



Data Collection in Zooarchaeology: Incorporating Touch-Screen, Speech-Recognition, Barcodes, and GIS

W. Flint Dibble

Author address: Department of Classics, University of Cincinnati, PO Box 210226, Cincinnati, OH 45221-0226, USA.
Email: dibblewf@mail.uc.edu

Received: April 14, 2015
Published: December 16, 2015

Volume: 6(2):249-257
© 2015 Society of Ethnobiology

Abstract: *When recording observations on specimens, zooarchaeologists typically use a pen and paper or a keyboard. However, the use of awkward terms and identification codes when recording thousands of specimens makes such data entry prone to human transcription errors. Improving the quantity and quality of the zooarchaeological data we collect can lead to more robust results and new research avenues. This paper presents design tools for building a customized zooarchaeological database that leverages accessible and affordable 21st century technologies. Scholars interested in investing time in designing a custom-database in common software (here, Microsoft Access) can take advantage of the affordable touch-screen, speech-recognition, and geographic information system (GIS) technologies described here. The efficiency that these approaches offer a research project far exceeds the time commitment a scholar must invest to deploy them.*

Keywords: Zooarchaeology, Digital archaeology, Archaeological database, Touch-screen, Database, GIS, Barcodes

Introduction

Modern technology commonly facilitates the process of archaeological data collection, especially on large datasets with thousands of entries. While research teams generally recognize the need for well-crafted, rigorous project-wide databases, too frequently individual researchers persist in using low-tech solutions (paper and pencil, train-of-thought word processing documents, or disorganized spreadsheets) for large datasets.

Zooarchaeological data recording is an essential, but time-consuming and tedious process. Detailing the attributes of an individual bone specimen – and all its potential value for interpretation – into database format requires utilizing many arbitrary codes or IDs representing excavation context, taxonomic status, anatomical location, tooth-wear stage, etc. (Driver 1992; Gifford and Crader 1977; Kansa and Kansa 2013, 2014). The potential for transcription errors can be high, especially among zooarchaeological assemblages where analysts are working under budgetary and time constraints, or in challenging field settings. A variety of new digital approaches for data collection offer high potential for a dramatic improvement in efficiency in the lab as well as a substantial reduction in the potential for data-recording error that is inherent in conventional lab practices.

Barcode labels, alongside legible text, enable a rapid and accurate printing and recording of excavation IDs. The cost of the label paper and scanner is easily offset by the time and errors an analyst can save by making this change. Integrating this coded information to a spatial representation of anatomical elements (a simple GIS), it is possible to design a data entry system where analysts can click or touch an anatomical zone or landmark, rather than memorizing codes. Touch-screen or speech-recognition enabled databases are simple to design through the use of buttons, which prompt users for information and answer questions. The change from pressing “g” for goat to touching a labeled illustrated button, or speaking aloud its caption, can increase efficiency and reduce a significant amount of error. Speech-recognition allows for hands-free recording and an investigator to remain focused on the material. None of these methods are innovative but the combination creates a synergetic data entry system with a wide range of potential use in various field and lab settings. However, it should be noted that, while the digital recording methods described below will reduce human errors, they will not entirely eliminate all forms of careless mistakes, nor the incorrect use of basic zooarchaeological methods.



Although few zooarchaeologists have strong backgrounds in programming or database design, there is much that one can do to tailor applications to specific needs with minimum training (Jones and Hurley 2011). It is possible to adapt the methods described here to fields other than archaeology, zooarchaeology, or zoology. It is also possible to adapt these methods for other software (Filemaker Pro) and non-windows touch screen devices (Android or iOS); however, this requires a different set of software, coding, or database-design skills (e.g., see the blog paperlessarchaeology.com by John Wallrodt for more suggestions on creating touch-screen Filemaker Pro databases). Microsoft Access (MS Access) provides a database interface known to many current practitioners, and the tools presented below simply enhance the user interface within one's own database. Crucial skills utilized here include a basic knowledge (but not expertise) of relational database design and the MS Access design interface, as well as a willingness to learn about coding specific actions (such as pressing a button) into an existing database.

While it is impossible to predict which pieces of technology will become obsolete in the medium- to long-term future, the methods presented here do not necessarily require a change to the structure of the data or its need for archiving. Given the inevitable challenges due to changing devices and software obsolescence, it is important that today's scholars should try to "keep up" with technology in order to stay current in the scientific world. Acknowledging these challenges, the methods presented here are designed to be low-tech, simple solutions in order to take advantage of current, widely available technology.

Redundant Data Collection Process

It is essential to first design one's data recording process prior to designing a complementary digital system. I designed this database to facilitate my recording of ca. 20,000 zooarchaeological specimens over the course of two years from the sites of the Athenian Agora, Azoria, and Nichoria in Greece. I developed the following step-by-step process (adapted from Halstead 2014) with the intention of increasing efficiency and reducing identification and data-entry errors:

- 1) Sort bags of zooarchaeological material into desired contextual assemblages (e.g., chronological, spatial, etc.) based upon research questions and preliminary observations.

- 2) Label all potentially identifiable specimens (as determined by the project's recording protocol) in each bag with a printed barcode tag containing a unique zooarchaeological ID (e.g., stratigraphic unit + sequential number)
- 3) Sort labeled specimens by anatomical element (e.g., humerus, femur), laid out on a table with all other specimens of the same anatomical element and the same contextual group.
- 4) Sort each anatomical element by taxon (e.g., pig, sheep/goat, equid) with reference to a representative comparative collection.
- 5) Sort each group into left vs. right vs. indeterminate sided.
- 6) If appropriate, sort each group by age or sex indicators, and/or proximal and/or distal halves.
- 7) Determine minimum counts of each anatomical unit (e.g., proximal pig humerus).
- 8) Following the project's zooarchaeological recording protocol, record each specimen into the database, organized by the above contextual, anatomical, and taxonomical groupings.
- 9) Repeat each step until the entire assemblage has been analyzed.

As in an assembly-line, the analyst focuses on only a few redundant variables at a time. Stackable trays or portable shelves enable a specialist to augment restricted table space, often needed for sorting large assemblages. By leveraging the redundancy in process and restricting the focus to just a few variables, it is possible to take advantage of existing technologies to improve and automate steps and correspondingly reduce operator error.

Human Error

A clear strength of an organized, efficient work flow that also takes advantage of digital approaches is a reduction in data-recording error. Human mistakes are especially grievous when they involve archaeological context. Such errors are common both when initially assigning IDs and when repeatedly transcribing IDs in the course of organizing, cataloging, analyzing and curating archaeological remains. Digitally produced and recorded IDs allow archaeologists to both reduce error and save time. These improvements can be demonstrated through identification and data-recording experiments.

To understand the frequency and nature of human errors that occur during analysis, we conduct-



B 4103. 04-59	B4103.04-59
B 3103. 09-2	B3103.09-2
B 3054. 05-84	B3054.05-84
B 4103. 14-44	B4103.14-44

1 7 2 7

A European 1, American 7, sloppy 2, or a 4?

Figure 1. Examples of handwritten mistakes from the timed data entry experiment.

ed several experiments. In the first, eight undergraduate students each labeled a collection of 40 lithics, not knowing they would be tested for error (Dibble and Dibble 2014). Following this step, each student then transcribed 40 IDs from their neighbor's assemblage. Finally, another student verified the transcribed IDs against the labeled lithics. Out of 320 total labeled lithics, 18 were transcribed incorrectly on the final sheet (a 5.6% error-rate).

In another experiment, 25 participants (PhD students and recent PhDs) each recorded the same set of 20 archaeological identification codes three times: once by hand, once with a keyboard, and once with a barcode scanner (Figures 1). This resulted in 500 uniquely entered records recorded in triplicate. Participants were aware that they would be timed and checked for errors. The author timed each participant's data-entry with a stop-watch (Figure 2, Table 1). Despite the fact that many participants said they would proceed slowly in order to avoid errors, they made frequent mistakes. Illegibility was the leading cause of a 4.4% error-rate for the 500 handwritten IDs, while typos presented a 2.6% error-rate for the 500 keyboard-input IDs and a 0% error-rate for the 500 barcoded IDs. These error rates can be compounded by the common excavation workflow, where a tag is first written out by hand (or a specimen labeled), and later keyed into a database. The above results are comparable to other published error rates from a variety of data-entry studies conducted in various professions where such rates fall around "a few percent" per cell on a spreadsheet (Panko 2008a). For more complex, multi-step tasks, perhaps analo-

gous to entering all the variables from an archaeological specimen, the error rate is generally far higher, ca. 30% of all records containing at least one error. Furthermore, the ca. 80% error detection rates observed in most proofreading studies suggests that many errors are not caught (Panko 2008b). Taking this logic to its ultimate conclusion, it is probable that ca. 6% of all recorded specimens in any given assemblage contain some form of error ranging from minor to grievous typos.

Therefore, given the tenacity of human error, it is crucial – especially when considering the scale of archaeological data-entry – to design robust analytical workflows. Ideally, these systems should be designed to minimize unconscious mistakes made when transcribing information onto data labels or databases. While some detailed suggestions are presented below, it is also recommended to use digital devices (calipers, scales, etc.) whenever possible since they can transfer data directly to the computer with few transcription errors (McPherron and Dibble 2002).

Barcoding Archaeology

Printed barcoded labels is a simple method already in use on many archaeological projects for reducing errors and speeding up archaeological labeling and recording processes in the field and lab (Dibble et al. 2007; Dibble and Dibble 2014; McPherron and Dibble 2002). Labels can be custom designed to include whatever printed text one wishes, with a barcode at the bottom of a tag representing an archaeological provenience or 'ID' (Figure 3).

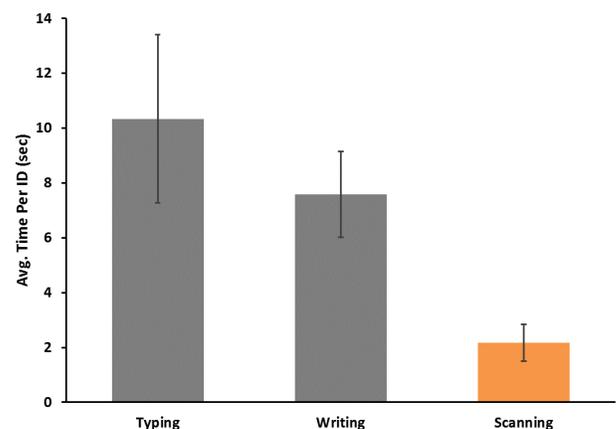


Figure 2. The time results for the data entry experiment in a bar graph showing average time per ID entered (25 participants entering 20 IDs each) with standard deviation bars.

**Table 1.** The results from the data entry experiment organized by participant, data entry type, and number of mistakes.

Participant	Typing	Errors	Writing	Errors	Scanning	Errors
1	211	1	140	3	30	
2	217		165		56	
3	146		128	4	37	
4	267		215	1	70	
5	172		125	3	47	
6	243	1	165		55	
7	234	2	119	1	35	
8	230	1	155	1	48	
9	160		142	1	50	
10	230		114	1	45	
11	243		175		50	
12	201	1	161		50	
13	220		232		32	
14	145		123		34	
15	158		115		33	
16	308	1	170	1	45	
17	200	2	138	1	36	
18	198		184		46	
19	126		110	2	21	
20	134	1	157		32	
21	167		110	2	38	
22	400	1	168		85	
23	209		145		42	
24	124	1	156	1	35	
25	225	1	176		36	
Total	5168	13	3788	22	1088	0

Therefore, even if barcodes become obsolete, data are still recorded in the layout one wishes on a tag. It is possible to affordably print tags (with or without barcodes) on paper, mylar, polyethylene, or many other materials.

Due to the idiosyncrasies of archaeological cataloging, research, and curation systems it is important to approach label creation and ID assignment in a project-by-project *ad hoc* manner. A printed labeling system needs to be robust enough to deal with potential joining fragments, “bags within bags,” and needs to readily integrate (through both textual and visual vocabulary) within the larger archaeological project. The barcode merely duplicates in a digital-readable format the standard conventions of a project.

Automating the printing of tags can save a significant amount of time and reduce error. For example, from an archaeological context with 75 identifiable specimens, it is simple to instruct the

computer to print tags for all 75 specimens at once, each with a unique-sequential ID. This eliminates 74 chances to record an ID incorrectly, and the printed labels can easily be checked as a block prior to assigning them to individual specimens and sorting the labeled specimens into larger contextual assemblages. In addition, scanning a barcode is a virtually error-free method for recording an archaeological ID. The risk of scanning the wrong barcoded tag is the same as data-entering the wrong tag. The tag includes any text one wishes, therefore, it is still possible for anyone to visually inspect a tag or do manual data entry.

A Button-Based Database

Most database entry forms used in archaeology rely upon a combination of drop-down boxes or text boxes. While both are useful, neither truly solves the issue of typos. While restrictions on variables in drop-down boxes do limit spelling mistakes, they do not adequately prevent an operator from mistyping and



Figure 3. Example barcoded labels designed in ArchCode, it is possible to create whatever template you want incorporating any fields in a table.

accidentally selecting a wrong choice from the restricted variables. Typos are grievous errors to commit because they do not allow the effective querying of one's results. In a very real sense, spelling counts when analyzing data.

After all, archaeological codes include not only spatial context, but frequently a large variety of variables coded into alphanumeric shorthand make archaeological datasets difficult reading. While data validation routines can identify and sometimes clean problematic data (Kansa and Kansa 2014), certain similarly spelled words or codes, in addition to numerals are notoriously difficult to retroactively identify and fix.

A simpler user interface involves designing a single recording form for each field in a table,

including buttons on the data-entry form corresponding to the most common responses for the field (Figure 4). Clicking a button enters the data and proceeds to the next entry form. If the data table is extremely complex it is possible, through VBA coding within MS Access, to order the data entry process sequentially or create forms that adapt to the entry as it progresses.

Once designed, button-based databases are easy to use with touch-screen and speech-recognition software and hardware available on affordable new computers. Buttons can be 'pushed' via clicking a mouse, touching a screen, or (if speech-recognition is activated) speaking aloud the caption (button captions in MS Access are automatically 'listened for' by the native Windows 7 and 8 speech-recognition software).

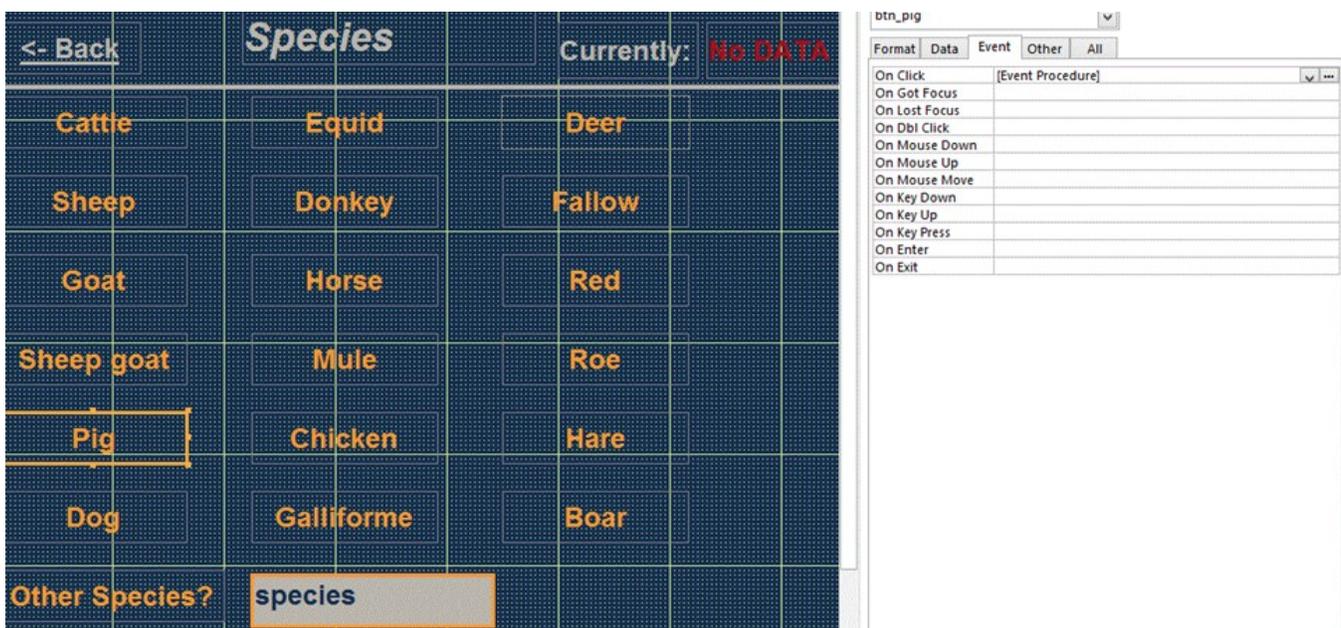


Figure 4. The Design View for FRM_Species with buttons for each common taxon.



Both touch-screen and speech-recognition are extremely easy to use as the verbal or tactile nature of data entry keeps the focus on the actual information one is recording, rather than struggling to transpose and type a code or awkward archaeological term.

Lastly, speech-recognition provides the added benefit of hands-free data-entry, meaning one's concentration can remain unbroken from the archaeological specimen. The native Windows 7 and 8 speech-recognition software is adequate for reliably recording information in English via buttons but is not adequate for recording 'freehand' sentences. Since the speech-recognition software is "listening" only for the captions of the buttons, it can swiftly and accurately recognize complex terms ("carpometacarpus") and distinguish between similar terms ("sheep goat" vs. "sheep" vs. "goat"). If the software is confused by a spoken command, it will ask the analyst for clarification. Therefore, speech-recognition can only be relied on for specific pre-designed responses, but not for populating text boxes.

Integrating Spatial-Anatomical Information within a Database

It is also possible to create a simple GIS of spatial-anatomical zones (e.g., crania, post-crania, hind limbs, etc.) overlain onto an illustration of a skeleton (Figures 5 and 6) whereby anatomical zones are linked to a button-based database. This enables clickable buttons to be placed on an image on the data-entry form, approximating the size of each zone. Figures 5 and 6 show zones defined for specific anatomical elements utilized for cutmark recording following anatomical templates provided by Popkin (2005) and zones defined by Dobney and Rielly (1988) (also see Orton 2010). These buttons and linkages can be adjusted depending on the focal species and eliminates the need to memorize, look-up, or mistype an arcane zone value. The buttons work with touch-screen or mouse.

Since the zones are saved in a standardized fashion (e.g., in a cutmark table), it is additionally possible to export these results to GIS software and run explanatory spatial analyses illustrating which element was most well represented in a given assemblage or which element had the highest frequency of cutmarks. GIS software can consider each zone as a polygon and there is no need to adjust them to a coordinate system (after all, each specimen is of a different size). It is necessary to record cutmarks within their own data table since there might be many

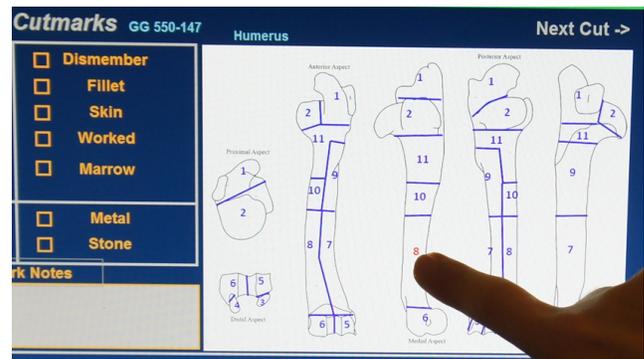


Figure 5. Recording the spatial zone of a cutmark (photograph by Jonida Martini).

cutmarks on each specimen. The example presented in Figure 6 from the Athenian Agora illustrates the high frequency of chop marks evident on the proximal anterior tibiae (Zone 4) from the removal of the patella (while no other zone had even 10 chops, the Anterior Zone 4 had 30 such examples). This butchery pattern, readily visualized through spatial analysis, highlights the introduction of a standardized butchery technique associated with the adoption of the cleaver in urban contexts during Classical period Athens (Dibble 2014). This example of a GIS visualization highlights the utility of combining different software while conducting data-entry and exploratory spatial (in this case anatomical) analyses in field and lab settings.

Time is Money: The Cost of a Digital System

Surprisingly, according to the above timed studies of data-entry, it actually takes longer to use a keyboard to type out a unique archaeological ID (A3102.03) than it does to write it out on paper (Figure 2). Cumulatively, it took 25 individuals 85 minutes to type 20 ID codes each (500 total entries). IDs or codes are often awkward to recall, let alone type, and thus they reduce time for zooarchaeological analysis. Moreover, specialists frequently record the same ID twice or more (the handwritten ID associated with an object, its data, whether it has been photographed or not, etc.). Therefore, each ID in a hypothetical 10,000 specimen assemblage might be written or typed two to three times for a total of well over 50 hours of work.

The experiment above suggests that scanning a barcoded ID is approximately 400% faster than writing it, and 500% faster than typing it. In addition, this speed does not account for the mental distraction of typing in ID codes, nor the time spent correcting



Athenian Agora Tibias: Chop Frequency by Zone

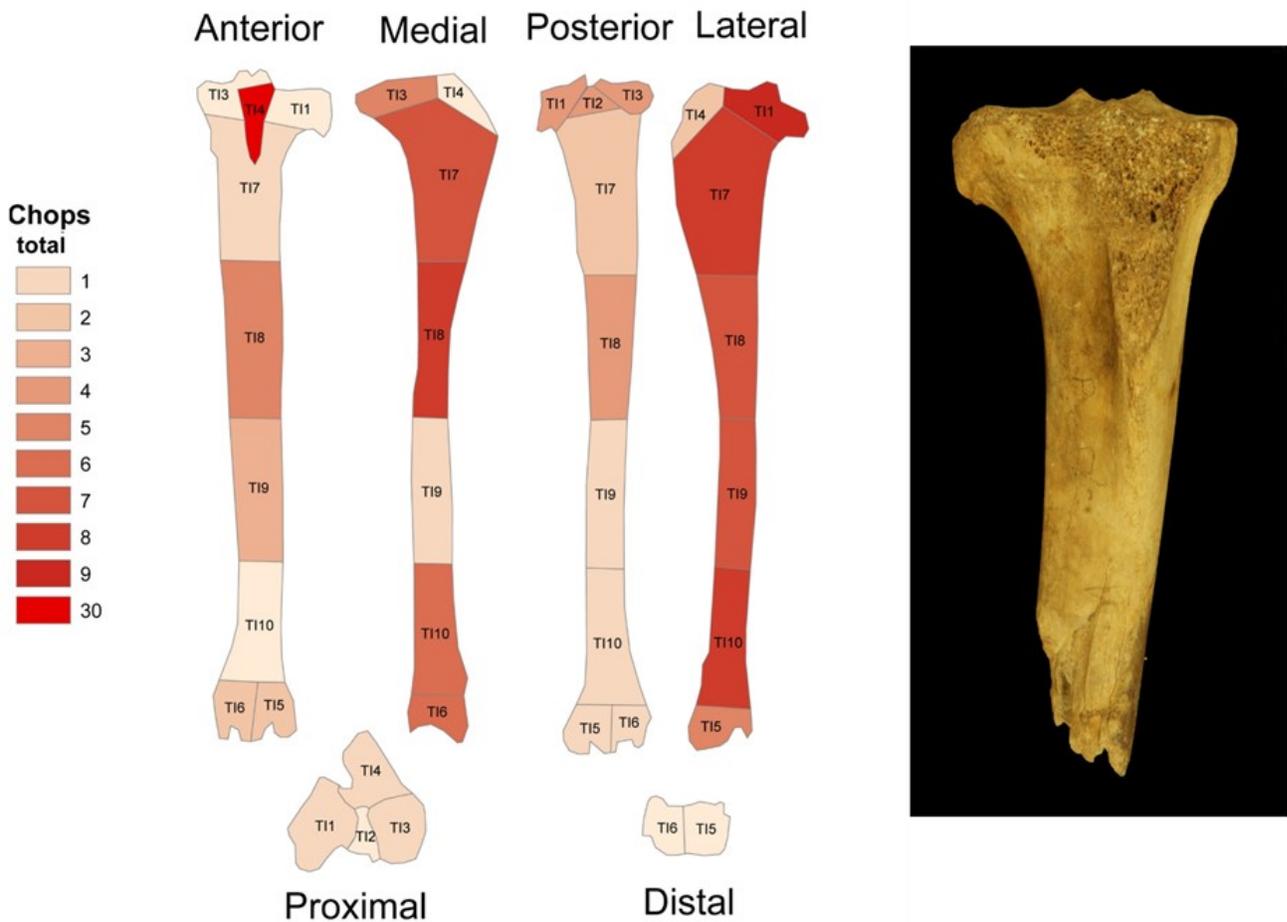


Figure 6. An example GIS Output of Chop Marks on Tibias from the Athenian Agora (most chops derive from Anterior zone 4 seen clearly in the accompanying photograph of a specimen taken by Jonida Martini).

human mistakes. This also does not include, as mentioned above, the time saved in automating the printing of labels, rather than laboriously writing them out.

Therefore, while implementing this digital system incurs some up-front expenses, the amount of time an analyst saves should provide a financial cushion (fewer travel and cost-of-living expenses incurred over the course of a project). Importantly, the analyst can carry forward the expenses in equipment and design time to future projects. The methods and materials described above are affordable to most scholars. Touch-screen functionality requires a Windows 8 handheld device with a touch-screen (ca. \$400 USD) running MS Access (academic license for

Office 365 ca. \$70 USD). Speech-recognition works better with an external microphone (ca. \$30 USD).

The expense of barcoding is also quite minimal, although this depends on what material one wishes the tags to be printed on. The program, ArchCode, co-designed by the author and Harold L. Dibble, is available for free at www.oldstoneage.com and has been tested on Windows 7 and 8. ArchCode should require no additional coding but will read/write to a single table in a MS Access database file (.mdb extension). It is possible to print out 10,000 labels on sticky label sheets to be integrated with each specimen in a small plastic bag for a total budget of under \$500 USD. However, these might be destroyed in the course of a field-project and need replacement.



Indestructible, archival plastic tags (polyester, polypropylene, or polyethylene) incur a larger expense but solve the above problem. The tags are more expensive (ca. \$600 USD for 10,000 tags) and a thermal-transfer printer (\$300+ USD) and printer ribbons are required to print on archival quality tags.

Conclusion

Affordably incorporating 21st century technology within archaeological data recording systems serves to increase the efficiency of field and lab based research and to decrease the incidence of human errors. The only real expense is an investment in time prior to designing a large data recording project; however, this investment enables researchers to maximize their data recording time, leading to a net-gain in research capacity. The database described in this paper is available for download at paperlessarchaeology.com. Hopefully, the examples presented above will convince scholars that it is worth the effort to create simple, yet powerful code to enhance one's database by incorporating a variety of current technologies. Each of these technologies need not replace a scholar's current data structure but rather enhance the custom data entry interface.

The utilization of barcodes, GIS, touch-screen, and speech-recognition, combined with a minimum knowledge of software programming, can help create a robust and efficient data entry system. While none of these technologies are new to archaeology, investing time to creatively deploy such data management technology can save researchers significant time and reduce mistakes. Improving the quantity and quality of zooarchaeological data will lead to stronger results and new research avenues.

Acknowledgments

The development of the database and software described above was made possible due to research funding provided by the American School of Classical Studies at Athens, the Archaeological Institute of America, and the University of Cincinnati. Laboratory supplies were provided by the Malcolm H. Wiener Laboratory of the American School of Classical Studies at Athens, the Institute of Aegean Prehistory, and the Azoria Project. The undergraduate lithic labeling experiment was conducted by Harold Dibble. Thanks are due to Harold Dibble, Sarah Kansa, Iain McKechnie, John Wallrodt, and two anonymous reviewers for comments on the paper. I also thank Sarah Kansa and Iain McKechnie for encouraging the submission of this paper.

Declarations

Permissions: None declared.

Sources of Funding: The study received funding from the American School of Classical Studies at Athens, the Archaeological Institute of America, and the University of Cincinnati.

Conflicts of Interest: None declared.

References Cited

- Dibble, W. F. 2014. Urban Butchery Patterns from the Athenian Agora and Azoria in Greece. Paper presented at the 12th International Conference of Archaeozoology: San Rafael, Argentina.
- Dibble, W. F. and H. L. Dibble. 2014. Barcoding Archaeology: Digital Methods for Error-Free and Rapid Labeling, Data-Entry, and Inventorying. Poster Presented at the 115th Annual Meeting of the Archaeological Institute of America. Chicago, IL.
- Dibble, H. L., C. W. Marean, and S. P. McPherron. 2007. On the Use of Barcodes in Excavation Projects with Examples from Mossel Bay (South Africa) and Roc de Marsal (France). *The SAA Archaeological Record* 7:33-38.
- Driver, J. C. 1992. Identification, Classification, and Zooarchaeology. *Circaea* 9:35-47.
- Dobney, K. and K. Rielly. 1988. A Method for Recording Archaeological Animal Bones: The Use of Diagnostic Zones. *Circaea* 5:79-96.
- Gifford, D. P. and D. C. Crader. 1977. A Computer Coding System for Archaeological Faunal Remains. *American Antiquity* 42:225-238.
- Halstead, P. 2014. The Faunal Remains. In *Nemea Valley Archaeological Project, Volume 1: Early Bronze Age Village on Tsoungiza Hill*, edited by D. J. Pullen, pp. 741-804. The American School of Classical Studies at Athens, Princeton.
- Jones, E. L. and D. A. Hurley. 2011. Relational Databases and Zooarchaeology Education. *The SAA Archaeological Record* 11:19-21.
- Kansa, E. C. and S. W. Kansa. 2013. We All Know That a 14 is a Sheep: Data Publication and Professionalism in Archaeological Communication. *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 1:88-97.



- Kansa, S. W. and E. C. Kansa. 2014. Data Publishing and Archaeology's Information Ecosystem. *Near Eastern Archaeology* 77:223-227.
- McPherron, S. P. and H. L. Dibble. 2002. *Using Computers in Archaeology: A Practical Guide*. New York: McGraw Hill.
- Orton, D. C. 2010. A New Tool for Zooarchaeological Analysis: ArcGIS Skeletal Templates for Some Common Mammalian Species. *Internet Archaeology* 28. Doi:10.11141/ia.28.4.
- Panko, R. R. 2008a. What We Know About Spreadsheet Errors. *Journal of End User Computing* 10: 15-21. Available at: <http://panko.shidler.hawaii.edu/My%20Publications/Whatknow.htm>. Accessed on February 2, 2015.
- Panko, R. R. 2008b. The Human Error Website. Honolulu, HI: University of Hawaii: Available at: <http://panko.shidler.hawaii.edu/HumanErr/Index.htm>. Accessed on February 2nd, 2015.
- Popkin, P. 2005. Caprine Butchery and Bone Modification Templates: A Step Towards Standardisation. *Internet Archaeology* 17. Doi:10.11141/ia.17.2.
- Wallrodt, J. 2015. Paperless Archaeology. Available at paperlessarchaeology.com Accessed on August 25, 2015.

Biosketch