

Construct a database or dataset of selected archaeological material/records, provide metadata statements in line with current Archaeology Data Service guidance, and a commentary explaining and justifying its structure and the categorisation/classification of the data.

Introduction

As I have yet to personally complete any kind of specialist work, and therefore lack the expertise to record and compile a dataset from an assemblage, this work will instead repurpose an existing dataset from a currently unpublished site. The site chosen, ‘Stibbe, Friars Causeway Lane, Leicester’ was an urban excavation located within the core of the old Roman and medieval town of Leicester and undertaken by University of Leicester Archaeological Services (ULAS) between 2016 and 2017 (Speed, forthcoming). The specific excavation data being utilised for this task is the medieval pottery data, with the analysis and reporting on this having been prepared by an external freelance specialist. The assemblage totalled 7,289 sherds from 506 contexts and spanned from the Late Saxon to the modern period.

The dataset was put together by the freelance specialist using their own custom database and, upon completion of the report, this data was then exported into an Excel spreadsheet format and forwarded to ULAS for its inclusion in the site archive. The primary issue with the data in this format however is that the specialist had input the data using custom codes for a number of elements of the assemblage (e.g. form, fabric, decoration), meaning that it requires consultation with multiple other documents in order to determine the specifics of the recorded data.

This issue and the fact that medieval pottery will be my future specialist focus meant that my approach to this task was centred around two main ideas – what information I believed would be key to include in a dataset when assembling one myself in the future, and how can I make the information in the dataset as accessible to other people consulting it, regardless of their level of specialist knowledge. Additionally, it was decided that the dataset would be formulated

in a Microsoft Excel spreadsheet as that is the favoured method for other specialists working at ULAS, as well as for archive deposition (Forward, 2023).

The blind selection phase

Before properly analysing the dataset to be repurposed, I wanted to take a blind approach to assembling a category list for the first iteration of my own version, in order to avoid being unduly influenced by the choices of the specialist who had prepared it. The initial steps involved consulting the guidance documents ‘A Standard for Pottery Studies in Archaeology’ (MPRG et al, 2016) and the CIfA ‘Toolkit for Specialist Reporting: Advisory Checklist’ (2020), specifically the ‘Full Analysis’ part of the checklist. From these I tried to figure out what my initial rough categories needed to be. The standard (2016) was especially clear on what aspects should be recorded and the trickiest part was trying to assess which pieces of information should be listed in their own columns, and which would be better grouped into a catchall ‘Additional info’ type category that I knew was a common feature in specialist datasets. The resulting list was still extensive and included: context number; sherd count; sherd weight; fabric type; whether the sherds were base, rims, or body pieces; the type of vessel; date range of the fabric; the forms of things like spouts, feet, handles etc.; surface treatment; decoration; potter or merchant marks; dimensions of the vessel; the number of vessels present; evidence of usage; overall sherd condition; and if sherds fitted together, in the same context or across contexts.

With these categories established I then tried to figure out the best column order would be. Ideally, I wanted to find a balance between having the columns flowing left to right in the order that the analysis of an assemblage would follow, while also keeping as many of the categories I considered the most likely to be of interest/importance to a non-specialist on the left-hand side of the spreadsheet.

The clear candidates for categories of primary importance to a non-specialist were context number, sherd count, sherd weight, fabric type, and date range. The first four of these are fortunately the elements I know are usually recorded first from passive observation of the process. The date range however often requires more of the remaining categories to have been assessed before it can be accurately determined, and so from a specialist perspective this would perhaps be one of the later columns filled in. Scrolling back to the left repeatedly after finishing assessment of the sherds just to input the date range would potentially become tedious work

though. Options considered as solutions to this problem included a simple column duplication, some form of conditional formatting i.e. using colours to indicate date ranges, or even attempting to code a macro into the spreadsheet's VBA to automatically move the data when it was entered. For now, the 'date range' category was placed in the penultimate spot before the 'Additional info' column.

Of the remaining categories, ones such as sherd type or the form of component parts seemed best suited to being subdivided into separate columns for each, e.g. one designating the number of rim sherds, one designating the number of base sherds, etc. The rest of the categories appeared as though they would work as individual columns though, and so they were quickly arranged based on a combination of my developing understanding of the assessment process, and the order they were listed in the guidance documents. Figure 1 shows the layout of the columns as they were at the end of this part of the exercise.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	CONTEXT NUMBER	SHERD COUNT	SHERD WEIGHT (g)	FABRIC TYPE	BODY SHERDS	BASE SHERDS	RIM SHERDS	RIM DIAMETER	BASE DIAMETER	HEIGHT	EVEs	VESSEL TYPE	HANDLE FORM	SPOUT FORM
2														
3														
4														
5														

	N	O	P	Q	R	S	T	U	V	W	X	Y
1	SPOUT FORM	FOOT FORM	SURFACE TREATMENT	DECORATION	POTTER OR MERCHANT'S MARKS	EVIDENCE OF USAGE	SHERD CONDITION	FITTED SHERDS	DATE RANGE	ADDITIONAL INFO		
2												
3												
4												
5												

Figure 1. Column arrangement of dataset prepared blind

Examining the dataset(s)

It was at this point that I finally did a proper examination of the dataset I would be working with, along with also making the decision to consult examples of two other recent datasets from separate pot specialists as references, just to see how far off course I had strayed. It was immediately obvious upon doing this that each of the three specialists had very different approaches to the task, and as such this would not simply be a case of moving the odd glaringly

misplaced column around that differed from the consensus. Instead, part of the challenge would be figuring out which approach seemed to work best on a category-by-category basis.

The first issue that became apparent though was one where I definitely differed from the consensus; I had completely neglected the ‘cut’ grouping, potentially leaving myself with a bunch of assorted contexts divorced of their relationships to each other, an embarrassing error to start things off with. Following consultation with my two reference datasets I then saw that they both included additional categories that, although I don’t have the data to fill in, seemed like they would have been useful for the dataset to have when trying to understand and characterise an assemblage: the area of the site the context was from (e.g. structure A, enclosure C), and the type of feature (e.g. pit, ditch). As this site has yet to complete its post-ex stage and be archived, I chose to add these to my dataset as ‘blank’ columns, to possibly be completed at a later stage.

Sherd count and weight were consistently placed in that order in all three datasets, as I’d arranged them. Their placement within the datasets varied a little but I was happy to keep them where they are for the workflow reasons stated earlier.

I opted to add an extra fabric column so as to include both the basic ware group and the specific fabric code, even though only one of my example datasets did this, simply because it adds to the accessibility for non-specialists. This decision also made clear that it wouldn’t make sense to split the ‘ware’ and ‘code’ columns for the fabric up, and since the specific fabric code might only be determined further into the analysis, coming back to the column later was inevitable. Therefore, if I was already doing that once it now made sense to simply shift the ‘Date Range’ column up to sit alongside it, which solved that conundrum. At this point I also noticed that I had ‘Vessel Form’ as a much later category than two of the specialists. Due to lacking experience this had felt like an identification that would happen later in the assessment, but I then recalled observing specialists looking over recently processed finds and the ‘form’ of pots was something they always deduced early on, at the same time as fabric. As such I shifted ‘Vessel Form’ to be closer to the fabric details.

Next was the issue of rim, base, and body sherds. In the dataset being repurposed these are not given individual sherd counts, whereas in one of the reference datasets each sherd type was assigned a separate row, and in the other they were given separate columns in line with my

arrangement. I think having separate columns gives a nice level of clarity, so I chose to stick with my original choice in this case, although it meant including further ‘blank’ columns. I had however missed another type of information from this portion of the dataset, the rim and base form, so this was added after the count for each of those groups respectively. Again, there was little consistency between specialists on the recording of dimensions, with only one recording base diameter and none recording height. After consulting the appendix of the guidance document (MPRG et al, 2016:33) it became apparent that base dimension has a somewhat niche usage, for pottery production sites, and so wasn’t a necessary inclusion. The rarity of height recording meant that that column was also excised. Rim diameter and Estimated Vessel Equivalents (EVEs) go hand in hand and so these were placed together and, although not recorded on the dataset being repurposed, it is recommended that along with EVEs a maximum vessel count is also recorded (MPRG et al, 2016:34), so another ‘blank’ column was included for this data.

Although the repurposed dataset included a couple of extra columns for the likes of ‘Surface Treatment’ (Glaze Colour, Slip Colour) and ‘Decoration’ (Incised Decoration, Applied Decoration, Slip Pattern), I felt like these could be covered adequately with a single descriptive column each, which also helped simplify the dataset’s appearance. Of the final categories (Handle Form; Spout Form; Foot Form; Potter or Merchant Marks; Evidence of Usage; Sherd Condition; Fitted Sherds) only ‘Foot Form’ now seemed too niche to have its own column, and so was withdrawn. What all three datasets also made obvious is that a final ‘Comments’ column was used as standard to cover any rarer occurrences (such as those categories excised in this phase), or unusual findings not covered by the dataset otherwise, so I was happy for my ‘Additional Info’ column to function in this way. I was then content with the remaining categories as I’d originally set them out, both content and order wise.

The Usability Stage

With the categorisations set it came time to add functionality to the dataset. The data being repurposed was not in an easily filterable format, so I needed to create a ‘Table’ in Excel to enable sorting and filtering of the data. I started by created a ‘dummy’ spreadsheet into which the repurposed data was added, along with transferring all of the information relating to the codes it employed onto the same sheet. The columns were then rearranged to match the format I’d selected for my dataset, and I proceeded to go through the process of applying a series of

VLOOKUP formulas to the coded entries to convert them into the format I desired. The data from the multiple 'Surface Treatment' and 'Decoration' columns was then combined into single columns using CONCAT formulas. Once all of the data was formatted as desired it was subsequently copied into the main dataset spreadsheet where it could now be converted into a 'Table' which would allow it to be readily manipulated as was necessary.

Conclusion

Although I am reasonably happy with the decisions made regarding the dataset and its categorisation, my lack of experience with inputting data into one means more of the decisions were made using intuition than I am entirely comfortable with. From a novice perspective the approach I have taken seems like it should function well, it was designed to be as ergonomic as possible, and to be understandable to those outside of the specialism. And while I can at least partially vouch for the latter, until it comes times to undertake an actual assessment of a pottery assemblage, I cannot be sure of its viability in the 'real world' of specialist work. What the notable differences between the datasets of the three specialists I consulted for this show though is that datasets are idiosyncratic entities, likely honed through years of trial and error to find what works best for each individual specialist. It will be interesting to return to this attempt in a decade and see how similar it is to my standard usage dataset at the time. Although I chose not to consult with other specialists directly for this exercise, I anticipate the influence of my mentors will probably show through strongly in whatever form it has taken on. I hope I'm at least still keeping the 'accessibility to non-specialists' factor in mind in any case.

Word count: 2,185

Bibliography:

CIfA, 2020. *'Toolkit for Specialist Reporting: Advisory Checklist'*, Chartered Institute for Archaeologists. Available at: <https://www.archaeologists.net/reporting-toolkit/checklist>, [accessed: 19/12/2023].

Forward, 2023. Personal conversation. December 2023.

MPRG, PCRG, SGRP, 2016. *A Standard for Pottery Studies in Archaeology*, Medieval Pottery Research Group.