Refining Archaeological Data Collection and Management

Michael Heilen and Shelby A. Manney

ABSTRACT

Most archaeological investigations in the United States and other countries must comply with preservation laws, especially if they are on government property or supported by government funding. Academic and cultural resource management (CRM) studies have explored various social, temporal, and environmental contexts and produce an ever-increasing volume of archaeological data. More and more data are born digital, and many legacy data are digitized. There is a building effort to synthesize and integrate data at a massive scale and create new data standards and management systems. Taxpayer dollars often fund archaeological studies that are intended, in spirit, to promote historic preservation and provide public benefits. However, the resulting data are difficult to access and interoperationalize, and they are rarely collected and managed with their long-term security, accessibility, and ethical reuse in mind. Momentum is building toward open data and open science as well as Indigenous data sovereignty and governance. The field of archaeology is reaching a critical point where consideration of diverse constituencies, concerns, and requirements is needed to plan data collection and management approaches moving forward. This theme issue focuses on challenges and opportunities in archaeological data collection and management in academic and CRM contexts.

Keywords: data collection, data management, data life cycle, data practice, Indigenous data governance, data integration, data reuse, synthesis

Globally, academic and cultural resource management (CRM) projects produce vast quantities of archaeological data. Most of these data are collected in compliance with preservation laws and regulations. A staggering amount of work has been done in US CRM since the passage of the National Historic Preservation Act of 1966 and the National Environmental Policy Act of 1969, for example (Altschul 2016a; McManammon 2018a, 2018b). Most of the money spent on US CRM efforts each year—$1.4 billion in 2020 (SRI Foundation 2020)—is devoted to performing basic compliance in response to individual projects when and where they arise. Despite the enormous scale and predictability of CRM work, there has been a limited effort to refine and operationalize data practices to meet today’s evolving management and research needs (Halford 2019; Schlanger et al. 2015). Many CRM efforts follow data collection standards and methods that have become stagnant and routinized without assessing their accuracy and reliability (Altschul 2016a; Halford 2019; Shott 1992). Moreover, many of the resulting data remain hard to locate, access, integrate, and reuse (Atici et al. 2013; Faniel and Yakel 2017; Kansa et al. 2019; Kintigh et al. 2015; McManammon 2018c; Richards et al. 2010; Watts 2011; York et al. 2018). As Schlanger and colleagues (2015:92) articulate, “We create data in greater detail and greater volume than ever before, yet we do not know what to do with much of it, and have few effective means of sharing it, managing it, analyzing it, or archiving it.”

Palabras clave: recopilación de datos, gestión de datos, ciclo de vida de datos, práctica de datos, soberanía de datos indígenas, integración de datos, reutilización de datos, síntesis
Projects such as ARIADNE and ARIADNEplus in Europe (https://ariadne-infrastructure.eu/) and Digital Antiquity (https://live-digant.ws.asu.edu/) and Open Context (https://opencontext.org/) in the United States have successfully archived and indexed many archaeological datasets and, in some cases, link datasets and provide tools for data integration. Nevertheless, many archaeological data are not collected either with their future integration and reuse in mind or in consideration of how they will be accessed and governed over the long term. In this journal issue, we envision a future where archaeological data are collected and managed in deliberate, purposeful ways that allow them to be (a) useful throughout the life of a project, (b) accessible and reusable after a project has ended, and (c) governed during and after a project in ways that honor the rights, needs, and requirements of diverse data constituencies and stakeholders. To do this, archaeological data will need to be intentionally modeled, collected, and managed in full anticipation of being accessed, interoperationalized, and reused by multiple constituencies for diverse and sometimes conflicting purposes. In this future, archaeological data will be collected and managed as living data that retain utility beyond the purpose for which they were generated—that is, data that are not only reusable but also “re-useful.”

As federal agencies such as the US Bureau of Land Management transition from “site-centric” to landscape approaches, they need the ability to rapidly integrate and interpret archaeological data from many projects over large areas and across jurisdictions. Synthetic and interdisciplinary research using “big” archaeological data will also require data to be integrated and managed at a scope, scale, and rate consistent with the level at which data are now collected (Altschul 2016b; Doelle et al. 2016; Halford 2019; Huggett 2020a). Yet, the bewildering variety of systems that archaeologists use to record and manage data are rarely designed to anticipate other practitioners’ research questions or how future projects or management approaches may reuse data (Kansa and Kansa 2021; Pavel 2010). This problem is further complicated because archaeological data collection methods often extract from or destroy nonrenewable resources, rendering data collection from the same context nonrepeatable (Lipe 1974, 2009; Pavel 2010, 2012; cf. Roosevelt et al. 2015). There are few opportunities for a do-over in archaeology and few examples of reproducible archaeological research.

Combining archaeological data from multiple projects to address a particular research question or management concern can be exceedingly time consuming and challenging in the current data ecosystem. Many questions may arise concerning how data were collected and recorded, what they represent, and how they can or cannot be successfully standardized, harmonized, leveraged, and reused. Some of these questions will be difficult or impossible to answer. The extra time and effort needed to prepare and validate a combined dataset detract from the amount of time and effort that can be applied to exploring and learning from the dataset. By failing to plan for long-term access, interoperability, and reuse, existing data practices diminish the ability to leverage archaeological data for research and management. Ultimately, this limits what can be learned from archaeological data about history, the environment, and the human experience. By contrast, quickly and confidently locating, accessing, and integrating data from multiple projects can dramatically increase research efficiency (Beagrie and Houghton 2013). Archaeologists need to find ways to exchange the vast amount of effort now required to wrangle data into useable and reliable data products with effort spent instead on interacting with data to gain knowledge and insight.

Refining archaeological data collection and management requires consideration of the entire workflow and data life cycle—and how data are to be governed throughout the data life cycle. Methods and practices implemented in one phase of the data life cycle will impact the outcome of other phases positively and negatively (Kansa and Kansa 2022; Yakel et al. 2019). Archaeologists and cultural resource management practitioners need to step away from the project-by-project mindset and view their efforts as part of a larger physical and metaphorical landscape of archaeological research and heritage work. Although the project-based, archival focus of the current CRM data ecosystem can provide acceptable management results in isolation, far better research and management outcomes could be achieved if archaeologists conceptualized and worked with data at broader and more inclusive scales. We must transition from thinking of data as a means to accomplish a project goal to thinking of data more holistically as valuable (and often public) assets that can provide extended benefits over the long term for multiple constituencies and purposes. This transition will need to consider heritage values more broadly, collecting data to be interoperable and reusable, and managing data following centralized, standardized, and inclusive methods. Archaeologists need to adopt an ethic of data preservation and extensibility.

The authors in this issue of Advances in Archaeological Practice all have a common theme to think “bigger” about archaeological data and the data life cycle. By bigger, we are referring not only to the scale at which archaeological data can be aggregated but also to how archaeologists transform data into narratives, meanings, and conclusions. Furthermore, archaeologists and cultural resource practitioners must consider how the collection, management, and dissemination of archaeological data impact, help, or imperil living descendant cultures and communities. From various points of view, the authors explore the complexities surrounding the data life cycle. The authors in this issue present examples of approaching archaeological data as living data rather than static siloed data that are collected, analyzed, and utilized for a singular short-term purpose.

This introductory article provides a foundation for the others in this issue by defining archaeological data, overviewing the data life cycle, and identifying some of the legal requirements surrounding their collection and management. We then outline what we see as archaeology’s “data problem.” Next, we discuss how articles in this issue address data collection and management challenges. We conclude with a “call to action” for all archaeologists to take concrete, positive steps toward improving archaeological data collection and management methods and practices.

**ARCHAEOLOGICAL DATA AND DATA PRACTICES**

Perhaps more than ever, archaeologists need to be cognizant of data practices—both their own and their colleagues—and how those practices affect research, management, and heritage. A preliminary step required for “thinking bigger” about data is for archaeologists to broaden their understanding of data as part of a
continuous life cycle. Although definitions vary, the data life cycle refers to the main stages and transformations that data take as they move from the planning of data acquisition to data recording, processing, analysis, interpretation, dissemination, curation, and reuse (Borgman 2019; Faniel et al. 2018; Liu 2021; McManamon and Ellison 2022; Williams and Williams 2019; Yakel et al. 2019). Archaeologists should plan for each stage of the data life cycle and consider, throughout this cycle, ethical implications and access.

Archaeological data are diverse and come in many forms. A great deal of archaeological data documents the location, characteristics, disposition, and context of material remains observed during field or laboratory studies. Archaeological data also include ethnoarchaeological studies, historical contexts, administrative topics, legal records, and computational data from analyses, experiments, and simulations (Huggett 2022; Kansa and Kansa 2021; Kintigh et al. 2018; Manwick and Pilaar Birch 2018). By themselves, archaeological data have no inherent meaning. Data must be processed, structured, and interpreted to derive meaning and context. To effectively share, integrate, and reuse archaeological data, metadata are needed that convey how datasets are structured, standardized, related, and encoded. Paradata are needed to document the methods, limitations, and contexts in which data were collected and analyzed (Atici et al. 2013; Börjesson et al. 2022; Huggett 2020b; McManamon and Ellison 2022). Without metadata and paradata, identifying and controlling for bias when integrating and (or) reusing data can be difficult or impossible. Controlling for bias requires transparent documentation of the methods used throughout the data life cycle, standardized data practices, and consideration of whether data collection methods are adequate to meet evolving research and management needs.

### ACCOUNTING FOR BIAS IN DATA COLLECTION

Some of the most significant impacts on archaeological data and the data life cycle stem from how data are collected and recorded (Yakel et al. 2019). If archaeologists put effort into rendering data more interoperable and reusable, they also need to consider whether the methods used to collect data will be effective for that purpose. Many conceptual, environmental, and methodological factors can bias archaeological data collection, thereby affecting the outcome of all subsequent stages in the data life cycle. For example, artifact identification methods, spacing intervals, lighting, artifact characteristics, density, and surface visibility affect what archaeologists identify and record during a survey. Researchers have studied and modeled the effects of several biases on field research, and many have proposed methods to identify and control for bias (Banning 2002; Banning et al. 2006, 2011, 2016; Gnaden and Holdaway 2000; Hawkins et al. 2003; Heilen and Altschul 2013; Plog et al. 1978; Schiffer et al. 1978; Shott 1995; Shott et al. 2002; Stewart 2006; Wandsnider and Camilli 1992). Yet, there has been little interest in accounting for bias in CRM practice and little emphasis on managing the kinds of metadata and paradata needed to control for bias when data are reused.

A significant source of bias stems from uncritical reliance on the site concept. CRM efforts routinely carve archaeological landscapes into arbitrary pockets of time, space, and data—that is, sites. Although sites can serve as convenient units of analysis and management, relying on them as fundamental data collection units can distort the interpretation of the archaeological record (Colwell and Ferguson 2014; Dunell 1992; Ebert 1992; Leckman and Heilen 2023; McCoy 2020) as well as frustrate reuse of primary data. Many survey approaches, for example, focus observation primarily on recording high-density scatters and rare artifact types. Wandsnider and Camilli (1992:182) point out that “the perception that the archaeological record consists of rare ‘hot spots’ in high artifact density and just a few dispersed artifacts is heavily reinforced by traditional discovery techniques.” Focusing on only one segment of the archaeological record can lead to erroneous interpretations of behavioral patterns and cultural processes and, ultimately, to misapprehension and mismanagement of the archaeological record. In this issue, Leckman and Heilen (2023) present a pedestrian survey method that allows systematic and transparent delineation of sites and site components required for legal compliance and management while simultaneously generating standardized, reusable data that archaeologists can aggregate into cumulative datasets suitable for addressing unanticipated research questions, with or without sites as units of analysis (Heilen and Murrell 2015; Miller et al. 2009; O’Leary et al. 1997; Seaman et al. 1988). They show how the method allows archaeologists to aggregate relatively fine-grained primary data from multiple survey projects to infer landscape-level patterns and processes, such as identifying precontact foot trails and mobility patterns across their area of study (Miller et al. 2018).

Another form of bias in data collection is that archaeologists often base survey results on a single snapshot in time. When another set of archaeologists resurveys an area, they often assume that the latest recording episode is the most authoritative and accurate (Heilen et al. 2008). Field experiments have shown that surface archaeology is not static, however. What is exposed and visible to archaeologists in the field varies through time due to disturbance, ongoing and episodic earth surface processes, and changes in field conditions, such as those following rain, plowing, vegetation change, wildfire, or erosion. Moreover, dynamic geomorphic processes and ground conditions differentially influence the observable distribution of artifacts and features across the landscape. For this reason, repeat observation of geomorphic landscape elements at different points in time can be essential to accurately and reliably interpreting a surface assemblage (Holdaway and Fanning 2008; Shott 1995; Shott et al. 2002). In this issue, Douglass, Holdaway, and Wandsnider (2023) consider what archaeologists can learn when they incorporate repeat observation and a geomorphological landscape perspective into field surveys. With revised field methods that account for the dynamic nature of archaeological landscapes, archaeologists can arrive at entirely new interpretations of culture history and land use. Without revised methods, they argue, the archaeological sample preserved for future generations is likely to be biased and unrepresentative (see, for example, Lipe 1974, 2009).

Data collection biases can also limit what archaeologists and land managers learn through consultation with stakeholders and how they apply that information to research and management. Most consultations are project specific and focus only on sites that may be affected by a project (Anschuetz and Dongoske 2017; Ferguson 2009; Watkins and Ferguson 2005). Landscape elements and characteristics that may be of equal or greater interest to stakeholders—such as important plants, water resources, habitats, viewsheds, soundscapes, and landmarks—are often ignored or
underappreciated (Colwell and Ferguson 2014; Ferguson et al. 2015; Heilen and Altschul 2019; Stoffle et al. 2009; Toupal et al. 2001; Zedeño et al. 1997). The way some archaeological investigations are conducted has the effect of omitting, ignoring, or diminishing Indigenous perspectives while simultaneously colonizing, appropriating, and disrupting Indigenous heritage and lifeways (Dongoske and Anschuetz 2021).

Sarah Schlanger (2023) addresses archaeology’s data problem in this issue from a land management perspective that broadly considers the diverse and sometimes conflicting values that Indigenous communities and other stakeholders place in cultural resources and landscapes (Lipe 2009; McManamon et al. 2016). Schlanger argues that managers must consider how different constituencies value a landscape and its characteristics, and how they prefer to experience those values. She recommends that data collection methods for managing culturally significant landscapes should include inventorying and monitoring landscape characteristics integral to stakeholders’ experience of place, instead of placing emphasis primarily on the documentation and avoidance of individual archaeological sites and other sensitive resources.

These are a few ways to refine data collection methods to produce more meaningful and reusable archaeological data, but they are by no means the only or the right way. The point of these articles is that if we are going to plan for a more visionary future that makes better and more holistic and ethical use of archaeological data, we must start with consideration of how the data are collected and whether those data will serve archaeology and the public well—both now and in the future. Revised collection methods will have to address biases, allow for interoperability, consider descendant community concerns, and anticipate refinements in subsequent steps of the data life cycle.

DATA MANAGEMENT

After data collection, the data life cycle usually ends prematurely, in “data death,” with the completion of a project. For archaeologists to think bigger, data must be planned, integrated, and managed more effectively at broader scales. The only way archaeology can prepare for future access and reusability of archaeological data is by managing its data assets at every stage of the data life cycle (Niven 2011). A data management plan that addresses the entire data life cycle is now considered essential for every archaeological project (Baker and Duerr 2017; Kansa and Kansa 2021; Kansa et al. 2019). However, academic researchers, land managers, and CRM contractors continue to manage a great deal of archaeological data without clear and detailed data management plans and without confronting difficult questions regarding how to care for archaeological data over the long term (McManamon and Ellison 2022; McManamon et al. 2017; Wright and Richards 2018). The data life cycle must not end with a publication or contract completion. Rather, archaeological data need to live on so that they are permanently curated and accessible, and so that they can be reused for research and management needs different from the immediate purpose for which they were initially generated.

Extensive efforts have been undertaken in Europe to improve the management and accessibility of archaeological data by integrating digital repositories and datasets across Europe with the ARIADNE and ARIADNEplus projects. Digital Antiquity and Open Context have been fulfilling similar archiving and cataloging functions in the United States. The vast majority of archaeological data have yet to find their way into a digital archive, let alone in ways that support their interoperability and reuse.

In the United States, many states and federal agencies (e.g., the Army Corps of Engineers, Bureau of Land Management, Forest Service, National Park Service) now have digital CRM databases containing data on thousands of projects and sites (Halford and Ables 2023; Wilshusen et al. 2016). Most store derivative, summary information about archaeological sites and project areas—not the primary data. The data within and between databases also varies in quality, reliability, consistency, standardization, and completeness. In many cases, access to precise locational data is restricted, diminishing the saliency and reusability of archaeological data for many avenues of research (Ortman and Altschul 2023; Robinson et al. 2019). Many primary data are difficult or impossible to access, buried in site cards and reports, or even stored on a personal computer, often without adequate metadata and paradata. As a result, the conclusions of an archaeological project often “cannot be immediately followed or traced back to primary evidence” (Pavel 2010:6). Even when data are available, many data continue to be modeled and managed on a project-by-project level, primarily for archival rather than analytical purposes, and following diverse data structures and formats, minimal standardization, and limited adherence to standard ontologies and typologies (Kansa and Kansa 2021; Kintigh et al. 2018; Marwick 2017).

Many archaeologists may not realize that in the United States, by regulation, archaeological data collected with federal funds or on federal property must be appropriately stored, made accessible to potential users, and managed in perpetuity (Cultural Heritage Partners 2012). As Clarke (2015:322) advises, “Recent changes in academia and federal funding have made it so that there is no longer a choice, and scholars are now legally responsible for preservation of research data.” Multiple federal regulations and statutes have authorized curation and access to data, records, reports, samples, physical artifacts, and other remains. These include the American Antiquities Act of 1906, the Archaeological Resources Protection Act of 1979 (as amended), the National Historic Preservation Act of 1966 (as amended), Federal Regulation 36 CFR 79 (Curation of Federally Owned and Administered Archeological Collections), public records laws, and National Archives regulations.

The state of archaeological data management in the United States has created significant barriers to effectively managing the archaeological record by lacking shared data standards and limiting access to information by those who need it (Halford 2019). The US Bureau of Land Management has taken steps to remove some of these barriers within its agency by creating the National Cultural Resource Information Management System (NCRIMS), as presented in this issue by Halford and Ables (2023). NCRIMS uses complex Python scripts unique to each contributing database to annually extract, translate, and upload data into the system. Although the system allows authorized users to rapidly assess cultural resource distributions over large areas following a landscape approach to managing public lands (Altschul 2016b; Clement et al. 2014), it also begs the question of why such a system is necessary. Are archaeological contexts and research questions in individual states so different from each other that they...
must be maintained following separate data standards and formats? Is it possible to develop a shared standard that facilitates the integration of most archaeological data while still allowing room for experimentation and refinement?

Archaeological data are increasingly born digital and could be originated and maintained within integrated data management systems following common standards (Niven 2011). As Sarah Kansa and colleagues (2019:2) note, when “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.” Archaeology can contribute uniquely to a deep-time perspective on contemporary issues with significant public policy implications, such as social inequality, migration, climate change, wildfire, violence, urbanization, disease, and health (Altschul 2016a; Altschul et al. 2017, 2018; Kintigh et al. 2014; Ortman 2019; Ortman et al. 2020). Yet, few steps have been taken to realize this capability (Arbuckle et al. 2014; Austin 2014; Benden and Taft 2019; Richards 2017; Schlanger et al. 2015).

One solution is adhering to the open data and open science principles (Kansa 2010, 2012, 2015; Marwick et al. 2017).

Open Data and Open Science

The “open” concept refers to making data, information, and knowledge maximally interoperable and freely accessible in open formats to anyone for any purpose, with no institutional barriers to their access. Open data are maintained using clear, unambiguous terms that are understandable by all; fully described to allow users to understand the data’s strengths, weaknesses, and limitations; and released as early as possible to allow time for data review and refinement. Crucially, open “data will be in their original, unmodified form and at the finest level of granularity available” (Moore and Richards 2015; UK Government Cabinet Office 2013: Principle 2 [https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex]; see also Kansa 2012). Proponents of open data in archaeology anticipate that improved access to primary archaeological data will provide greater accountability, transparency, and repeatability in archaeological research. This change could, in turn, revitalize archaeological research, lead to new insights and creative reuse, and improve disciplinary interaction (Kansa 2012; Marwick et al. 2017; Moore and Richards 2015).

For archaeological data to be more open, data products must be findable, accessible, interoperable, and reusable (FAIR; Wilkinson et al. 2016, 2019). In this issue, Nicholson, Kansas, Gupta, and Fernandez (2023) emphasize that FAIR principles apply to a wide range of primary and derivative data products. However, their application in archaeology remains minimal for a variety of reasons. There are currently few incentives for individual researchers or contracting firms to share data, and there are concerns that data sharing will expose errors or preempt publication, fail to credit data creators, and encourage misappropriation or mishandling of sensitive data by those who gain access (Atici et al. 2013; Marwick 2017; Marwick and Pilaar Birch 2018; Moore and Richards 2015).

Eric Kansa (2012:507) warns that withholding archaeological data undermines the integrity of the archaeological record and can amount to the “personal, self-aggrandizing appropriation of cultural heritage.” By contrast, sharing primary data accompanied by adequate metadata and paradata can support reproducibility of research results (Marwick 2017). It can also “foster adherence to more consistent and comparable approaches to the recording and reuse of data” (Kansa et al. 2019:2). For shared data to be usable, standard vocabularies and “better methods to model, create, clean, and document” are needed, along with reward systems that encourage good practice in generating and sharing reusable data (Beebe 2017; Heckman et al. 2020; Kansa et al. 2019:2).

A concern is that open data initiatives place inadequate attention on “the impacts of colonialism on the practice of science and have overlooked the interests and rights of Indigenous peoples when it comes to data ownership, sharing, and knowledge creation” (Gupta et al. 2020:539). Applying FAIR principles to archaeological data requires that archaeologists account for the colonial roots of archaeological research, fundamental differences between scientific and Indigenous epistemologies, and the sovereign rights of Indigenous Peoples to control their heritage, including controlling any data related to that heritage (Dongoske and Anschuetz 2021; Ross et al. 2011). Similar concerns are emerging among other stakeholder groups, such as African American and Asian American communities seeking greater control over and access to heritage. Consequently, although there is a growing interest in making archaeological data FAIR, there is a concomitant need to decide how archaeological and other heritage data should be governed.

Data Governance

Data governance frameworks require the archaeological community to address numerous questions about how data can be accessed, shared, processed, documented, and controlled; what permissions and permission systems are needed; and who has the right to decide on data practices. Indigenous data governance is an emerging topic discussed in this issue by Gupta, Martin, Supernant, Huddleston, and Elvidge (2023) that affects all aspects of the data life cycle. Indigenous data are data of any form that “impact Indigenous Peoples, nations, and communities at the collective and individual levels,” including data on their environment, history, and heritage (Carroll et al. 2021:1; Kukutai and Taylor 2016). The United Nations Declaration on the Rights of Indigenous Peoples reaffirms that they have a right to access, enjoy, practice, revitalize, and preserve their heritage, including their right to control Indigenous data. Yet, many data are not even findable or accessible to the communities that might claim them as part of their heritage and have instead “been kept from Indigenous communities, often inadvertently” (Nicholas and Bannister 2004:339). Archaeologists, archaeological institutions, cultural resource firms, and government entities control most archaeological data in the United States, Canada, and other countries. Advocates of Indigenous data governance insist that Indigenous communities have a sovereign right to control Indigenous data and govern how they are collected, accessed, used, and disseminated (van Schilfgaarde 2020). As Nicholas and Bannister (2004:339) explain, “Whoever owns (or controls records of) the past also owns or otherwise shapes the future of that past.”

These concerns have led to the development of the CARE Principles for Indigenous Data Governance (https://www.gida-global.org/care; see Gupta et al. 2023, this issue). First drafted in 2018, the CARE Principles promote collective benefit, authority to control, respect, and ethics in practices involving Indigenous data. Establishing Indigenous data sovereignty through data governance frameworks would reposition Indigenous Peoples from being
subjects of inquiry to being active participants and leaders in heritage research, policy, and management (Carroll et al. 2021). To some archaeologists, Indigenous data governance frameworks may appear to divest archaeologists of control over how archaeologists collect, use, and share data. On the other hand, such frameworks also provide an opportunity to decolonize archaeology and redress past wrongs while fostering greater collaboration with Indigenous Peoples and greater respect for Indigenous rights and interests. Adherence to Indigenous data governance frameworks could also help archaeology achieve greater relevance by (a) refining data practices to better support the protection of Indigenous rights, interests, and heritage; and (b) enriching research and policy with knowledge and perspectives that Indigenous communities choose to share (Atalay 2012, 2020; Laluk et al. 2022; Larson et al. 2020; Ogar et al. 2020). In the future, data governance frameworks will likely be viewed as essential tools for guiding archaeological data practices and the data life cycle.

RECONCEPTUALIZING ARCHAEOLICAL DATA AS LIVING DATA

The life cycle of archaeological data often results in the data being put to rest and largely unavailable for reuse, even though US regulations—and those of many other nations—require archaeological data to be preserved based on regulated data access and storage standards and in a manner that makes data both accessible and reusable in the future (Kintigh 2018).Archaeological data need to be reconceptualized as living data that retain utility beyond the immediate purpose for which they were generated. Ideally, academic archaeology and CRM would benefit from adopting a more unified set of methods and standards with baseline assumptions that allow for more robust interoperability and comparison among datasets while simultaneously providing room for experimentation. CRM needs to shift from thinking of data as an administrative record of individual projects, sites, and deposits to thinking of data as part of a living, integrated digital ecosystem that includes the following:

- Published ontologies and data dictionaries
- Collaborative tools and methods for creating ontologies and normalizing and harmonizing data
- Tools and methods for uploading, citing, cross-referencing, and linking datasets
- Permission systems and data tags for controlling access and use of culturally sensitive or restricted data
- Standards, templates, and guidelines for recording the metadata needed to locate, display, and analyze the data—as well as the paradata needed to understand the methods, processes, and quality associated with them
- Centralized systems for ethically ingesting, managing, and disseminating archaeological data at scales consistent with the scale at which data are collected

A final stage in the data life cycle is data reuse. In this issue, Ortman and Altschul (2023) confront the problem that there currently is no system in the United States or elsewhere for integrating and digesting archaeological data at the scale and scope in which archaeologists collect data, let alone in ways that make archaeological data directly useful for the social science research community and other disciplines. In other social science disciplines—such as demography, economy, geography, and sociology—a well-staffed government agency such as the US Census Bureau is responsible for translating raw data into standard data products for regional- and national-scale research. Although researchers and organizations have made significant advances in integrating archaeological data, they tend to follow one of three strategies: (1) aggregating data of a restricted type, such as radiocarbon dates (Bird et al. 2022; Kelly et al. 2022) or tree-ring dates (Robinson and Cameron 1991); (2) creating digital archives or indexes for finding and accessing archaeological datasets (Kansa 2010; Kintigh et al. 2018; Richards 2017; Sheehan 2015; Wells et al. 2014); or (3) creating integrated research databases for a particular region or research problem (Hegmon et al. 2021; Mills et al. 2020; Neusius et al. 2019; Pluckhahn and Wallis 2017). Ortman and Altschul (2023) present a vision to develop a national archaeological data service for the United States that would be built around the most common and consistent data that archaeologists record—spatial data—with built-in mechanisms for protecting the sensitivity of those data and governing how they may be accessed and used (see also McKeague et al. 2020).

CONCLUSIONS

This journal issue is ultimately about refining data practices to improve archaeological data’s durability, utility, and ethical reuse. In effecting change in practice, archaeologists must recognize that different incentives exist for different data constituencies, such as federal CRM programs, CRM contractors, academic archaeologists, Indigenous communities, and other disciplines. Despite the substantial investment of government time and expense in the industry, there are few incentives within CRM to experiment with new methods or synthesize data across many projects unless required by a project. Refining archaeological data practices within CRM and for many academic projects may ultimately require implementing data management plans, and practices that (a) adhere to demonstrably effective, ethical, and sustainable data standards and systems; and (b) provide for long-term data access, interoperability, and reuse.

A comparative framework for assessing the costs and benefits of archaeological data collection and management methods is needed to determine which methods will be the most effective and worthy of investment. Archaeologists should consider the entire workflow when refining methods, including the human and digital workflows involved in generating, managing, integrating, analyzing, transforming, curating, and sharing data. Moreover, Indigenous rights and perspectives should be considered, along with the values placed on cultural resources by different constituencies and user groups. Education in digital literacy and ethics, as well as behavior change, are needed to realize change in archaeological data practices. In collaboration with other stakeholders, archaeologists must decide which data are critical to collect and how they should be standardized, encoded, related, and made available to researchers, stakeholders, and the public. What makes these data useful, and to which constituencies are they relevant? How will they be controlled and used now and in the future?
Refining data collection and management practices across the data life cycle will take time, behavioral change, disciplinary “buy-in,” and systemic reform. Multiple tensions will need to be resolved, including deciding on the purposes and goals of archaeological data and data practices, and the degree to which data access and use should be open or restricted. A perfect system with all its parts seamlessly integrated and deconflicted will not emerge spontaneously in whole cloth. Even so, it is time that archaeologists evaluate how their data practices affect different components of the data life cycle and begin collecting and considering data for this more visionary future.

Acknowledgments

We would like to thank the authors of the articles in this issue for their insightful and visionary contributions and for their persistence and dedication in conceptualizing and advocating for refinements in archaeological data collection and management. This article and many other articles in this issue are based on an organized symposium, sponsored by the Arizona Army National Guard, that was held virtually in April 2021 in the midst of a global pandemic during the 86th Annual Meeting of the Society for American Archaeology. We are grateful to all the symposium participants for their contributions to the topics discussed in this issue, including authors who were unable to participate here. We greatly appreciate the many colleagues who peer-reviewed articles in this issue. No permits were required for the work described here.

Funding Statement

We would like to thank the Arizona Army National Guard for its generous support—in the form of staff time, resources, and funding—in helping this special issue come to fruition and in providing funds to support this issue, including authors who were unable to participate here. No permits were required for the work described here.

Data Availability Statement

No original data were used in this article.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

REFERENCES CITED


Kintigh, Keith W., Katherine A. Spielmann, Adam Brn, K. Seljuk Candan, Tiffany C. Clark, and Matthew Peebles. 2018. Data Integration in the Service of...