Title
ArchaeoSTOR: The Development and Utilization of a Web-Based Database for the Field and Lab

Journal
Near Eastern Archaeology, 77(3)

ISSN
1094-2076

Authors
Gidding, Aaron
Levy, Thomas E
DeFanti, Thomas A

Publication Date
2014-09-01

DOI
10.5615/neareastarch.77.3.0198

Peer reviewed

ArchaeoSTOR:
The Development and Utilization of a Web-Based Database for the Field and Lab

Artifact storeroom organized using ArchaeoSTOR for inventory management. Photograph by A. Gidding.

Aaron Gidding, Thomas E. Levy, and Thomas A. DeFanti

The inventory of archaeological material is an old practice and basic to the method of all research and analysis. The whole purpose of finding archaeological material is to be able to keep track of its provenience and describe the cultural history of that object. Fundamentally archaeology is about things and the relationships between things and in turn how those relationships map onto social organization (Hodder 2011). As an important institution for the long-term storage of archaeological material, the museum has always been used as the ideal storage place for archaeologists. The museum is able to both show material off to the public and privately maintain a larger research collection, out of the view of the museum visitor. The curation of material in museums is meant to help guide the interpreter to understand the relationship between artifacts and come to some kind of synthetic conclusion regarding the material. By creating a way to understand the relationship between different kinds of artifacts, meaning is produced but, in a semi-prescribed way, the visitor is able to draw their own connections between the artifacts as they choose. Behind the presentation of artifacts are thousands more artifacts, stored and organized for later study or analysis.

Like many other excavation projects, the UC San Diego Edom Lowlands Regional Archaeology Project (ELRAP) seeks to function similarly to a museum. Every excavation season ELRAP researchers excavate tens of thousands of artifacts that all need curation. For ELRAP, nearly every artifact excavated is also sent back to the UCSD Levantine and Cyber-Archaeology Laboratory for long-term storage and analysis. Since 1998, when ELRAP "went digital," every artifact has been collected with geo-spatial metadata, Over the years, new digital technologies have been adopted and adapted to the digital data collection workflow (fig. 1). However over the course of many seasons the amount of material that formed the collection was becoming difficult to manage. More than a decade of research led to an influx of material. To add to the complication of storing the physical material, the organization of the digital metadata that was necessary to help draw out the relationships between these excavated things grew increasingly complex as new technologies were implemented. Given the increased mandate that is made by many large funding agencies regarding data sharing, such a proposition became especially difficult, especially given our project's desire to make a useful interface for public access to our data.

In response to the various pressures outlined above, long term storage of both physical and digital material and the need to publish that data in a curated fashion, the web application ArchaeoSTOR was developed (Gidding et al. 2013). It was designed to meet primarily the needs of the ELRAP by inventorying any item excavated with the ability to make easier references between sites, time-periods, and years of excavation. As such, ArchaeoSTOR facilitates the management of both the physical and the digital data from excavation season to season. The metadata required to describe that management is stored in a PostgreSQL database that can be shared either as the raw data itself or through a number of functions within the web application that provide the user with other ways to interact with the data. In this sense, the stated goal of ArchaeoSTOR was to produce not only a way to manage the data, but also a means to share that data in a museum-like way, so that the visitor can understand the direction of our interpretations, but also make their own.

Presentation of digital data is more meaningful in the current research landscape because of the increased resolution and power that analysis of archaeological artifacts has using modern technologies. The digital data that is associated with the artifacts occupies the space as its own kind of thing within the model of a relationship of things. The metadata for artifacts is varied into a number of different kinds of categories. The natural description that has existed throughout the history of archaeology is the spatial relationship between artifacts, now represented in digital forms. Other kinds of description that have always been around include qualitative analysis used by the excavator to leverage their personal knowledge of the material record. New forms of data that also can be embedded include photographic representations, three-dimensional representations, and data from digital chemistry. These newer forms of digital data allow for characterization in a unprejudiced way, preserving a more reproducible qualification of the nature of a given artifact.

The Description of Things

Unlike in the museum, the current desire is not for simply a prescribed route to understand the relationship between archaeological material, but also the extensibility to make new connections and develop relationships that otherwise might not have been observed. Describing archaeological material has historically been a difficult task due to the wide number of ways that a given artifact could be interpreted. This was one place where the ArchaeoSTOR project saw an opportunity to leverage new digital technologies as a backbone for how to model the management of data (figs. 1–2). This means that within the database there is a set of explicit relationships drawn out between different data hierarchies that maintains multi-scalar affiliations between

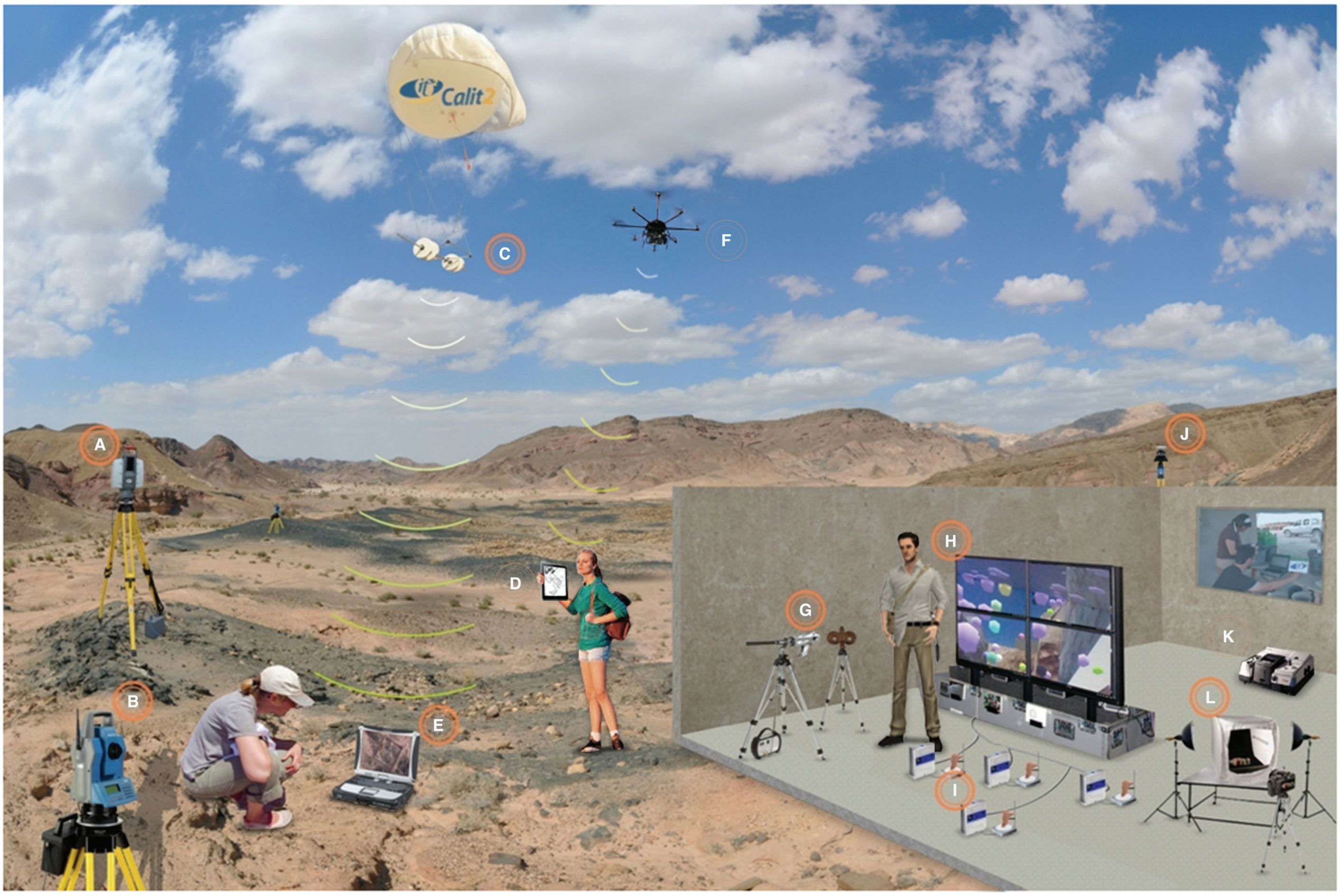
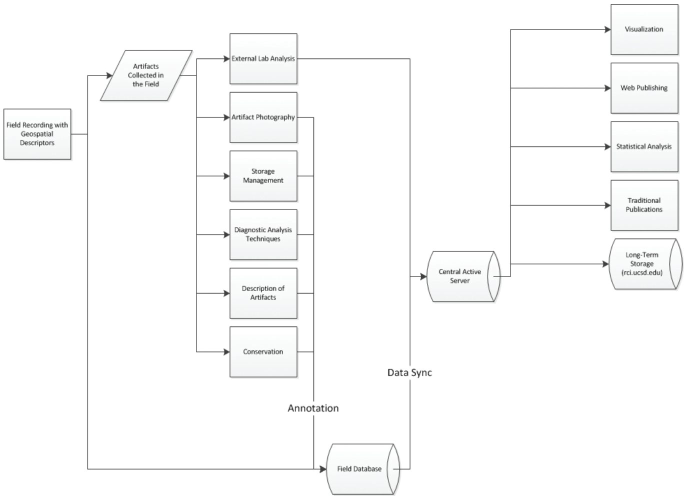


Figure 1. Graphic model of the various components of the ELRAP field recording methodology. A. LiDAR Scanner B. Total Station C. Aerial photography platform D. Digital note taking (OpenDig) E. Control for aerial photography platform F. Octocopter forensic imaging platform G. XRF H. Nexcave portable display I. Nextengine 3D scanners J. dGPS K. FTiR L. Digital object photography. Figure made by T. E. Levy and S. Blair, courtesy of UC San Diego Qualcomm Institute.



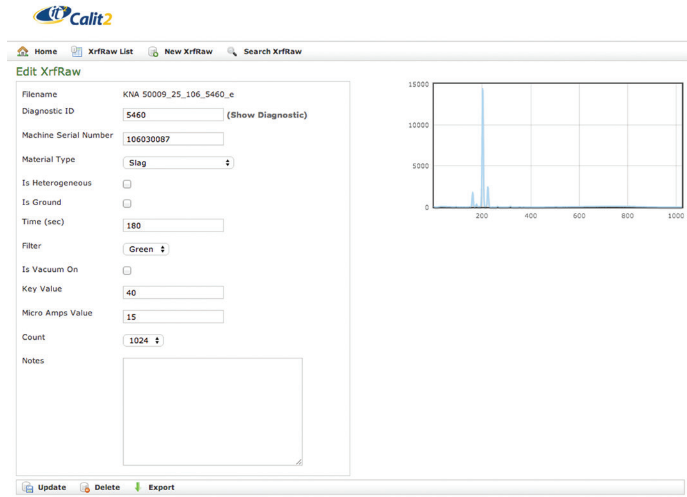


Figure 2 (top left). A simplified data flowchart illustrating the means by which different kinds of metadata is added to artifact description. Figure made by A. Gidding, courtesy of UCSD Levantine and Cyber-Archaeology Laboratory.

Figure 3 (top right). The XRF data interface in ArchaeoSTOR. The resulting curve from an XRF scan illustrated on screen for easy reference to other shots taken from the same or different samples. Figure generated by A. Gidding, courtesy of UCSD Levantine and Cyber-Archaeology Laboratory.

different levels of data. In the end these means of making connections within the database between all of the excavated material allows for the establishment of interpreted difference between assemblages that make up archaeological inference.

Digital chemistry is one of the most exciting new sources of data for archaeological fieldwork (Weiner 2010). Not only does data from techniques like pXRF (Portable X-Ray Florescence) and Fourier Transform Infrared Spectroscopy provide relevant data to archaeological excavations that otherwise would be difficult to come by, it also provides a means to collect comparable raw data (fig. 3). The comparability of that raw data, however, requires a certain amount of metadata that allows the data analyst to be able to normalize the datasets and ensure proper data comparability. It is providing this scalar metadata that ArchaeoSTOR excels.

As an integrated database ArchaeoSTOR also includes the recording of macrographic features, the traditional means of collecting data. The way that the micrographic data is stored is as a subset of other descriptions of the macrographic features. As a result the database records multiple levels of descriptions for the individual artifacts from which the micro-analysis derives, but also the more general assemblage that is necessary to contextualize other data types.

When considering the scope of the ELRAP, to understand the social evolution surrounding metal production in the lowlands of Edom, the process of normalizing data across the years of data collection can be difficult. It is not enough to note the machine that produced the data and methods of analysis, both important factors. Also important are the circumstances of collection and the associated materials. What this has meant for ArchaeoSTOR is creating an interface where the integration of data from multiple data streams (fig. 1) could be used in conjunction in order to make dynamic models that include multiple varieties of data (fig. 4). Simply put, the idea here is allowing researchers ready access to the full corpus of data generated as a part of archaeological fieldwork in the effort of generating more complex analyses, ideally analyses that transcend sub-disciplinary boundaries within archaeology.

The Location of Things

We began by highlighting the importance of understanding the relationship between things, but it important to note that for the archaeologist fundamental to unraveling that relationship between things is the space between them. So far, that space has been discussed in terms of understanding the difference between assemblages through the descriptions of artifacts through a number of means. However, the control of space is one of the best ways that archaeologists have to both analyze and organize their data. The importance of space is not a revelatory assertion, but the use of space as the means to organize all data is not necessarily universally applied for the entire data structure by all projects. Collecting spatial data and using it for organization has been central to the ELRAP for more than a decade now, meaning that the transition to using that data for the purpose of ArchaeoSTOR has been fairly easy (Levy et al. 2010).

The implications of using space as a means to manage all data relationships allow for the relationships between artifacts to be easily drawn out and represented. Even without an integrated database it would be possible through popular GIS software (ArcGIS or QGIS) to tie together data from an excavation in order to draw out spatial relationships. However, one of the advantages that using an integrated database offers is a lighter weight interface that provides a means for front end that can return spatial queries without external software (fig. 5). This is useful for a number of applications that go beyond standard data analysis.

Part of every excavation is recording the circumstances of excavation, but the recall of that data with everything else that was excavated is not straightforward. The problem being that all of the data that is collected about what was excavated with details of that material is not necessarily stored in the same place, and as that data is analyzed by various specialists within the excavation each working within their metaphorical silo. A multi-user, integrated database like ArchaeoSTOR facilitates the bringing together of all of these components and this is to the benefit of

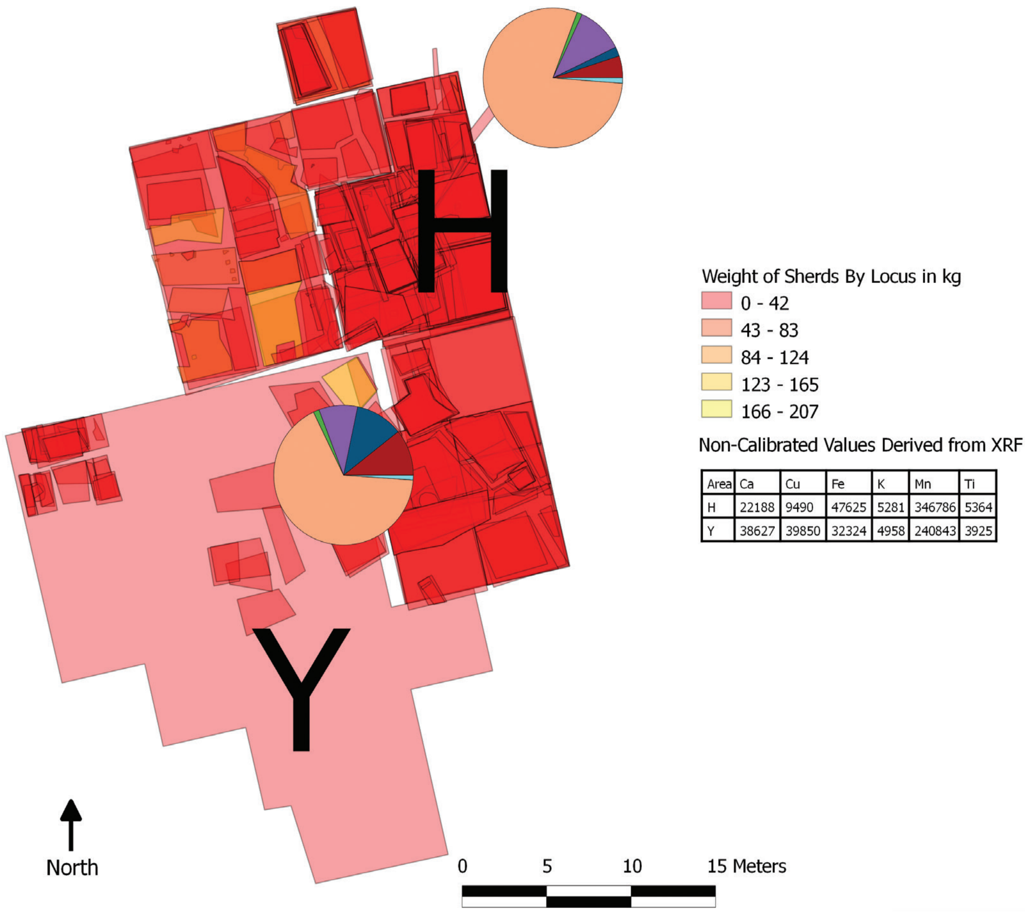
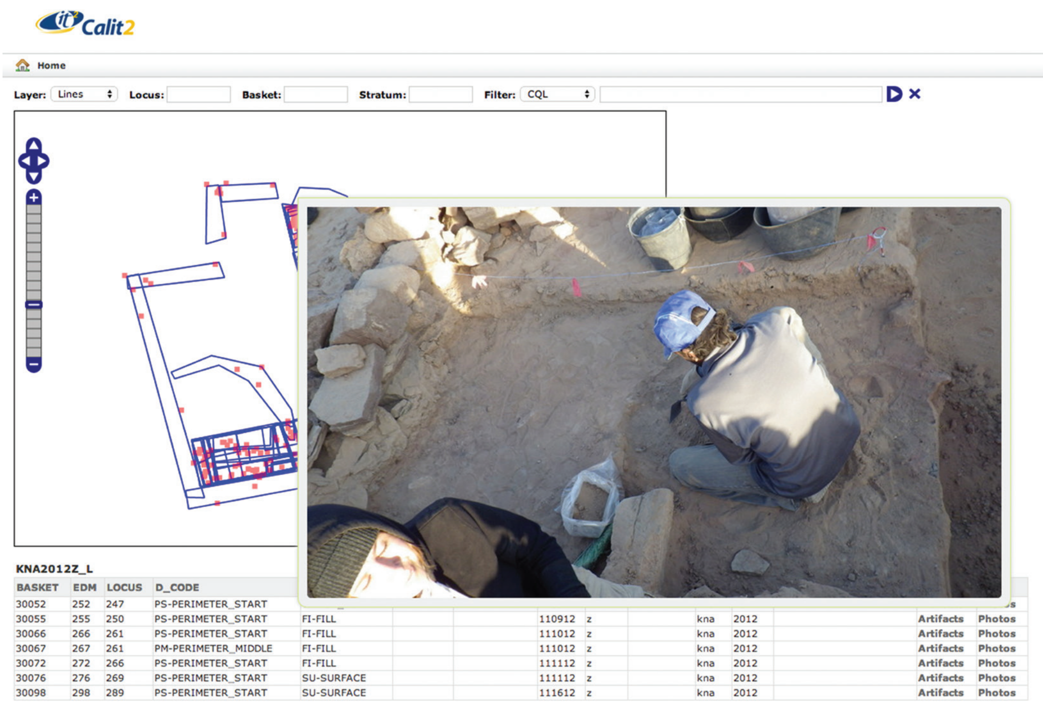


Figure 4 (above). Map of pottery and pXRF data layers visualized in a QGIS together to show variations in data between excavation areas.
Figure 5 (below). Illustration of the spatial query system. The view presented is after clicking on a spatial unit. At the bottom are all of the artifacts associated with that spatial unit, with a way to navigate to entries. To the right is an image of the excavation unit selected when it was first opened, with the ability to cycle through all images taken of the unit during excavation. The central display is the representation of all of the polygons and points excavated. All of this data is of course stored in the database. Figures generated by A. Gidding. Photograph within by I. Jones, courtesy of UCSD Levantine and Cyber-Archaeology Laboratory.



the supervisors of the excavations. As seen in figure 5, in addition to the various kinds of data describing artifacts or things that are collected, images from the excavation are also displayed allowing the supervisor to actively explore the data without the use of an external program, placing that data within the excavation context for easy recall of what happened earlier in the excavation process. Additionally, this allows for the excavator to pull all of the relevant data about the artifacts from the excavated contexts and make use of it for reports and other publications, bringing together the data of things with their locations (fig. 6).

Finally, within the context of curation the location of things takes on extra importance in a literal sense. Because excavations generate thousands of finds, each of which require a variety of analyses, tracking the physical location of all artifacts is important. The tracking of artifacts naturally happens within ArchaeoSTOR within the context of how the data is stored. Every level of artifact has its own unique ID that can be tracked within the system and referenced during analysis. This means that the long term curation work, the work of maintaining the artifact beyond the use-life as an analyzed object in a museum is built into the basic functions of the system. This aim is built into the core operations of ArchaeoSTOR which helps the web application work well both in the field and lab settings for ELRAP.

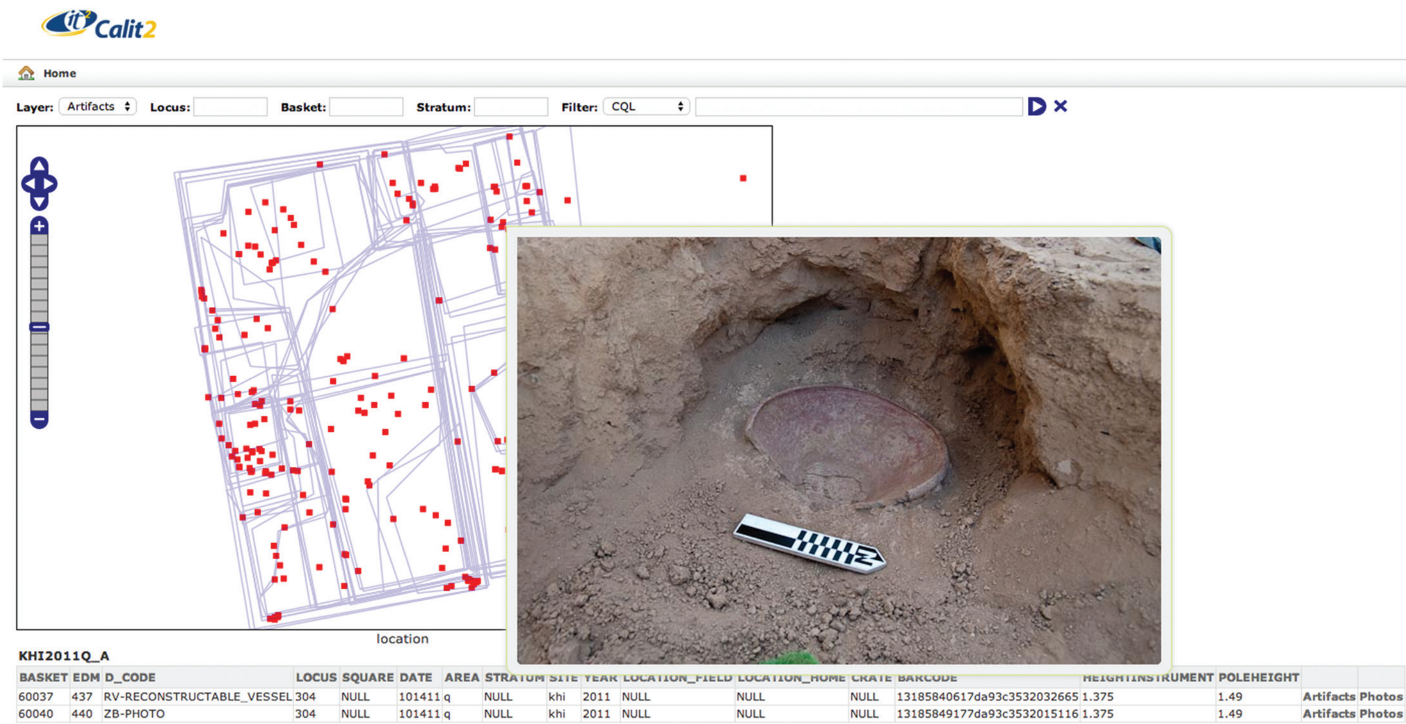


Figure 6. Further illustration of the spatial query system, this time highlighting the selection of an artifact found *in situ*. Figure generated by A. Gidding. Photograph within by A. Gidding, courtesy of UCSD Levantine and Cyber-Archaeology Laboratory.

ArchaeoSTOR as a Tool for Data Exploration

The model that the ArchaeoSTOR project looks to exemplify is one in which the onus is on the data scientist to work as an observer and active participant in data exploration. As in the museum, the curator can help guide interpretation and analysis, but the data is there for different means of differentiation to be evident. The interlocutor is able to make their own decisions about interpretation, albeit with the understanding that the data was structured in a way by the data creator. This is where all of the described metadata annotation helps with annotation and ideally allows future researchers to move analysis of the data forward in novel ways in the model of other e-Science applications.

When taken as a whole ArchaeoSTOR provides for the integration of many disparate data varieties integrating the discussed aspects of artifact description and location. Figures 3, 4, and 5 highlight the array of data that can be utilized in tandem in order to describe different kinds of entanglements in archaeology. Although the figures are from different excavations, the ability to take what is not superficially obvious in the artifact description or XRF (fig. 3) and combine it with macroscopic data from the site in figures 4 and 5 means that the relationship between what is observed in laboratory and field settings can be more easily connected. An example of this from our research in the Edom Lowlands can be taken from more recent excavations at the site of Khirbat Faynan where ArchaeoSTOR was used over multiple excavation seasons to inventory material and then highlight statistical differences in excavation areas that impacted the interpretation of periodization and the formation processes at the site (Levy et al. forthcoming).

The utilization of ArchaeoSTOR for ELRAP is ultimately about facilitating the ease of movement of data between various use case scenarios and that data includes the physical objects themselves. Leveraging digital techniques cannot just be about new visualization and metadata, but also using them to benefit the management of the physical material is important as well. It is through the museum analogy as a guiding principle that ArchaeoSTOR is able to fulfill that obligation. Not only is the data available for cutting edge research in the digital realm, allowing data scientist to make what use they may of the large variety of data produced, but the physical material is also well taken care of in case the digital methods of analysis are not enough.

References

Gidding, A., Y. Matsui, T. E. Levy, T. DeFanti, and F. Kuester. 2013. ArchaeoSTOR: A Data Curation System for Research on the Archeological Frontier. *Future Generation Computer Systems* 29(8): 2117–27.

Hodder, I. 2011. Human-Thing Entanglement: Towards an Integrated Archaeological Perspective. *Journal of the Royal Anthropological Institute* 17(1): 154–77.

Levy, T. E., M. Najjar, Bennallack, K., I.W.N. Jones, A. Gidding, R. Shafiq, M. Vincent, M. Howland, and A Shoemaker Forthcoming 2012 Excavations at Khirbat Faynan: Preliminary Report. *Annual of the Department of Antiquities of Jordan.*

Levy, T. E., V. Petrovic, T. Wypych, A. Gidding, K. Knabb, D. Hernandez, et al. 2010. On-Site Digital Archaeology 3.0 and Cyber-Archaeology. Pp. 135–53 in *Cyber-Archaeology*, ed. M. Forte. Oxford: Archaeopress.

Weiner, S. 2010. *Microarchaeology*. Cambridge: Cambridge University Press.