Extended Abstract

Archaeological Research is often found to be heavy on reasoning prowess and could be significantly aided by data intensive analyses. Excavations provide archaeological researchers with material objects and features that serve to imagine and reconstruct the past. When such hypotheses are put together to draw up the life and times of our ancestors, disagreements between various hypotheses often arise; upon such circumstances, researchers need to go back to the evidences to critically examine and refine their positions. Consider the case of discovery of some remains of a structure that seems to have characteristics of either a ship or a farmhouse. The absence of any evidence of a water body in the premises could be used to weaken the hypothesis that the remains were those of a ship and strengthen the farmhouse assumption. Digitized documentation of the remains could help speed up this such hypothesis framing and decision making. A decision support system to enable and ease such research should hence be able to provide real-time data dicing capabilities and provide insightful analyses of various subsets of data chosen (according the users’ interests). Coming back to our example, based on the knowledge that pebble like round stones are found close to water bodies, the researcher should be able to query the database to find the number of such artifacts that were found close to the remains in question. If the number of such pebbles (or round stones) are found to be not as abundant as expected close to a water body, the researcher could focus more on the farmhouse hypothesis. In this paper, we focus on our experiences towards building a digital database of archeological findings that would aid such analysis. The assemblage of artifacts which are often too fragmentary and disconnected, demands corroborations and permutation combination analysis which could be achieved through a well conceived digital data base. It can contribute towards scientific sketching of the lost past.

Documenting archeological data digitally is often a challenging task. This is due to various factors such as heterogeneity of data types, possibilities of
error in data entry and due to the nature of incremental build up of artifact metadata. Often data is recorded in Microsoft Excel spreadsheets. Instead, in the 2011 Season of Pattanam Excavations, we switched to a software that is custom-built for data entry. We now outline our considerations in moving towards such a method (contrasting it with conventional Excel spreadsheet style data entry).

**Error Free Data Entry:** Consider two trenches 1 and 2 that have 2 and 3 loci respectively. However, an artifact could be mistakenly labeled as belonging to locus:3 in trench:1 (whereas trench 1 is known to have only 2 loci). Such errors are less likely to be detected by the data entry operator who has limited knowledge about the excavation trenches. These errors, obviously, could easily propagate to the Excel sheet too; due to the absence of any sophisticated error checking mechanism. However, in the custom built software that we have developed, there are provisions to incorporate such domain knowledge about trenches and loci; such knowledge is used by the software to disallow entry of such erroneous data.

**Canonicalization:** The same artifact may be casually referred to, by different names. As an example, broken glass bead and broken bead of glass could be used interchangeably. Canonicalization is important so that artifacts that are referred to by either of these entities are identified to be the same, this is obviously important while performing deep analysis on the data. In our software, we include a dictionary (presented as a droplist) of canonical artifact type names with provisions to choose from among them during data entry; this decreases the likelihood of minor variations in type names from passing off as different types. It may be noted that such dictionary based selection of artifact types cannot be incorporated in Excel.

**Parallelization of Data Entry:** The software solution for data entry is web-based and can be accessed by any server on the network. In large excavation sites, this aids parallelization of the data entry process enabling multiple users to enter the data from different terminals at the same time.

**Easy Updating:** When an artifact such as a bangle is found, some details are recorded in the Tally Sheet that is filled up at the excavation site on a day to day basis. However, certain finer details (e.g., width of the bangle, diameter etc) are recorded only at a later time since such measurements are taken within facilities that have equipment to record accurate measurements. In an Excel based data entry system, when more details including laboratory/expert analyses are available at a later date, the corresponding record has to be found from among the many records available and then updated. This is cumbersome and error prone and could be easily avoided in a software solution that allows for efficient search. We enable such functionality by providing multiple ways to search through artifact records and drill down to the one in question, making the updating easy and almost error free.

**Seamless Linkage of Heterogeneous Data Sources:** Artifacts, in addition to being described by their various attributes such as type, shape and color, often have an accompanying photograph and/or an illustration. In a software solution, we provide with functionality to view all attributes pertaining to
an artifact together, in a single page. Such linking with data sources such as photograph files is obviously beyond the scope of a spreadsheet software such as Excel.

In this paper, we describe how the aforementioned features have been enabled in our software, with illustrative screenshots that describe functionalities of interest. We also cover our experiences in building such a software with the hope that it would contribute towards compiling a set of replicable models for digital documentation of archeological data.