The Chicken or the Egg: Highlighting the Importance of Beginning with Deliberate Database Design

Jenna Milinsky, Martin P. Walker, Elizabeth Albee, Miranda Campbell, Abigail Huffman, and David G. Anderson
Department of Anthropology, University of Tennessee, Knoxville

Introduction
One of the basic elements of any current archaeological research project is the database, this could not be more apparent than when looking at the National Science Foundation’s grant applications and the required Digital Management Plans, or such growing projects as the Paleoindian Database of the Americas (PIDBA) (Anderson et al. 2010) or the Digital Index of North American Archaeology (DINAA) (Wells et al. 2018). Despite the increasing importance of the proper management of data within current research projects, discussion of careful and deliberate database design is rarely noted. This poster aims to present an optimal database design will allow for all current and any future research agendas, as well as being able to integrate seamlessly into GIS platforms.

The specific data that this database is constructed around are from the University of Tennessee, Knoxville’s ongoing Late Prehistoric archaeological project at the Topper Site (38AL23). The Topper Site is a multi-component site located in Allendale County, South Carolina. Excavations at this site stretch back several decades and have explored the Paleoindian occupation in the area extensively. More recently, attention has turned to the late Precontact presence in the area. With a surge of new data resulting from UTK excavations, questions arose concerning data management strategies.

Outside of the more well known, larger archaeological databases, such as PIDBA or DINAA, many project-based databases usually conform to the following model: 1) research question development, 2) fieldwork/data collection, and 3) analysis, with database design being based upon what is needed to complete the specific analysis being conducted.

Unfortunately, this approach often leads to databases that are designed almost exclusively for the specific scenario within which they are developed. This leaves any others who wish to utilize the database forced to start from the beginning and recreate their own database from the raw data to obtain what is required for their new analysis. The careful outline of the structure and flow of this database, with the goal of future adaptability, in addition to ensuring the pre-planning of data fields at the outset of the project, has been vital in the efficient management of both field and laboratory time and practices.

Methods
The database presented here was constructed within Microsoft Access 2010. One of the reasons for this choice was the ability within Access to build graphic user interface (GUI) forms (see Feature Form in figure 1; Lithics Analysis Sheet in figure 3; Ceramics Analysis Sheet in figure 4) that allow for easy and accurate data entry. This is especially necessary for any ongoing, multi-year project that includes users of all training levels. Within the form building tool in Access we can make certain field mandatory entry, we can make some fields hierarchical, and we can even create limited lists of typologies to limit possible subjective inputs.

Another factor that led to the use of Access for this database was the ability to relate multiple relational tables to each GUI form, enabling the creation of multiple, specific databases with one pass of data entry. This increases the processing of raw data in an efficient manner, as well as increases the capacity of our database to handle multiple parameters of analysis and data sets. Additionally, this streamlines crossover with GIS packages via the syncing of specific raw data into geodatabases. Figure 4 illustrates our ability to gather basic information regarding units and features that housed ceramic artifacts while also gathering more specific data about these ceramics such as counts and weights but then having these data be stored in unique data tables. These unique data sets enable for efficient linkages to GIS packages that then allow for the creation of maps such as the lithic debitage density map in figure 3.

Discussion
Edward Tennant, in his work with archaeological data and GIS, states, “Datasets used in the creation of living documents should meet four criteria: (1) that the dataset remain accessible by more than one person; (2) that it can integrate with other types of data such as those from the natural sciences; (3) that it is easily updated with future research data; and (4) it results in the creation of accompanying documentation” (2007). These are some of the things we aimed to include in our database construction. To aid in this goal, we worked to create a database that was accessible, adaptable, and compatible with cross-discipline analyses. In creating a full and comprehensive database we avoid leaving out information that could be vital for future research and that is impractical or impossible to add at a later time. To maximize accessibility, we intend to make this database open source to enable other researchers to access the information in order to employ the data with their own research projects. This is one of the most important aspects to functional and adaptable database design. By making raw data widely accessible, we enable the largest capacity of future research projects and analyses, allowing for greater collaboration and furthering the development of our understanding of human history.

Acknowledgements
Thank you to Achroma, Inc. for their continued support of the archaeological research on their property and for being such excellent hosts. Thank you to Southeastern Paleohemeric Survey (SEPS) for providing logistical, equipment, and volunteer assistance in the field. A mountain of gratitude and praise to Cayla Colclasure for providing the GIS maps and showing what this puppy can do. Lastly, many thanks to the students and volunteers for the diligent help processing materials and the hard work and time put into collecting the data for this project.

References