From the Site to Long-term Preservation: A Reflexive System to Manage and Archive Digital Archaeological Data

Maria Esteva, Jessica Trelogan, Adam Rabinowitz, David Walling, Stephen Pipkin; University of Texas at Austin; Austin, TX, US

Abstract

A collaboration between the Texas Advanced Computing Center and the Institute of Classical Archaeology at the University of Texas at Austin has resulted in the development of strategies and a cyberinfrastructure model for the management and preservation of 1) primary archaeological data, 2) the process history generated as those data are analyzed, manipulated, and interpreted, and 3) the final interpretive results. This project draws on the concept of reflexive archaeology to map the lifecycle of archaeological research, and the evolving archive that such research creates, to provide a richer view of both data and process that will make it easier for the data to be reused and to evaluate the interpretations.

Introduction

Current archaeological research increasingly relies on digital tools to produce more and richer documentation. These tools also allow documentation to be shared in dynamic, contextual forms. At the same time, digital solutions have created a new set of serious problems. Without sufficient metadata and a sustainable long-term preservation strategy, original digital objects risk becoming indecipherable, which places at risk the irreplaceable evidence produced in the course of the fundamentally destructive process of excavation. Furthermore, since digital objects can easily be renamed, edited or transformed, the connection between original data, repurposed versions, and final interpretations is easily lost. Without access to the original evidence on which interpretations are based, it becomes difficult to question or revise those interpretations. Recognizing this, the Institute of Classical Archaeology (ICA), a research unit at the University of Texas at Austin (UTA), became concerned about the data it has generated during fieldwork over several decades. Many of ICA's recent projects have produced extensive and complex digital data which have tended to be managed haphazardly, with little metadata and little systematic organization.

An ongoing collaboration between ICA and the Texas Advanced Computing Center (TACC), a center at UTA that provides advanced computing resources for research development, has resulted in the design of a cyberinfrastructure model for the management and preservation of 1) primary archaeological data, 2) the process history of those data as they are analyzed and manipulated for publication, and 3) the final interpretive results. Various existing archaeological data-archiving initiatives [1, 2] treat archaeological data as static facts, either in their "raw" form as originally collected or in a final "clean" form as eventually published. The ICA-TACC model draws on the concept of "reflexive archaeology" to map the lifecycle of archaeological documentation. It considers the dynamic interplay between documentation, study, and interpretation, thus providing a richer view of both data and process. By preserving process as well as data, the project seeks to make it easier for the data to be reused and the interpretations evaluated.

The Nature of Digital Archaeological Documentation

Traditionally, the production of archaeological knowledge has involved two stages. In the first stage, research is carried out through survey or excavation, and primary data (about artifacts, structures, stratigraphic layers, biological remains) are recorded, first in their spatial context and in relation to other material, and then as individual objects of inquiry. Between field seasons and after fieldwork is complete, the primary data are analyzed by specialists and eventually synthesized to produce a diachronic interpretation of a site or landscape, which usually serves as the basis of a final publication. While most archaeologists recognize their responsibility to produce and preserve accurate documentation of the primary data, the fluid stage of analysis and selection that lies between the data and the final publication is rarely preserved or presented for public consumption.

In excavations that use a single-context recording system, documentation strategies usually center on the stratigraphic unit or "context"—that is, the traces of a single action or event, as they are preserved in the soil. Each context is associated with all the material found within it; this material is typically separated into "bulk finds" (objects that occur in large quantities and that are more informative in the aggregate than as individual items, such as ceramics or animal bones) and "special finds" (objects that are infrequent and especially informative as individual items, such as coins or metal utensils). Each context is assigned a unique code, and its physical properties are documented with photographs, sketches, measured drawings, descriptive text, and spatial coordinates. Both bulk and special finds are associated with this context code during collection. In the South Region at Chersonesos (our case study project), special finds were also assigned their own unique codes, noted in a register, and located spatially using a laser theodolite (Total Station). They were then passed to a project registrar, who provided a written description and took measurements before sending them to the conservation laboratory, where they were photographed before, during, and after cleaning or conservation. Finally, a subset of the most important and interesting special finds was accessioned into the collections of the Chersonesos Museum.

The information collected about contexts in the field is typically the only record of a given stratigraphic event, since these must be removed—and thus destroyed—in the course of excavation. The finds, on the other hand, retain an independent physical existence, but their meaning depends on their association with a stratigraphic context and with the other objects from that context. If information about those relationships is lost, it becomes impossible to interpret the material or the excavation. The situation is further complicated as objects move from field to storeroom. Documentation strategies rarely reflect the decision-making processes by which certain special finds are selected as worthy of unique identification and description, and post-excavation processes such as conservation and museum storage are not always included in an excavation's documentation pipeline. In fact, every site presents a set of unique circumstances and contextual elements that play a role in the linearity and completeness of its documentation strategies. The study of excavated material presents an even more complex situation. It is very difficult to capture processes of analysis and interpretation that take place after each field season, often continue for several years after field research, produce documentation in diverse media, and involve specialists working in several different countries.

Until recently, most primary data were collected by archaeologists on physical media. Apart from the artifacts themselves, documentation took the form of paper recording sheets or notebooks; photographic negatives or positives; and plans, sections, and object drawings in pencil or ink on paper or mylar. In the last two decades, however, much of that documentation has been supplemented with or replaced by digital files. While some projects still prefer to begin with a "hard copy" stage involving paper contextual recording sheets and hand-drawn plans, these forms of documentation too are usually digitized at a later stage, in order to integrate them with the "born-digital" components now included in most recording systems. The digital files replicate earlier documentation (e.g. in the case of digital photos) or significantly enrich it (e.g. in the case of relational databases, GIS, and 3D models). While a paper archive could be consigned and stored indefinitely, digital data are more vulnerable to the passage of time and there is a significant risk of loss. At the same time, unlike paper records, the components of these datasets can be manipulated and transformed without leaving clear traces of their transformation. As a result, the endless possibilities for repurposing and duplication afforded by digital data are also responsible for a deluge of archaeological information that is becoming more and more difficult to manage and access.

The Chersonesos South Region Data Collection

ICA has been conducting excavation, conservation, and research projects at Chersonesos since 1994. Each of these projects has produced a collection of inter-related sets of documentation, much of which is now "born digital". The data-collection methods employed in ICA's field research have themselves evolved extremely rapidly, even throughout the course of a single excavation project, as software, hardware, and the skills of researchers have changed.

To meet the needs of faculty and researchers for data collection services, and to contribute to the potential of data-driven research to make discoveries, in late 2008 TACC established a group dedicated to the management of UT data collections. This group builds and maintains large data-management and storage resources, and consults with collections' creators in all aspects of the data lifecycle, from creation to long-term preservation and access. The group members are specialized in database management, GIS, scientific data formats, metadata, storage architecture, archiving, and long-term digital preservation.

Archaeological Workflow

The first step of the TACC-ICA collaboration was to study the documentation workflow and evaluate, as a pilot project, the Chersonesos South Region data collection produced in the course of excavations conducted between 2001 and 2006. We identified the following active workflows feeding digital data into the collection: 1) field collection: original/raw digital photographs, spatial data related to objects and contexts collected with Total Stations, remotely-sensed landscape data, and scanned plans and sketches; 2) data processing during and after the field season: the data entry of textual records into a SQL database, the extraction and management of spatial and graphic information in a GIS, and the creation of 3D models using photogrammetry; 3) off-season studies/analysis: specialist reports, statistical and geospatial analysis, materials study, data cleaning, the creation of additional 3D models; and 4) creation of resources for publication and presentation: digital materials manipulated or produced (e.g. photographs cropped and color balanced, illustrations inked and digitized, formal maps created from spatial data) for public dissemination. The digitization of various paper records, including hand-drawn plans, sketches and illustrations, field recording forms, and field notebooks, is a component of all four of these workflows.

Fieldwork

At the site, observational data of many kinds are collected at different stages during excavation and conservation. We call this "original/raw data," and different pipelines exist for the different documentation types (e.g. image documentation, spatial, documentation, and text documentation, etc.). Data objects, both born-digital and digitized, are renamed according to a basic naming convention and are stored in directories in a portable server that is carried back and forth to the field. Images and their associated metadata, such as photographer, date, and subject, are entered daily into the Archaeological Recording Kit (ARK) database, which is also hosted on the portable server and used to manage image and contextual data [3]. While the data are captured effectively during the excavation, differing practices between the excavation and conservation teams make it difficult to trace the documentation of special or bulk finds through the conservation and museum accessioning processes.

Data Processing

Both during and after the field season a considerable amount of data processing occurs. Data entry of recording sheets and field notebooks is done daily, as is the scanning, georeferencing, and digitization (creation of vector layers) of hand-drawn plans within the GIS. 3D models are extracted from photographs and from Total Station data. This stage of the data-processing work-flow is, in fact, where the bulk of the work is done to generate the spatial data that is housed within the GIS and the contextual information stored in the database.

Off-season

After a given field season is over, the portable server is brought back to ICA in Austin, plugged into the network and accessed via a webserver. At this time, research continues and a large, and dispersed team of specialists in different continents and studying different materials accesses the original raw data. In the course of study, digital objects may be repurposed on the personal computers of various participants and new data are created. Specialists studying and conserving materials in Ukraine produce artifact drawings and photographs, reports and descriptive tables. Additional resources are also often digitized from older excavations, the surrounding landscape, or comparable sites and, although not part of the "original" dataset, are an integral part of understanding the site in context. These new objects are not always re-integrated into the server or standardized in terms of naming conventions.

Publications

Presentations, annual reports, and specialized journal articles are produced throughout the life of the excavation, generating in the process a new batch of data derived from original materials. The publication workflow takes place across different storage devices and ICA's institutional server. In this case, considerable storage resources are needed to host high quality images. In addition, ARK will constitute a web-publication interface that researchers and the public will be able to browse and query to find additional information on objects and contextual relationships, and frame new questions about the excavated material.

When the final publication of the excavation project, including catalogues and final interpretations, is completed, the collection ceases to develop and will be considered static. Additional data can still be derived from the static collection in future study or as new excavation areas are opened in surrounding or related areas, but they would be considered part of a separate dataset.

The Collection Assessment

At the start of the TACC-ICA collaboration, since the files were distributed among various storage devices and only some staff members knew their functions, the assessment was a manual process. Sketches of the workflows and a thorough inventory allowed us to determine the scope and variety of digital objects in the collection. Assessments were made of: their location across storage devices, whether they were copies or versions of raw data, whether they existed in ARK, and the existence of corresponding metadata within ARK or elsewhere.

Of roughly 2 terabytes of data, almost half corresponds to repeated copies/back up files or original data distributed amongst various portable storage devices. The disconnect between the workflows of ICA staff members and those of their off-site collaborators was apparent: the contents in the storage devices were often not appropriately or consistently labeled or indexed, most digital objects outside of ARK were not associated with any metadata, and the relationship between them was impossible to trace. In the past, attempts to implement a record-keeping system for specialists and staff members failed because the system was considered cumbersome and inconvenient.

The security of the systems that host the data was also evaluated. We learned that once in Austin the portable server was not automatically backed up, and that ICA staff had neither control over their institutional server nor the support necessary to conduct more serious data analysis activities including multi-user GIS and database maintenance. In sum, the physical size and the dispersed nature of the collection was becoming unmanageable.

The situation observed at ICA is a typical case of digital data collections maintenance. Similar problems are found in other types of data collections, such as biological or image databases, that grow in size and complexity to the uneven rhythm of research grants, constant technological advances, and fluctuations in the availability and interest of specialized staff. Yet these collections form the basic units of analysis for the research they reflect, and as they grow more disorganized, they begin to stall, complicate, and interrupt the research they should be generating.

Upon the completion of the assessment, we defined the concept of an evolving archive and devised a "reflexive" management and archiving system that would integrate the different collection stages by allowing the archiving of data at all the different stages of the archaeological research process.

An Evolving Archive

Over the last sixteen years of work at Chersonesos, original data from several expeditions have been transformed for analysis and publication at the same time that new evidence was being gathered and incorporated into subsequent cycles of analysis. These juxtaposed processes have created an "evolving" collection in which the relationships between the data objects are blurry.

The challenge resides in managing the data so that the distinctions between the different archaeological documentation processes and their stages, between the different workflows, and consequently between original and repurposed data are transparent. In this way, while the evolving collection will be considered "static" when the final publication of a given project is issued, the archive that it leaves will reflect the evolving aspect of the archaeological research process.

In recent years, two basic models have emerged in archaeological data preservation: 1) the centralized model, in which a single repository handles data after they are finalized and consigned in standard, archival formats; and 2) the decentralized model, in which individual projects curate their own data however they can. Neither of these models is particularly well-suited to document the transformations digital data undergo between collection in the field and their final publication. The centralized repository model works well as a final clearinghouse for static data at the end of a given project, but cannot address the needs of on-going research with an evolving digital archive. The decentralized model may be better-equipped to deal with the idiosyncratic, "messy" middle-ground of evolving collections, but, as this project demonstrates, data organization across the research project is often neglected because the data management strategies are too burdensome for the archaeologists. This limitation leads to delays and confusion in research as a result of the "mess", and affects the quality of the final presentation of the data. A new model is clearly necessary.

Archaeological theorists have long acknowledged that the connection between data and interpretation is rarely linear: instead, the two are tied together by processes of filtering and inference that are specific to any given project or group of collaborators. Traditional models for archiving and publication pay little attention to the documentation of these filters and inferential chains, but they can usually at least be reconstructed from the raw materials. Digital models, on the other hand, have the potential to preserve those processes more fully or erase them altogether. In seeking to preserve inferential connections and record the transformations undergone by data between collection and publication, we have used "reflexive archaeology" as a conceptual framework to design a system that stresses data management and archiving intervention from the early stages of data generation.

Archiving and Reflexive Archaeology

The idea of "reflexive archaeology" was first put forth by Ian Hodder [4]. Hodder argued that the way in which archaeological data were collected, and the biases and choices of the archaeologists doing the collecting, had as much impact on the interpretation of remains as the "objective" data themselves. Archaeologists therefore had to recognize and make explicit the social and methodological contexts in which their data were gathered, in order for others to reuse those data effectively [5]. By mapping reflexive methods to the digital archaeological data management and archiving processes, we will make explicit the ways in which data are gathered, categorized, labeled, and repurposed to fulfill different roles and provide different functionalities, and the ways in which digital objects in the collection relate to each other. Thus, these data will be able to be reused for new studies of the site; to understand the reasoning of the original researchers; to validate or re-interpret existing research; and ultimately, we hope, to make new discoveries.

In essence, we wish to make it possible for the future user of this archive to engage in the "archaeological" investigation of the digital archaeological data, uncovering not only the completed object (the final dataset), but also the processes that connect that object with the initial collection of raw data and the context in which those processes took place. Another archaeological metaphor helps to describe what we are trying to preserve: the *chaîne opératoire*, a term used to encompass the entire range of human actions, choices and practices involved in the transformation of raw materials into a finished technological product [6]. The concept is often applied to prehistoric lithic industries, for example, by archaeologists interested not only in the flint source or the finished tool, but in the series of human actions that connected the two. Similarly, our archival framework seeks to make available not only the raw material and the final technological product, but also the paths that lead from one to the other.

Thus the system includes both reflexive and processual components. On the one hand, it records the workflows through which data were originally collected, providing documentation of the context of production. On the other hand, it documents the processes of transformation and manipulation of those data, allowing the eventual user to connect the "raw" data with the interpretations and final versions derived from them.

A Reflexive System

We are in the first stages of implementation of this "reflexive" system, which will function both as an archive and as an active data management system. Its core is the storage component, in which data are continuously deposited at the different stages of the archaeological research process until no more data will be added or modified. The integral parts of the storage system are 1) a file naming convention and 2) a hierarchical directory structure. These two components are designed to provide contextual and descriptive information about each digital object and their relationships to other objects in the collection.

While there are other data management technologies involved in the workflows—for example, the use of ARK for data management on the site—the reflexive system is the central repository for all the data types. The goal is that, independent of the platform in which they will be stored, the primary data and their metadata will be preserved for the long term, rendering transparent the evolving process from which they were generated. In the next sections we describe the system's components and how they work.

Storage Archive

The archived collection is hosted within an instance of iRODS [7] running on TACC's data-centric application resource Corral. Corral consists of 1.2 PB of online disk and a number of servers providing high-performance storage and database and web application hosting. iRODS is data grid software that allows administrators to specify 'rules' for automating tasks according to a given archiving agreement. For this data collection, these 'rules' include 1) generation of file checksums, 2) automatic replication/backup and 3) the execution of scripts. The scripts are used to: a) validate file names, b) extract Dublin Core encoded metadata from the file naming convention and hierarchy, information stored in ARK, and third party tools such as DROID [8], and b) registration of the extracted metadata into the iRods metadata catalog. Additionally, iRODS provides authentication and authorization mechanisms as well as a variety of client interfaces for managing and searching for data in the collection.

Once in iRODS, the data cannot be re-written. Command line and UI interfaces allow users to query the metadata catalog for data of interest. In this way, a digital object may be accessed from a different workflow from which it originated, repurposed for analysis and or publication, and re-entered to the system as a new object.

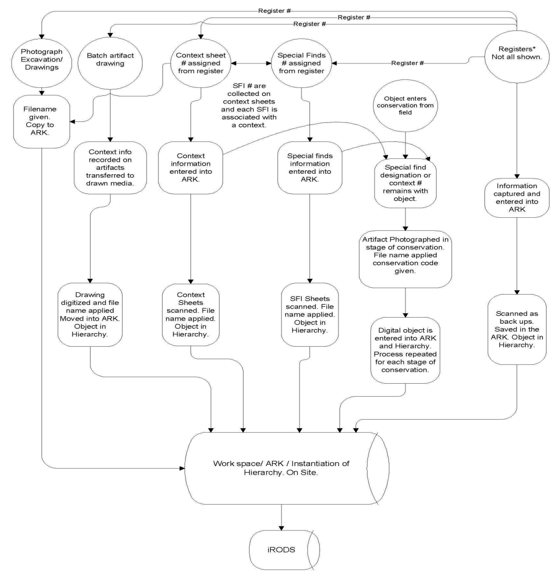

Figure 1. *Field workflow*

Reflexive Workflow

The pipelines for different types of data were studied to determine the moment in which data objects should be named and ingested to Corral during field, off-season and publication workflows. In general, it was decided that ingest should happen as close as possible to the moment at which data are created or repurposed to avoid unnecessary duplication, accumulation of data, risk of loss, and further confusion.

Figure 1 shows a diagram of the field workflow. The circles on the top of the diagram represent the types of data (born-digital and digitized) captured. Through the diagram it is possible to follow the objects as they are gathered, recorded, and entered to ARK. The workflow also shows the time at which objects should be named and ingested to iRODS. In this case, if the network capacity is available at the site, objects can be directly ingested to iRODS at the end of each working day. Otherwise, ingest should happen as soon as the server is brought back to ICA.

File-names

We revisited and improved ICA's existing file-naming convention to better describe the data objects and allow automatic parsing of metadata from filenames. The new convention describes provenance, documentation type, project and date, process (of object documented), and process stage (of digital object). Below is an example of the file-naming convention for a photograph of a special find taken at the site after the object was conserved.

[ica_ua]_[id]_[sfi_ch05sr_3065]_a9_o.tif

The code [ica_ua] indicates provenance as ICA and Ukraine. The code [id], image documentation, identifies the type of documentation (separate codes are used for tabular data, reflectance transformation image, 3D models, etc.). The middle section [sfi_ch05sr_3065], which corresponds to the item key/item value pair within the ARK database, indicates what the image is, in this case a special find; the place and date of excavation and project to which it belongs to (i.e., the site code: Chersonesos, 2005, South Region), and the register number given to that special find on the site. This object code ties individual items to related ones in ARK in a one-to-many relationship, as there may be various images of a given special find, or (as recorded on the site) the special find can be linked to another code (such as context) that describes the spatial context in which it was found. The next part of the file name [a] designates the process of the object being recorded, which in this case means "after conservation," and the [o] expresses the process history of the digital file, in this case, that it is the original image taken at the site, as opposed to an [r], for an image "repurposed" (e.g. color corrected) for publication. Because many photographs of the same object may be taken at various process stages, the numbers with the process code indicate quantity. In this case, this is the ninth picture taken of that special find after conservation.

Directory Structure

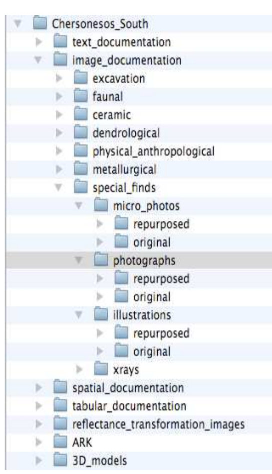

Figure 2. *Directory structure for Cherosonesos South Region data*

Figure 2 above shows a partial view of the directory structure that highlights the hierarchy created to represent the types of image documentation that are gathered about special finds in the field.The hierarchical directory structure serves to categorize the data as they are gathered and produced in the different workflows, close to the point at which they were created. It may exist within and potentially outside iRODS depending on the convenience of each workflow. So, for example, in the field, where Internet connection to the iRODS server may not be reliable, the hierarchy (which can be ported to a removable storage device) serves as a place-holder in which data objects are deposited until they can be batch ingested to iRODS utilizing available iRODS interfaces.

Top-level directories are labeled according to documentation type, some or all of which are produced during the different workflows. For each type, the sub-directories within reflect the materials subject to that type of documentation at the site, and then the different kinds of information that are collected during the further study of these materials. The last directories in the path indicate whether the data are original or repurposed for analysis and or publication.

Because ICA has accumulated five years' worth of data from the Chersonesos South Region excavations, the staff spent three months identifying, renaming and categorizing files that have already been batch ingested to iRODS. We have reported that attempts to make collaborators follow record-keeping practices during off-site and publication processes, have failed. To facilitate these procedures, we identify the precise moment in which data should be named and ingested in each workflow and will implement a web-service that includes prompts to aid file naming and data classification to the directory structure.

Metadata

In designing the scripts to automatically extract descriptive metadata, we mapped the codes of the filenames and the labels in the directory structure to selected elements in the Qualified DC standard, and created a data dictionary to specify the use of each. We pull the relationship between a special find and the context in which it was found from ARK and, if it exists, its description. Below is an example of a metadata record for a special find.

```
<dc>
    <creator>ICA</creator>
     <title>special_finds</title>
    <spatial>Chersonesos South</spatial>
    <relation>image_documentation</relation>
    <type>Photographs</type>
     <isPartOf>sfi_CH05SR_3065</isPartOf>
    <isPartOf>After conservation 9</isPartOf>
     <source>Original</source>
     <description>Lock</description>
    <description>2 elements: Element 1
    comprises 2 plates 0.7 cm. </description>
    <date>2005:08:23</date>
    <format>.JPG</format>
     <identifier>sfi_CH05SR_3065_O9_m.JPG</identifier>
     <publisher>ICA</publisher>
  </dc>
```

The DC element 'isPartOf' is used to relate the data objects in the collection, for example by belonging to the same context or being generated at a given stage of the workflow. As a group, these metadata files render the different objects, processes, and stages of the archaeological research for a particular project and represent the evolving nature of the collection.

Preservation

As mentioned, the goal of this project is to preserve the evolving archive over time independently from software platforms, so if the data and metadata need to be transferred or migrated, or if they become detached from the ARK database or separated from each other, processes that originated the data and the myriad relationships that connect objects can still be reconstructed.

At the moment, our preservation plan involves bitstream preservation via local and off-site replication, and gathering technical documentation that will allow making preservation decisions. Upon ingestion, technical metadata from each object will be extracted with DROID and, like the descriptive metadata, stored as part of the collection and registered in the iRODS metadata catalogue. All data are replicated to Ranch, TACC's long-term mass storage solution with a capacity of 10 PB, and off site to TeraGrid funded resources specially supporting large scientific data collections.

Since some information in ARK is not specific to any one object within the collection, and thus can not be stored within the extracted metadata for a given object, regular MySQL dumps from the database are ingested to iRODS and are replicated along with the rest of the collection.

Cooperation Model

The cooperation between ICA and TACC is modeled after the service concept used to support users of High Performance Computing (HPC) resources on the TeraGrid, the NSF-funded cyberinfrastructure in support of open science [9]. In this model, supercomputing centers across the nation create and maintain advanced computational resources, allocate them to research projects, and provide training and help to the scientists working in those projects. The benefit is that scientists use these complex systems without having to develop or maintain them.

In this case, we have determined the system's requirements and design, and are now in the implementation phase, one in which both teams work closely together. In the future, administratively, ICA will have access permissions to manage the different applications and to add data, while TACC will continue maintaining the overall infrastructure (servers, networks, storage), allowing researchers to work independently without having to focus on overall systems maintenance.

Conclusion

The use of a "reflexive" framework focuses attention on the process of the production of archaeological knowledge, not just the raw material and the final product. This is all the more important in a digital environment, where data are both more likely to be transformed and more threatened by the disappearance of contextual information and metadata. This solution establishes a sustainable strategy for long-term preservation, while creating more robust tools for the investigation of the data at all stages of analysis. On the level of preservation, the creation of a structure from which metadata can be automatically extracted avoids a labor-intensive, error-prone and often incomplete process of manual metadata entry. The resulting archive will preserve both the contextual relationships of the data and a record of their transformation in a standardized, human-readable format. At the same time, it will support the public presentation of the dataset in an easily-browsed format through the ARK interface. These two overlapping systems will allow the user to browse the dataset in order to develop new questions, and to search the archive in order to explore the connection between original data and interpretation. The architecture described here thus provides the flexibility necessary to preserve both the digital archaeological data themselves and a full record of the path they followed between collection and publication.

References

[1] Mc Manamon, F. and Kintigh, K., "Digital Antiquity: Transforming Archaeological Data into Knowledge," SAA Archaeological Record 10.2. March, 2010: 37-41.

[2] Richards J.D., "Preservation and re-use of digital data: the role of the Archaeology Data Service", Antiquity (York) 71:1057–1059, (1997).

[3] Eve, S. and G. Hunt. "ARK: A Developmental Framework for Archaeological Recording." Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin 2-6 April 2007.

[4] Hodder, I., "'Always Momentary, Fluid, and Flexible: Toward a Reflexive Excavation Methodology," Antiquity 71: 691- 700, (1997).

[5] Hodder, I., ed. "Towards Reflexive Method in Archaeology: The Example at Catalhoyuk," British Institute of Archaeology at Ankara Monograph No. 28. MacDonald Institute for Archaeological Research (2000).

[6] Schlanger, N., "The Chaîne Opératoire." In Archaeology: The Key Concepts, eds. P. Bahn and C. Renfrew. Routledge: 18-23, (2005).

[7] iRODS: Data Grids, Digital Libraries, Persistent Archives, and Real-time Data Systems, https://www.irods.org/index.php (April 16, 2010).

[8] DROID. Computer software. Vers. 4.0. The National Archives of the United Kingdom, PRONOM, (2006), http://droid.sourceforge.net/. (April 16, 2010).

[9] Teragrid, http://www.teragrid.org (April 16, 2010).

Authors Biography

Maria Esteva has a PhD in Information Science from the University of Texas at Austin (2008). She is a Research Associate at TACC where she is member of the Data and Collections Management and the Data Analysis and Visualization Groups.

Adam Rabinowitz is Assistant Director of ICA and Assistant Professor of Classics at The University of Texas at Austin. He holds a PhD (2004) from the University of Michigan. Since 2004, he has co-directed excavations in the South Region of Chersonesos.

Jessica Trelogan is a Research Associate at ICA specializing in GIS and remote sensing for archaeological applications. She received an MA in Classics (1994) and has been working with spatial technologies at ICA's excavation and survey projects since 1998.

David Walling has been working at TACC since earning his BBA from The University of Texas at Austin in 2004. He specializes in database applications and information management.

Stephen Pipkin is a Masters Degree Candidate in Information Studies from the University of Texas Austin.