship of data requires understanding the workflows and human roles in data management. Putting effort now into thoughtful data management and broad data-literacy training means we will be able to make the most of the "bigger" data that archaeologists now produce. An important aspect of this reorientation will be to look beyond the boundaries of our own research projects and information systems. Future research, teaching, and public engagement needs will also compel us to explore how our data articulates with wider contexts—within and beyond our discipline.

Keywords: data management, data literacy, collaborative research practices, digital publishing, professional preparation

RESEARCH DATA AND ITS MANAGEMENT

The term "data" means different things to different people. For the purpose of this discussion, we consider data to mean the primary documentation of our research—including our field notes, images, and 3D models, as well as the entities modeled, described, and organized in our GIS files, spreadsheets, and relational databases. This discussion centers on the connections (or lack thereof) of various types of data created and managed by different people. Our hope is that greater attention to research data management and increased data literacy will make these different types and sources of data useful (and usable) as a more cohesive whole, thereby opening up new research opportunities.

Growing interest in archaeological data reflects a broader realignment in research priorities and expectations. This realignment involves a shift in focus from "final" products—such as books, journal papers, and reports—to a growing interest and concern over promoting quality, professionalism, and ethics in the research process that leads to those final products. For example, the recent granting requirement for data management plans (DMPs) helps to signal that demonstration of quality and professionalism in the process of research is becoming a more expected outcome of scholarship. Consequently, data management, ethics, and safety issues in field work (Colaninno et al. 2020; Colwell 2016; Heath-Stout and Hannigan 2020; Leighton 2020), community archaeology (Gonzalez 2016; Gupta et al. 2020), and even greater transparency in data analysis with reproducible research (Marwick 2017) all reflect a growing recognition that *how* one arrives at research outcomes is at least as important as the outcomes themselves.

The practice of developing and implementing a DMP is still relatively new in archaeology. DMPs tend to be project specific, and they vary in how they address topics such as data security, backups, data quality and validation, and provisions for long-term data preservation and access. We should emphasize that virtually every aspect of archaeological practice involves some element of data management, and this can no longer be considered as a separate, secondary concern in carrying out archaeological research projects. In other words, it is increasingly inadequate to see research data management as simply a "checkbox" requirement for grant applications.

If we consider data management as integral to all aspects of archaeological research, it should receive greater scrutiny and attention beyond the grant application DMPs, and this adjustment of priorities raises important considerations around data access and reuse:

- Do our data structures and schemas (the way we organize data) adequately represent the phenomena we want to observe and record?
- How do the data we create through excavation, survey, or archive and collections studies articulate with the data created by our colleagues, who are working on different systems and platforms both within and outside of our discipline (e.g., Beebe 2017; Faniel

Advances in Archaeological Practice 9(1), 2021, pp. 81–85

DOI:10.1017/aap.2020.55

et al. 2020)? What are the research opportunities and methodological challenges involved in cross-referencing and integrating data at small, medium, or large scales (Figures 1 and 2)?
- Have we included sufficient documentation (metadata and paradata) to enable others to understand and reuse our data for replication or to ask new questions?
- How do we get teams of collaborating researchers to identify and name items in ways that are consistent and unambiguous? For example, does "Trench IV" mean the same thing as "Trench 4"? And if so, how is that made explicit in databases—especially in different databases curated by different members of a project team (Figure 3)?
- What research questions can be explored by citing and using data made accessible through generalized systems, such as tDAR or Open Context, and through special-purpose platforms such as (among others) the Canadian Archaeological Radiocarbon Database (CARD), the Pleiades gazetteer of ancient places, the Digital Archaeological Archive of Comparative Slavery, the Chaco Research Archive, Kerameikos, and Nomisma, or various government- and CRM-maintained data sources?
- How do we navigate the landscape of opportunities for public engagement while reducing risks of race- or gender-targeted harassment and violence when presenting archaeological arguments and data in social media?
- What is the role of digital data in addressing the ongoing "curation crisis" (Kersel 2015) in archaeology?

Research conducted by the Open Context team and colleagues as part of the National Endowment for the Humanities–funded Secret Life of Data (SLO-data) project highlights additional frictions in archaeological practice around the integration of excavation data and the data produced by later specialist and lab studies (Faniel et al. 2020). These are problems that cannot necessarily be solved with a tool or some other technological fix. This is because they are related to human workflows and communication. Most archaeologists use various digital tools to collect information during excavation, survey, and analysis. Despite the ubiquity of digital data collection, professional engagement with data may not be widely distributed across a team, and many research teams relegate data management responsibilities to a single specialist. This lack of integration of data literacy skills into archaeological training undermines data quality and inhibits later reuse of archaeological data. Data specialists are often temporary staff members and—for various reasons that include the precarious nature of short-term funding—they may need to leave a project, taking their unique skills and expertise with them. If data management skills are not more widely distributed, projects may find themselves unable to maintain critically needed workflows and information systems.

THE ROLE OF DATA LITERACY AND INCLUSIVITY IN THE FUTURE OF ARCHAEOLOGICAL DATA

In addition to the human workflow and communication challenges around digital data, many archaeological projects are focused only on their own project's database and data management and preservation strategies—that is, there is not yet much discussion of bridging data across projects or larger-scale data integration, which

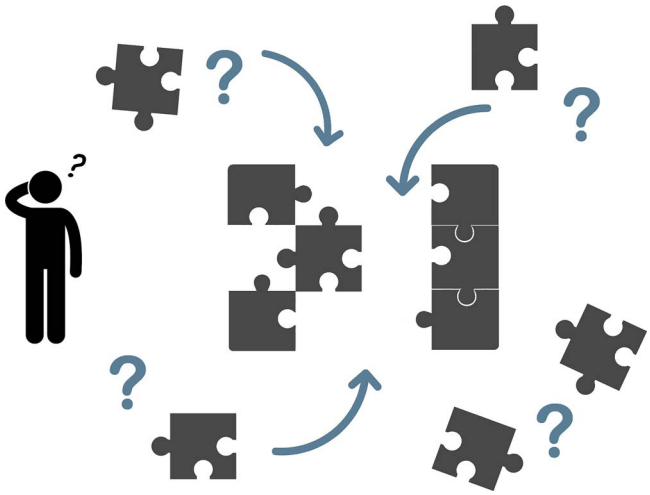

FIGURE 1. There are many data-related challenges in contributing to "bigger picture" questions. Different projects and individuals record data under different recording and sampling protocols. What methodologies and theoretical frameworks do we need to bring disparate datasets together in meaningful ways? How do we promote consensus in areas where comparative analysis is fruitful while still encouraging innovative approaches to data description and modeling?

is what might start to get us closer to "bigger data" in archaeology. Increasing expectations around data access and reuse raises new professional development and training needs. Largely due to mandates from funders, a great deal of archaeological research data is now being archived, but the discipline still lacks good examples of its reuse. If one aim of preserving data is for others to be able to use it, we need to increase practitioners' skills in accessing and using data. Broadening data literacy skills over the next decade will help us realize the full potential of archaeological data.

"Data literacy" has many definitions. A 2015 white paper by Rahul Bhargava and colleagues for the Beyond Data Literacy Workshop describes data literacy as "not primarily about enabling individuals to master a particular skill or to become proficient in a certain technology platform. Rather it is about equipping individuals to understand the underlying principles and challenges of data" (Data-Pop Alliance 2015:8). This description is useful because it does not promote a specific technique, service, or technology, but instead promotes deeper appreciation of foundational issues of empowerment, critical thinking, and argumentation. This white paper continues its description of data literacy with the following:

> This understanding will in turn empower people to comprehend, interpret, and use the data they encounter—and even to produce and analyze their own data. This can only be achieved by considering data literacy . . . a means toward a necessary reinvention of community engagement and empowerment—towards what we term data inclusion [Data-Pop Alliance 2015:8].

It also emphasizes "the ability to read, work with, analyze, and argue with data" (Data-Pop Alliance 2015:14), and this emphasis

https://doi.org/10.1017/aap.2020.55 Published online by Cambridge University Press

FIGURE 2. Even on the same excavation project, teams working in different trenches may use different terminology, or they may record at different levels of specificity and detail. The SLO-data project found that differences in documentation occurred especially where teams worked in trenches located far apart (Faniel et al. 2020). Having an individual whose role is to spend time in each trench can help with consistency in data collection and ease of data integration across various parts of a project. What other practices can research teams adopt to make data collection more consistent and cohesive? Photo credit: "D4 Figure 2 from Turkey/Kenan Tepe/Area D/Trench 4" by Bradley Parker and Peter Cobb, Kenan Tepe 2012, in Open Context.

on argumentation and critique helps situate data as a central concern for scholarship rather than technical proficiency with a specific software application. Archaeology offers rich opportunities to promote broader data inclusivity. Archaeologists must grapple with the interpretive challenges of incomplete and often ambiguous data (Faniel et al. 2013; Kansa 2012). Explicit documentation and justification of potentially contestable analytic steps can strengthen the rigor of archaeological knowledge claims. Emerging reproducible research practices emphasize use of open data together with greater transparency and public documentation of analytic and interpretive steps in research workflows with data (Marwick et al. 2017). If presented through accessible forms of pedagogy, reproducible research in archaeology can empower students to peer into the black box and evaluate the steps used in making claims with data.

Changing professional publication practices and norms to better cross-reference rich media and digital data with textual narratives and arguments can similarly promote broader data literacy. Recent hybrid publications in archaeology that incorporate digital media, such as 3D models and explicitly modeled Linked Open Data drawn from across the web, highlight future possibilities for communicating archaeology in engaging and analytically rigorous ways (e.g., Counts et al. 2020; Opitz et al. 2016; also discussed in Opitz [2018]). Leveraging these examples and drawing on expanding data literacy, new forms of integrating, visualizing, and communicating archaeological data will increasingly enhance the value and analytic power of our archaeological publications and reports. As data—especially well-contextualized and linked data—become more integral and expected in our professional communications, our community can make the many incremental contributions required to address the types of "grand challenge" questions raised by Kintigh and colleagues (2014).

DATA-DRIVEN RESEARCH IN THE FUTURE OF ARCHAEOLOGY

Archaeologists, like practitioners in other disciplines, must navigate both the digital and the physical worlds. After all, much of our data models and describes landscapes, sites, deposits, objects, and other aspects of physical reality. We need better approaches to more clearly and explicitly associate our data with these physical things. A new initiative, the Internet of Samples (iSamples) project, aims to better manage linkages between physical scientific samples—including archaeological objects and ecofacts—and data about these objects. The project will develop services to provide globally unique identifiers for material samples, as well as critically needed metadata shared across multiple disciplines, including archaeology (https://isamples.org/). This is

https://doi.org/10.1017/aap.2020.55 Published online by Cambridge University Press

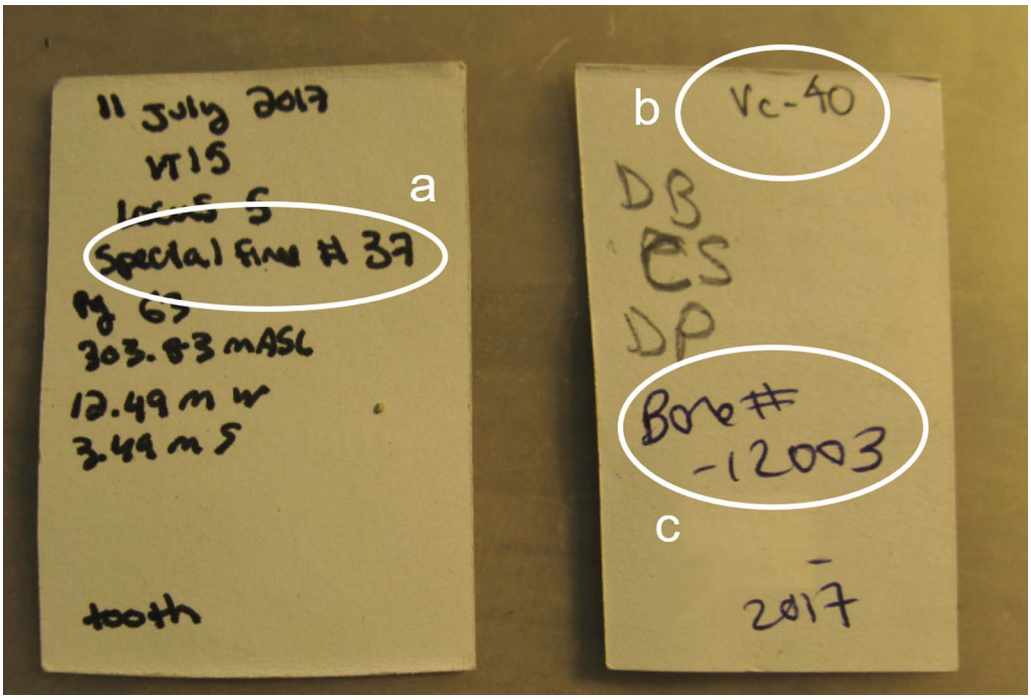

FIGURE 3. Multiple identifiers can be assigned to an individual artifact, ecofact, or sample by different researchers working at different times. This image shows three different identifiers assigned to a single object found in an excavation: (a) an identifier assigned to the find in the field and recorded in the field notebook; (b) another identifier assigned in the conservation laboratory and entered into the conservation notebook; (c) and yet another assigned by the specialist, perhaps years later, and entered into the specialist database. How do we track these various identifiers reliably and ensure that they are all properly associated with the single physical object to which they refer? How do we ensure that the observations made on that object are brought together? Photo credit: Sarah Whitcher Kansa.

one example of how archaeology can participate in informatics programs that can catalyze greater cross-disciplinary collaboration.

Data literacy requires more than familiarity with statistical methods and R or Python programming skills. Ethical, interpretive, and theoretical considerations cannot be divorced from data collection, modeling, analysis, presentation, and preservation. Consequently, data literacy must emphasize the ability to interrogate datasets and analytic steps in order to consider their implicit biases, blind spots, and tacit assumptions. Data literacy also helps to develop thinking about how data can and should relate to a broader context. That broader context can involve Linked Open Data kinds of modeling and integration with data curated by other expert communities. It can also mean greater critical awareness of how data should be interpreted and debated in different venues with different audiences.

The essays in this forum will explore a variety of topics, such as machine learning, big data, Linked Open Data, agent-based models, network analyses, 3D digitization programs, gaming, social media, preservation, and access. Like other areas of archaeological inquiry, these topics need debate informed by different schools of theoretical thought as well as Indigenous, feminist, and other perspectives. As demonstrated by this forum, broader and more inclusive participation in shaping data-driven research agendas promises to rekindle a flowering of creativity and renewed intellectual excitement in archaeology.

REFERENCES CITED

Beebe, Caroline
2017 Standard Descriptive Vocabulary and Archaeology Digital Data Collection. *Advances in Archaeological Practice* 5:250–264. DOI:10.1017/aap.2017.15.

Colaninno, Carol E., Shawn P. Lambert, Emily L. Beahm, and Carl G. Drexler
2020 Creating and Supporting a Harassment- and Assault-Free Field School. *Advances in Archaeological Practice* 8:111–122. DOI:10.1017/aap.2020.8.

Colwell, Chip
2016 Collaborative Archaeologies and Descendant Communities. *Annual Review of Anthropology* 45:113–127.

Counts, Derek B., Erin Walcek Averett, Kevin Garstki, and Michael Toumazou
2020 *Visualizing Votive Practice: Exploring Limestone and Terracotta Sculpture from Athienou-Malloura through 3D Models*. The Digital Press at the University of North Dakota, Grand Forks.

Data-Pop Alliance
2015 Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. *Data-Pop Alliance White Paper Series*. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Internews. Electronic document, https://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf, accessed November 23, 2020.

Faniel, Ixchel M., Anne Austin, Sarah Whitcher Kansa, Eric Kansa, Jennifer Jacobs, Phoebe France
2020 Identifying Opportunities for Collective Curation during Archaeological Excavations. *International Journal of Digital Curation* 15(1). DOI:10.2218/ijdc.v15i1.699.

https://doi.org/10.1017/aap.2020.55 Published online by Cambridge University Press

Faniel, Ixchel, Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel
2013 The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. In *JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 295–304. DOI:10.1145/2467696.2467712.
Gonzalez, Sara L.
2016 Indigenous Values and Methods in Archaeological Practice: Low-Impact Archaeology through the Kashaya Pomo Interpretive Trail Project. *American Antiquity* 81:533–549.
Gupta, Neha, Sue Blair, and Ramona Nicholas
2020 What We See, What We Don't See: Data Governance, Archaeological Spatial Databases and the Rights of Indigenous Peoples in an Age of Big Data. *Journal of Field Archaeology* 45:S39–S50. DOI:10.1080/00934690.2020.1713969.
Heath-Stout, Laura, and Elizabeth Hannigan
2020 Affording Archaeology: How Field School Costs Promote Exclusivity. *Advances in Archaeological Practice* 8:123–133. DOI:10.1017/aap.2020.7.
Kansa, Eric C.
2012 Openness and Archaeology's Information Ecosystem. *World Archaeology* 44:498–520.
Kersel, Morag M.
2015 Storage Wars: Solving the Archaeological Curation Crisis? *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 3:42–54.
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
2014 Grand Challenges for Archaeology. *PNAS* 111:879–880. DOI:10.1073/pnas.1324000111.
Leighton, Mary
2020 Myths of Meritocracy, Friendship, and Fun Work: Class and Gender in North American Academic Communities. *American Anthropologist* 122:444–458. DOI:10.1111/aman.13455.
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, Jade d'Alpoim Guedes, C. Michael Barton, Lynsey A. Bates, Michael Baxter, Andrew Bevan, Elizabeth A. Bollwerk, R. Kyle Bocinsky, Tom Brughmans, Alison K. Carter, Cyler Conrad, Daniel A. Contreras, Stefano Costa, Enrico R. Crema, Adrienne Daggett, Benjamin Davies, B. Lee Drake, Thomas S. Dye, Phoebe France, Richard Fullagar, Domenico Giusti, Shawn Graham, Matthew D. Harris, John Hawks, Sebastian Health, Damien Huffer, Eric C. Kansa, Sarah Whitcher Kansa, Mark E. Madsen, Jennifer Melcher, Joan Negre, Fraser D. Neiman, Rachel Opitz, David C. Orton, Paulina Przystupa, Maria Raviele, Julien Riel-Savatore, Philip Riris, Iza Romanowska, Jolene Smith, Néhémie Strupler, Isaac I. Ullah, Hannah G. Van Vlack, Parker VanValkenburgh, Ethan C. Watrall, Chris Webster, Joshua Wells, Judith Winters, and Colin D. Wren
2017 Open Science in Archaeology. *SAA Archaeological Record* 17(4):8–14.
Opitz, Rachel
2018 Publishing Archaeological Excavations at the Digital Turn. *Journal of Field Archaeology* 43:S68–S82. DOI:10.1080/00934690.2018.1505409.
Opitz, Rachel, Marcello Mogetta, and Nicola Terrenato (editors)
2016 *A Mid-Republican House from Gabii*. University of Michigan Press, Ann Arbor. Electronic document, https://hdl.handle.net/2027/fulcrum.n009w229r, accessed January 6, 2020.

AUTHOR INFORMATION

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

https://doi.org/10.1017/aap.2020.55 Published online by Cambridge University Press