Archaeological Data and Small Projects
A Case Study from the Pyla-*Koustopetria* Archaeological Project on Cyprus

William R. Caraher, University of North Dakota
Samuel Fee, Washington and Jefferson College
R. Scott Moore, Indiana University of Pennsylvania
David K. Pettegrew, Messiah College

Delivered at the 114th Annual Meeting of the Archaeological Institute of America.
Seattle, Washington
January 6, 2012

In recent years, there has been a growing emphasis on the curation and preservation of digital data produced by archaeological projects across the Mediterranean and the world. The model projects in this process have traditionally been large, wealthy, and able to leverage the considerable resources necessary to digitize substantial bodies of archaeological data and to develop pioneering methods to collect and analyze data with digital tools. These projects have often created elaborate, bespoke applications to collect, organize, and disseminate highly-visible, long-term digital archives. The standards established by these projects over the past 25 years have exerted significant influence over conversations regarding data formats, legacy terminologies, and ontologies in archaeology. The results of efforts by projects like the American excavations at the Agora - to name one well-known example – represents a significant example of how a large project can leverage their archival and archaeological material to produce an important digital collection. The Agora with its sister site at Corinth took advantage of the institutional support of the American School of Classical Studies under whose administrative structure both projects work. That such large projects can develop significant and substantial digital collections is hardly remarkable. The audience, resources, and institutional support will consistently produce results that are model for archaeologists working across the Mediterranean more broadly.

This paper is not about these large projects as impressive and important as they are. My paper today will focus on smaller projects. This is, in part, because smaller projects face challenges that reveal the intersection of common sets of institutional and technological constraints. This paper uses the notion of a "small project" as shorthand for projects that have limited infrastructure, small numbers of contributing scholars, finite life-spans, and typically small budgets for technology. As an archaeologist, I realize both the perils and utility of typologies, but what I'll try to consider are projects with commitments to digital approaches, but, in general, have to rely more on off-the-shelf software and have less access to the core infrastructure necessary to ensure the long term preservation and accessibility of this data. In fact, these projects tend to function in rather ad hoc or DIY ways and tend to invest relatively little in the traditional archaeological infrastructure by relying on existing infrastructure and devising projects that privilege low-impact methods which collect very few finds, reveal only small parts of sites, and generally leave little in need of curation, conservation, or long-term storage.

At present, attention to the social and technological environment of small projects is particularly appropriate in light of the changing political, economic, and methodological landscape of Mediterranean which seems less likely to sustain future long-term, large-scale projects. While the influence and legacy of existing large scale projects will persist, the future of Mediterranean archaeology appears to be smaller, shorter term, more intensive projects.

To explore some of the issues facing smaller projects (and by implication, I think, the future of archaeology), I am going to use my project, the Pyla-Koutsopetria Archaeological Project or PKAP, as an example. PKAP is a 10-year-old project focused on a multi-period site on the south coast of the island of Cyprus. The main focus of our work has been on an intensive pedestrian survey conducted from 2004-2009 at the site and three seasons of rather small scale excavations in 2009, 2010, and 2012 of excavations. The basic data from the survey and excavations were collected on paper forms. These forms were then keyed into a relational database in Microsoft Access. Spatial data from the survey and excavation was stored in ArcGIS. Paper forms were scanned to Adobe System's PDF format and images were generally produced in TIF and JPeG formats. Until very recently, we have maintained our data in commercial, proprietary software formats, and these collections will require significant post-processing to produce archival collections.

Data, Organization, and Experience

As a small project seeking to streamline data collection and analysis, we found ourselves privileging immediate utility over commitments to platform agnostic best practices. By using Microsoft and ESRI software, we limited our users to Microsoft based computers and we did very little to optimize data recording for multiple users or for future online dissemination or publication. As a result, we created a data entry bottleneck at the point where we converted the archaeological information collected in the field into digital data for analysis. It is important to emphasize that this was not a function of selecting programs incompatible with simultaneous, multiuser data entry functions, but rather that we lacked technical expertise and server-side support to make these programs work in the most efficient way.

Off-the-shelf components for data collection and analysis are not necessarily a bad choice, but they do structure the way that archaeological information is collected and disseminated. The utility of Access and ArcGIS derives from their ubiquitous distribution, functionality, and relatively easy access. At the same time, however, both programs privilege a single-user interface that structure the practice of data collection and analysis. While data collection in the field - on either a survey or excavation form - remains an dispersed endeavor including the entire field team or trench, the absence of server-side and development support for a bespoke or even modified version of standard software led data entry to fall to a single person. As a result, archaeological information moved from the collective responsibility of the project to the unique

responsibility of a single data curator who also came to have a significant influence over data structuring and interpretation. At the point of digitization, the data curator smoothed the irregularities possible on a paper form to allow them to conform to standards enforced through the fields in the digital database. This process occurred away from the field and the field teams, but we preserved the paper forms to maintain the transparency in this process.

It is worth noting that in 2012, with the assistance of Sam Fee from Washington and Jefferson College in Pennsylvania and with the support of Messiah College, we beta tested a bespoke iPad application that made it possible to collect the data traditionally recorded on a paper form directly to an iPad. Interestingly, the novelty and appeal of the iPad decentered the data collection process at the trench. Most trench supervisors took turns with students to record data in the iPad while reserving the work of recording on the more traditional paper forms to themselves. The resulting data collected on the iPads was synced nightly via a gmail account and provides us with a dataset that we can compare with the more authoritative narrative produced by the trench supervisor. The combination of a portable tool for digital data recording and at the edge of the trench moved one step closer to the convergence of the embodied process of creating archaeological data and the process of translating experience into digital records for analysis.

Creating Data

We were, of course, aware that producing archaeological data - in any form - requires careful attention to metadata that describe the nature and structure of the data. Small projects especially those that seek to explore highly specialized or limited research agendas face particular challenges when balancing the efficiency of focused data collection against the responsibilities of "best practices" and the opportunities to contribute to future or larger scale research questions. Traditionally, large projects have exerted a substantial influence over best practices and, quite naturally, set the terms for how the information collected by small projects can contribute to significant archaeological questions. In short, small projects that produce data comparable with larger bodies of archaeological information add value to their dataset and to extend the significance of the project beyond the limited research questions often at the core of smaller projects. Moreover, adopting the processes, terminologies, and technologies developed by larger project spares small projects the time and energy to develop these key aspects for themselves.

Our work at Pyla-*Koutsopetria* addressed a rather, limited set of research questions, and the project lacked any obvious and practical obligation to pre-existing data conventions (in other words, we were not beholden to earlier fieldwork at the site in anyway). To avoid producing an insular dataset, we adopted a vocabulary to describe our material and procedures that was consistent with a larger projects. We adapted our excavation techniques and our survey forms from larger existing projects as well,

namely Corinth Excavations in Greece and the Eastern Korinthia Archaeological Survey. While our adaptation of these projects' templates for data collection saved us significant time and helped us avoid recreating the wheel, in the realm of digital data sharing data structures and even ontologies pose some risks.

We described our finds using the chronotype system terminology pioneered by the Eastern Korinthia Archaeological Survey in Greece and The Sydney Cyprus Survey Project on Cyprus and adapted most recently for use in the Troodos Environmental and Archaeological Survey Project. We have since exported this terminology to the site of Polis-Chrysochous on the far western part of the island and at Athienou immediately to the north of site. While each of these projects prepared their datasets using a wide unique methods and sampling strategies, the common naming system has make it easier to compare the presence of artifact types between them and, when sampling strategies and environmental conditions allow, to compare assemblages.

Critiques of our desire for data consistency and comparability have emerged in recent efforts to compare excavations and survey datasets "side-by-side". The central issue remains how much emphasis we should place on creating and exploiting seemingly comparative datasets or whether the greater harm comes from denaturing archaeological results in the name of an illusory homogeneity. If comparative approaches are fraught with potential methodological and interpretative issues, then we should be equally hesitant to generalize data structures, terminology, or practices in the name of efficiency alone. These remain challenging and open questions that effect small projects in particular ways.

Data Preservation

Despite an awareness of these issues, adopting standardized data structures and terminology nevertheless appeals to small projects with the advantage of improving the usefulness of data for comparative study or for facilitating the aggregation of data for larger discussions of regional or transregional trends. These advantages, however, extend to more than just a perceived utility of comparable data for analysis, but also to the long-term preservation of the growing body of born digital and digitized information. Scholars have long been aware that archaeological information with broad utility is much more likely to be distributed widely, shared and, as a result, preserved; for digital data utility is all the more significant. As scholars use digital data, it gets duplicated and preserved as multiple copies and the more the community finds utility in a dataset the more likely they are to share investment in its preservation.

These practices are particularly significant for smaller projects which, like PKAP, lack the resources to invest in long-term, in-house, data archiving. For project with limited resources the key concerns are ease in data transfer, long-term data integrity, and the visibility of our data to our specialized archaeological community. While we have not solved all of our archiving problems, there is little doubt that a small project

like ours must share our data and embrace a model of distributed storage. It remains less clear how to achieve the kind of visibility and accessibility necessary to achieve this goal.

Conclusions

Small project have limited resources and distinct – if not uniform – forms of organization. These two limitations shape the way in which projects adopt and implement technology and structure their data. Recent scholarship has become increasingly attuned to how archaeologists and projects produce knowledge. Photography, modern survey tools, computers, relational databases, GIS applications, and global connectivity have all shaped our expectations for how we analyze, publish, and preserve data collected from the field.

While I admit that my case study is idiosyncratic, it nevertheless represents some of the ways in which small projects rely on off-the-shelf components that shape the ways in which participants engage with the data collection and analysis process. The limited and distinct character of many small project datasets has influenced the structure of small project data. The way in which archaeological knowledge is structured in a digital environment impacts the immediate and potential utility of small project data and this, in turn, has implications for the longterm preservation of this material.

The deeply interconnected nature of data collection, utility, and preservation highlight the social character of archaeological knowledge and reminds us that the tools we use shape the knowledge that we produce.