THE EFFIGY MOUNDS SCANDALS · DISAPPEARING DIGITAL DATA · A LEGENDARY ARCHAEOLOGIST

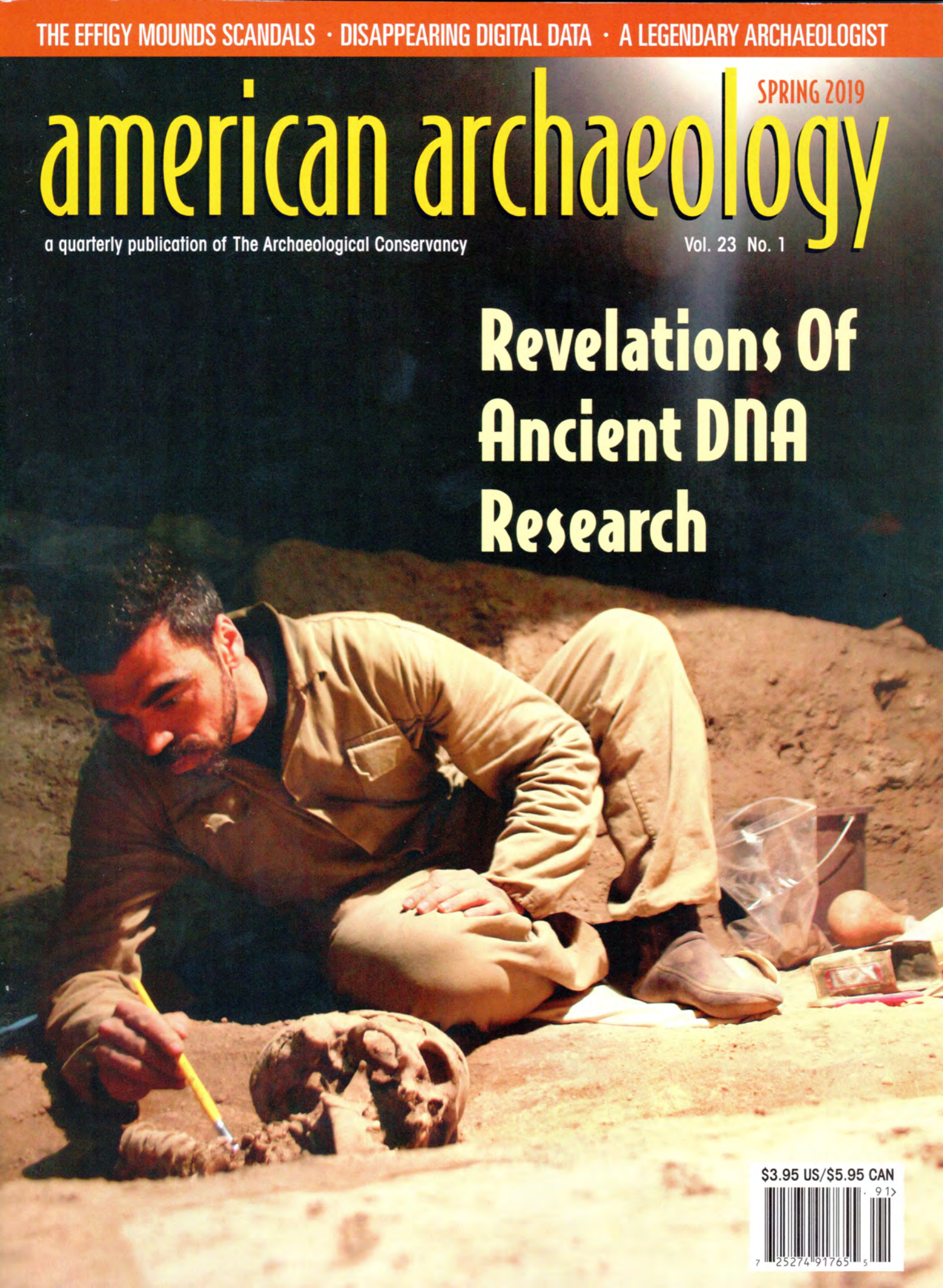
SPRING 2019

american archaeology

a quarterly publication of The Archaeological Conservancy

Vol. 23 No. 1

Revelations Of Ancient DNA Research



$3.95 US/$5.95 CAN

91

7 25274 91765 5

Over the years, archaeological digital data has been stored on devices such as magnetic tape and various disks shown here, none of which can be read by today's computers. If the data can't be accessed, it has, for all practical purposes, disappeared.

COURTESY OF US ARMY CORPS OF ENGINEERS / PHOTOGRAPHER GEORGE GONZALEZ

MICHAEL "Sonny" Trimble has seen many things in his nearly thirty-two-year career as an archaeologist with the U.S. Army Corps of Engineers. He worked on archaeological records and objects from African Burial Ground National Monument in Manhattan, which dates to the seventeenth and eighteenth centuries. He delivered the contested remains of Kennewick Man to the Burke Museum in Seattle, where the 8,500-year-old skeleton was kept for twenty years before being reburied. He also led the forensic investigation of several mass graves in Iraq, and then faced Saddam Hussein in court, detailing the scientific findings that helped convict Hussein of crimes against humanity.

When Trimble worked in Iraq he was protected by American soldiers. When some of those soldiers returned home, their tours of duty over, they contacted Trimble in hopes that he would return the favor by helping them find employment. Though Trimble wanted to help, there wasn't much that he could do.

But then a thought occurred to him. The Army Corps of Engineers' is the custodian of vast archaeological collections that, according to him, are second only to the Smithsonian's. The collections represent the tremendous amount of archaeological work that was conducted as the Corps built dams and developed other major projects. So Trimble initiated the Veterans Curation Program to train veterans to help catalogue the Corps' collections. In laboratories around the country, veterans now photograph, scan, and publicly archive information about everything from a prehistoric quartz quarry in Georgia to a rockshelter in Indiana. The data ends up in a digital repository hosted at Arizona State University in Tempe, known as the Digital Archaeological Record, or tDAR. Sensitive information, such as the specific location of certain sites, is withheld—but the veterans upload everything else to the public site. "That's the platform where everybody can get to this stuff no matter what their interest is," said Trimble, who retired in December. "It democratizes it."

Over the past few decades, researchers in the United States have performed tens of billions of dollars' worth of cultural heritage work and recorded nearly a million archaeological sites. Federal laws such as the 1966 National Historic Preservation Act require that the resulting data be preserved for future generations. Keith Kintigh, an archaeologist

Disappearing Digital Data

Archaeological projects are generating huge amounts of digital data that researchers are ethically, and in some cases, legally, obligated to preserve and make accessible. But the digital data is often lost due to imporper curation.

By Alexandra Witze

at Arizona State University and a founder of tDAR, pointed out that the act's preamble says that preserving the nation's heritage "is in the public interest so that its vital legacy of cultural, educational, aesthetic, inspirational, economic, and energy benefits will be maintained and enriched for future generations of Americans."

Nonetheless, most archaeologists aren't curating their digital data at public repositories like tDAR. Kintigh stressed that proper curation is more important than ever because now much of this information is, so to speak, "born digital" and exists in no other form. Without it, future generations of scientists won't be able to reanalyze and synthesize the information and make fresh discoveries of their own. "It's a tragedy that we're not adequately capitalizing on the potential uses of the data," he said.

INSTITUTIONS face many challenges in making their data accessible. "Yes, we all have these problems. Punch cards, floppy disks, and every other outdated form," said Wendy Teeter, curator of archaeology at UCLA's Fowler Museum. "It's a huge problem that is getting worse." Storage media, like CDs, physically deteriorate over time and eventually become obsolete. "You get someone with an archaeological report on a 3.5-inch floppy disk, and no one can access it anymore," said Alison Rubio, an archaeologist with the Air Force Civil Engineer Center in San Antonio, Texas.

Even if the files can be opened, the data may not be retrievable if they were stored using outmoded software. Files need to be backed up and checked regularly to be sure they have not become corrupted, and migrated as software changes. And if the files are not adequately documented with metadata (data that provides information about the data), they may be useless. Digital files require much more active curation than a physical collection of artifacts does.

Most archaeological collections managers are not experts in managing data. They don't have an IT staff that can keep them up-to-date in the latest technologies, or a budget that allows them to purchase expensive cloud-based storage. That means the data are inaccessible without a visit to the repository, they are not checked for readability, the files are not migrated to remain useable as new standards arise, and there is no review to make sure that the metadata

are sufficient to make the files useable by another researcher. Eventually the data become inaccessible and, for all practical purposes, disappear.

"My estimation is that the problem is widespread," said Danielle Benden, owner of Driftless Pathways, a small museum consulting firm in Wisconsin, and chairwoman of the Society for American Archaeology's committee on museums, collections, and curation. Hoping to help solve this problem, Kintigh and others founded tDAR as a repository where archaeologists could deposit their data. They envisioned it as a place where agencies could easily upload archaeological reports, datasets, and images that would then be made accessible to other researchers. Funded in its early stages by the National Science Foundation and the Andrew W. Mellon Foundation, tDAR went live in 2010.

Today, it contains more than 395,000 documents, 21,000 images, and 1,700 databases that document the results of some 1,100 archaeological projects. Researchers can upload their files and metadata themselves, or the staff at the ASU's Center for Digital Antiquity, the organization that operates tDAR, will help them do it. This kind of detailed curation takes time and energy. "It's not just something you do after a day of field work, or in your spare time," said Francis McManamon, an archaeologist at Arizona State and the founding director of tDAR. "It's something that needs focus and attention, just the way that caring for physical collections does."

tDAR has not taken off the way its founders had hoped. Some agencies and researchers are using tDAR or other repositories to make their data discoverable, accessible, and usable, but many are not. "We've been increasingly frustrated over the years," said Kintigh. He and McManamon said they frequently talk to project managers who say they want to make their archaeological data public, but who nonetheless don't put their material into tDAR or other public repositories. "We've tried the arguments that this is your professional obligation and your legal responsibility," Kintigh said. "But we're not getting the action that is urgently needed."

Many academic funding organizations, such as the National Science Foundation, require scientists who are

COURTESY US ARMY CORPS OF ENGINEERS

A veteran takes a photograph of a lithic object that was recovered during an archaeological project associated with the U.S. Army Corps of Engineers. The Corps' Veterans Curation Program, which is now ten years old, trains veterans to digitize data from the Corps' huge archaeological collections. The digital data is then uploaded to tDAR.

Sierra M. Bow, a Ph.D. student at the University of Tennessee, uses a pXRF machine to analyze paint on rock art at Painted Bluff, in Alabama. The digital data from this project was indexed by the Digital Index of North American Archaeology and hosted by Open Context to link with related data and publications from across the web.

STEPHEN YERKA

submitting a grant proposal to include a plan for how they are going to manage the data. A 1990 set of regulations known as 36 C.F.R. 79 lays out how federally-owned archaeological collections should be curated. That regulation and other federal laws require that records must be maintained and made accessible—and that includes digital records. So agencies that don't ensure the long-term preservation of the data for which they are responsible or don't make those data publicly available are breaking federal law, Kintigh argued.

For years McManamon has been meeting with groups such as the Advisory Council on Historic Preservation, which helps federal agencies meet their legal requirements under the National Historic Preservation Act. He also meets regularly with federal agency archaeologists to argue that their data need to be in a public repository. "Almost every archaeologist and cultural resource manager agrees that digital data curation is needed," he said. "In many offices they say they just don't have the funding to do it."

But Kintigh is unconvinced by this excuse. He noted that there are about 40,000 federally-mandated projects done every year, all of which have funding. The agencies simply need to dedicate a small fraction of the project's funds (he estimated one to three percent) to digital curation, just as they do for artifact curation. (tDAR charges $5 or $10 for users to upload a file themselves, depending on the number of files. Additional charges are incurred if tDAR's staff upload the files and add metadata.)

David Clarke, the federal preservation officer for the Federal Highway Administration, agreed with Kintigh. His agency funds federal highway projects, including any archaeological work associated with these projects. The curation of archaeological collections, both digital and physical, is done by the states in which the excavations take place, but it's Clarke's responsibility to make sure that federal laws such as 36 C.R.F. 79 are observed. "Most states have a curation fee that the Federal Highway Administration builds into the cost," he said. There should also be a similar fee for digital

Preventing Data From Disappearing

The Air Force, which manages more than 23,700 archaeological sites across the country, has been uploading its legacy data to internal databases and to tDAR. Alison Rubio, who helps oversee the Air Force's cultural heritage collections, said that one great advantage of using tDAR is that its staff will preserve the material forever. As long as tDAR exists, it will maintain its records, transferring them as old digital materials—like floppy disks—expire and new ones take their place. "That's where we see a big benefit," said Rubio. She also noted that the tDAR database redacts the locations of sensitive archaeological sites, so that the Air Force can keep the information secure.

That ability to redact sensitive data also appeals to Daron Duke, an archaeologist with the Far Western Anthropological Research Group in Henderson, Nevada. He has uploaded information to tDAR from several projects his team has done for federal agencies. "As a scientist, I feel obligated to preserve and disseminate the data that I produce," he said. "Our industry produces massive amounts of data that are generated for the public good." By keeping site locations secret, tDAR allows researchers and the public to learn about archaeological findings.

tDAR is not the only public archaeological database. Universities and libraries often host databases that are specific to one project or one area, such as the Chaco Research Archive at the University of Virginia that specializes in findings from Chaco Canyon in New Mexico. And the University of York, in England, hosts the Archaeology Data Service, which has accessible archives of many publicly funded archaeological projects in the United Kingdom. Curators at the UK system and at tDAR work closely together. They have co-published guidelines for good practice in curating digital data, such as how to choose the best formats for saving digital photographs and how to decide what to include in a project archive. Each is probably the closest thing to a public national archive of the data gathered during cultural-resource investigations, according to tDAR's Frank McManamon.

tDAR uses equipment like this book scanner to digitize information.

tDAR also collaborates with Open Context, a site that publishes archaeological research data and is more of a publishing platform than a repository. Open Context has posted records of 1.5 million items from 115 projects so far. Open Context's approach allows different types of science to be done, said its co-founder Eric Kansa, such as displaying and comparing artifacts from many different regions and many different time periods. Open Context is also home to the Digital Index of North American Archaeology, which collects records mostly from state historic preservation offices. In 2017, scientists used the index to assess how rising sea levels are threatening more than 13,000 archaeological sites across the American Southeast. Some of those sites are also indexed in tDAR, so Open Context and tDAR link to one another at relevant points in their databases.

Easy access "to digital archaeological information can be an incredible efficiency for new research and agency compliance efforts," said Emily Palus of the BLM, who added that her agency uploads some of its data to tDAR. Kintigh echoed Palus, noting that an abundance of accessible data allows researchers to explore broad questions that span multiple sites. For instance, he and his colleagues recently used tDAR and other databases to analyze the occurrence of animal bones at archaeological sites across the American Southwest. The scientists analyzed more than 300,000 individually recorded specimens from thirty-three separate sites dating between A.D. 1150 and 1500. Being able to integrate the information from multiple databases among tDAR's records allows researchers to explore questions such as the conditions under which persistent human presence on the landscape affected the abundance of large game animals such as deer and antelope.

While Danielle Benden believes that archaeologists should deposit their data in a safe and long-term environment, she asked, "what happens if tDAR goes away some day? Then what?" McManamon said that Digital Antiquity and tDAR are not about to disappear. The repository is focused on providing long-term access to, and preservation of, its data for future uses. It also has an agreement with the ASU Libraries' institutional repository to curate tDAR content if necessary. User fees and other income now make up more than half of its annual operating budget, so it is becoming less reliant on grants. —*Alexandra Witze*

CENTER FOR DIGITAL ANTIQUITY

COURTESY OF ARIZONA DEPARTMENT OF TRANSPORTATION

The prehistoric site Las Capas was excavated in advance of the expansion of Interstate 10 in Tucson, Arizona, between 2009 and 2013. The project was funded by the Federal Highway Administration, which is responsible for making sure that federal curation laws are observed. However, not all states comply with the curation laws, according to an FHA official.

data curation that his agency would pay for, but he's not aware of any states charging such a fee.

Clarke described the states' compliance with federal curation laws as "a mixed bag," noting that "there are some places that are not in compliance." The states often prioritize the physical collections, offering the excuse that they don't have "the energy, effort, or time" to properly curate their digital data. "It's a weak argument and a copout," said Clarke, "but I think it's the reality." Though Clarke has the responsibility of overseeing the states' compliance with federal curation laws, he doesn't have the power to enforce compliance.

Emily Palus of the Bureau of Land Management (BLM) believes the situation is more complicated than Kintigh would have it. "I don't know if it's as simple as moving money from one thing to another thing," she said. Palus, BLM's acting division chief for Cultural, Paleontological Resources and Tribal Consultation, noted that most of the archaeological projects on her agency's lands are not funded by the BLM. This is because the archaeological work results from projects such as energy exploration and extraction that are conducted by third parties, and they have to pay for the archaeology. The BLM issues permits for the work and oversees how it's conducted, but it has no say as to the amount of money spent or how it is spent. Palus said the money budgeted for archaeological projects is often insufficient to cover their actual costs, which include digital curation.

Kintigh acknowledged that curating the huge backlog of so-called legacy data from past projects requires that officials find additional funds, which can be challenging. The Veterans Curation Program, which entails the curation of both digital and physical legacy data, costs about $6 million annually and is paid for by the Corps of Engineers. (Trimble was unable to say how much of that money is devoted exclusively to digital data curation.) When federal "agencies can find spare money" they take proper care of their legacy data, Trimble said, but he added that these agencies have to contend with numerous other mandates and they often lack the money to fulfiill them.

McManamon believes that it may take a high-profile legal case to force agency managers to start properly curating their data. "In discussions, archaeologists consistently recognize the importance and value of digital data curation," he said. "However, they are just beginning to understand what good digital curation entails and to find ways of incorporating it into their projects. They had better speed up this aspect of their research or they will find that the data they have created has disappeared."

ALEXANDRA WITZE *is a science journalist in Boulder, Colorado. She is a frequent contributor to* American Archaeology.