

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

In the era of worldwide digital capability and interconnectivity, datasets – and facilitating easy, open access to them – have become increasingly important in archaeological investigation. Where once data dissemination was limited by its eligibility to be circulated in reputable publications, word-of-mouth if unpublished, the geographical location of hardcopies and practicable archiving technologies, the past four decades have seen increasingly rapid advances in storing and sharing archaeological datasets online.

Initiatives such as GO FAIR and the Journal of Open Archaeology Data [JOAD] seek to rectify the “inconsistencies found between different post-excavation products deposited in digital archives” encountered by May, Taylor and Binding (2023, Section 3.6). Advocating for data to be findable, accessible, interoperable, and reusable [FAIR] and providing a platform for openly available, professionally archived, peer reviewed, citable archaeological datasets and papers with a potential for high reuse is *de rigueur* (GO FAIR, 2022; JOAD, 2023). Consequently, the aims of major repositories (such as the Archaeological Data Service [ADS] and Digital Antiquity) to improve current preservation practices at the risk of permanently losing a significant amount of the archaeological record can be more readily met by archaeologists from all sub-disciplines operating nationally and, where possible, internationally (ADS, 2023).

The term ‘dataset’ will be used in reference to any set of data compiled or stored within a structured format, including Microsoft Excel; ‘database’ will refer to datasets compiled or stored in a database programme, such as Microsoft Access. ‘Archaeological project’ will be used as defined by Barclay *et al.* (2016, Section 3¹). Descriptions of field names in my dataset are outlined in the metadata document (Pearson, 2023a; Pearson, 2023b). Where applicable, acronyms will be introduced and used thereafter.

¹ Due to differing page numbers in the printed and online *Standard*, (sub)section numbers are used.

Despite efforts to encourage reusability in their design, archaeological datasets suffer from several drawbacks. A lack of consistency between archaeological companies/organisations has existed for years, and although guidelines from trusted bodies aim to rectify this issue, the guidance itself sometimes varies. Projects with limited scopes and resources can have detrimental impacts on datasets; a project with narrow aims may limit future use, especially when budgets do not allow for flexibility about what else might benefit other researchers, present and future, archaeological and interdisciplinary.

Central to FAIR data is logical and consistent metadata. In this essay I outline the key categorisations and classifications within my dataset and explain their relevance to the project's aims, justify the overall structure, comment on my reasoning for any changes and, where applicable, acknowledge what I would do differently in future projects and why.

The archaeological project and the pottery dataset

To provide a comprehensive commentary of my dataset (Pearson, 2023a), its structure and the categories used, the archaeological investigation should be defined. The assemblage documented contains pottery recovered from an archaeological excavation by Cotswold Archaeology [CA] in early 2023 known as "Land west of Coventry Road, Baginton" in Warwickshire. The site was approximately 500m south of The Lunt, a Roman fort dating to c. 60-80 AD. The archaeological evaluation established "there was significant Roman activity in the wider area around the site, predominantly focused north of the site and south of Lunt Roman fort" (CA 2019, p. 5). Furthermore, the dataset consists of pottery dating up to the end of the Roman period, c. AD 410. I was responsible for recording and analysing prehistoric and Roman pottery recovered from the excavation; post-Roman pottery was transferred to CA's medieval pottery specialist. The Roman assemblage is predominantly mid-to-late Roman, with a few early vessel forms. The selection strategy, as Historic England (2015) guidelines address, was to process, record, and include in the report all pottery found on-site. The recording criteria and methods are more extensive than those

utilised during an evaluation; the requirements are determined by an assessment of the project, its research aims, and the assemblage:

Analysis is the process of recording pottery to the level of detail appropriate for interpreting the assemblage within the framework of the project and wider research aims (Barclay *et al.*, 2016, Section 2.4).

Whilst pottery specialists have some autonomy, my dataset's structure and categories were largely predetermined by CA policy; for example, writing full pottery descriptions ('Pink grog-tempered ware') are not required and a column containing 'Fabric' codes (**PNK GT**) suffices. At CA, archaeological excavation data is recorded in a Microsoft Access database, which forms the project's archive. However, I used Microsoft Excel to analyse the pottery data. I adopted a CA template, with freedom to tabulate additional relevant information beneficial to my recording, analysis and reporting.

Prior to my involvement, the post-excavation team logged preliminary data about the pottery: deposit context numbers, registered artefact [Ra.] and sample numbers if applicable, artefact type (i.e., 'pottery'), sherd count, collective weight per context/Ra./sample (in grammes), and a section for comments. The Finds Log was prepared by adapting the 'Documentation' standards outlined by Barclay *et al.* (Barclay *et al.*, 2016, Section 2.2.2). Following a brief assemblage assessment, I supplemented these basic categories with additional information (specific or generalised fabrics, date ranges, and sherd condition) to bring it in line with CA's evaluation-level recording standards, and adherence to ClfA's *Toolkit for Specialist Reporting* (2023) at Level 2 where applicable (CA, 2022). This conforms to the recommended 'Basic Record' (Barclay *et al.*, 2016, Section 2.4.5); it is deemed 'appropriate' and adequately fulfils evaluation reporting requirements. Physical pottery clippings and key texts, most notably *The national Roman fabric reference collection* [NRFRC], were employed as reference material for fabrics and dates (Tomber and Dore, 1998). Reference resources enable and encourage standardisation, which in turn increases reusability. Classification using NRFRC's universally recognised codes in my dataset, distinguished from unsourced fabrics by a bold font, also adheres to the 'Findability' principle for cited works.

The following guidance encapsulates my reasoning for including supplementary categories: to enhance the dataset and create more detailed entries where appropriate.

The Basic Record is the very minimum that is required to produce an informed interpretation of a pottery assemblage but it is recommended that this be enhanced by adding some of the other methods described in the Detailed Record. [...] It is possible that some parts of an assemblage will merit detailed recording, while some can be recorded to a less detailed level (Barclay *et al.*, 2016, Section 2.4).

These categories, primarily featured in CA's pottery datasets from archaeological excavations, include rim diameter [diam] (with units), estimate vessel equivalent [EVEs], minimum number of vessels [MNV], and sherds [Shds] exhibiting soot, burnt food residue [bfr], limescale, other residue [res], decoration or burnishing. Data for the first three categories are derived solely from rim sherds, deemed to merit more detailed recording. Rim sherds inform the 'Form' categories, which often hone the dating.

Residues and surface treatments are frequently interconnected with forms and uses, and whilst 'use' is not included as a category, this information is analysed, and conclusions drawn (where reasonable) in the reporting phase. The capacity for interpretation is the main reason 'use' and similar subjective classifications are absent; objective measurements and physical characteristics form the basis of the dataset, with citable inferences about form and date. By intentionally excluding interpretations that could be subject to revision, it increases the data's reusability, allowing each researcher to draw their own conclusions, even if they choose to refer to research papers associated with the dataset/metadata.

I initially created this dataset as a CA employee, utilising existing, approved structures, including underscored field names. During the recording phase and repurposing the dataset for this assignment, I made enhancements, visual aids, and fixes, which I will elaborate upon in the following paragraphs.

Deviations from Cotswold Archaeology's template

When recording data, Microsoft Access databases are used for excavations, as per CA requirements. However, there are drawbacks associated with Access. I decided Excel was more useful to analyse the data in greater detail, partly due to the filter and pivot table features, and the pros espoused by Graham *et al.* (2020, Section 2.4). My line manager recommended this approach when I first undertook an excavation assemblage. I copied the existing data from the Access document into an Excel template; the completed spreadsheet was eventually imported back into the project's Access database, replacing the initial table.

To this template I added 'Feature' and 'Phase' columns. Information about feature types (ditches, pits, etc.) can be retrieved by creating an Access query using the finished table. However, knowledge of features is required for analysis, and, in my experience, the dataset can still be in flux; thus, it is more practical to incorporate it at the recording stage, despite the manual work and risk of human error. The sequence in which the site phasing was completed was at odds with guidance that "before starting recording ... specialists [should] have the information required [including] updated structural and phasing information" (Barclay *et al.*, 2016, Section 2.4.1). I was responsible for providing the publication officer with pottery date ranges to inform their phasing (Pearson, 2023c). Despite being contrary to the guidance, the reliance on datable wares to phase the site more accurately rendered it necessary.

Advice from my line managers

During the recording and reporting process, CA implemented quality assurance measures. I received expert assistance via my line managers' feedback, advising me to make the following changes to my dataset:

- re-word several vessel descriptions
- combine vessel types where possible and use fewer vessel codes, to improve usability through simplification.
- refer to samian vessels by their Dragendorf or Curle typologies (Webster, 1996), to enable a greater level of comparison with other datasets.

Modifications for this assignment

To construct a suitable dataset for critical commentary, refinement of my spreadsheet was prudent. I removed the 'Pivot table' sheet because it was superfluous to requirement, then edited the 'Comments' column to ensure brevity, relevance, and consistency. The 'Warwickshire Fabric Series' column was deleted. Whilst it is CA's policy to record county-specific codes, the only widely acknowledged Warwickshire series comprises post-Roman wares, rendering it non-applicable for my dataset (Soden and Ratkai, 1998). I would reinstate a regional fabric code column for eligible assemblages. The final significant change was erasing cell fill colour. When analysing assemblages, I find colour-coding crucial data effective for quick reference, such as vessel rims, illustrated sherds, and contexts containing $\geq 500\text{g}$ of pottery. This was explained to the publications officer, who called the information "super useful!" (Pearson, 2023c).

Ideal revisions for future datasets

I would implement two significant amendments. Foremost I would combine the 'Form_specific' and 'Form' columns. The 'Form_specific' category was used for work-in-progress to denote characteristics while recording, before patterns emerged; this helped to identify common traits and create a typology (see the 'Form Codes' sheet; Pearson, 2023a). The 'Form' column was added to enable vessels to be filtered by this unique typology's codes, created solely for writing the pottery report. The 'Form_specific' column was edited for consistency; it is replicated data that should be erased.

Budget-permitting, I would include additional rows for base sherds, similar to those for rim sherds (one per vessel), to record their universal characteristics (columns A-E, I-K and O-P), forms (F-H; using a typology similar to Brudenell (2012, p. 124, Figure 4.3)), diameter (L), EVEs (M), and MNV (N) where possible. Detailed recording of base sherds was not relevant to the post-excavation analysis of the 'Baginton' site, thus no budget was allocated to recording, analysing and reporting. However, this information may be useful to future researchers. It is not always possible to "compile a project archive that includes all material with potential to inform future research and other enquiries", but this

should not deter archaeologists from advocating for the means to adhere to the standards we are professionally and ethically committed to upholding (Barclay *et al.*, 2016, Section 2.1.3).

Conclusions

This essay has explained and justified the structure of my prehistoric and Roman pottery dataset. Standards can insist on basic inclusions, and guidance often enhances the categories/classifications featured in datasets, however there is no “one-size-fits-all-projects” approach or layout. The best practice ultimately necessitates addressing the specific requirements of individual investigations, including time and financial constraints. I adhered to the fundamental (and many advisory) specifications applicable to my assemblage (Barclay, 2016; ClfA, 2023; PCRG, 2010), as well as internal CA guidance. I benefited from the flexibility to devise and refine my own vessel codes, rather than attempting to shoehorn them into an existing typology. Regrettably, this generates potential inconsistencies between similar datasets; nevertheless, the pros outweigh the cons. These discrepancies increase the value of metadata statements, enabling researchers to outline their datasets and draw parallels with others. This work contributes to the existing rationale for categorising data and goes some way towards enhancing our individual and collective professional understanding of ‘advisable’ versus ‘essential’ inclusivity within datasets.

Word Count: 2,112 words

References

Archaeology Data Service [ADS] (2023) *Guides to good practice: About these guidelines*. Available at: <https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/digital-archiving/about-these-guidelines/> (Accessed: 28 December 2023).

Barclay, A., Booth, P., Knight, D., Evans, J., Brown, D. H. and Wood, I. (2016) *A standard for pottery studies in archaeology*. Available at: <https://medievalpottery.org.uk/wp-content/uploads/2019/12/A-Standard-for-Pottery-Studies-in-Archaeology.pdf> (Accessed: 20 December 2023).

Brudenell, M. (2012) *Pots, practices and society: an investigation of patterns and variability in the post-deverel rimbury ceramic tradition of East Anglia*. PhD thesis. University of York. Available at: <https://etheses.whiterose.ac.uk/14230/1/629465.pdf> (Accessed: 04 January 2024).

Chartered Institute for Archaeologists [CIfA] (2023) *Toolkit for Specialist Reporting*. Available at: <https://www.archaeologists.net/reporting-toolkit> (Accessed: 03 January 2024).

Cotswold Archaeology [CA] (2019) *Land west of Coventry Road, Baginton, Warwickshire: Archaeological Evaluation*. Available at: <https://reports.cotswoldarchaeology.co.uk/content/uploads/2019/11/661240-Land-west-of-Coventry-Road-Baginton-Warwickshire-Report1.0-text.pdf> (Accessed: 29 December 2023).

Cotswold Archaeology [CA] (2022) *Prehistoric and Roman Milton Keynes office pottery fabric series (reformatted)*. Cotswold Archaeology [CA]. Unpublished.

Go FAIR (2022) *FAIR Principles*. Available at: <https://www.go-fair.org/fair-principles/> (Accessed: 28 December 2023).

Graham, S., Gupta, N., Smith, J., Angourakis, A., Reinhard, A., Ellenberger, K., Batist, Z., Rivard, J., Marwick, B., Carter, M., Compton, B., Blades, R., Wood, C. and Nobles, G. (2020) *The Open Digital Archaeology Textbook*. Available at: <https://o-date.github.io/draft/book/arranging-and-storing-data-for-the-long-haul-databases.html> (Accessed: 31 December 2023).

Historic England [HE] (2015) *Archaeological and Historic Pottery Production Sites: Guidelines for Best Practice*. Available at: <https://historicengland.org.uk/images-books/publications/archaeological-and-historic-pottery-production-sites/heag019-pottery-production-sites/> (Accessed: 29 December 2023).

Journal of Open Archaeology Data [JOAD] (2023) *Home: About this journal*. Available at: <https://openarchaeologydata.metajnl.com> (Accessed: 28 December 2023).

May, K., Taylor, J. S. and Binding, C. (2023) *Stratigraphic Analysis and The Matrix: connecting and reusing digital records and archives of archaeological investigations*. Available at: <https://intarch.ac.uk/journal/issue61/2/full-text.html> (Accessed: 28 December 2023).

Pearson, L. (2023a) *Assignment Dataset*. Cotswold Archaeology [CA]. Unpublished.

Pearson, L. (2023b) *Metadata for Assignment Dataset*. Cotswold Archaeology [CA]. Unpublished.

Pearson, L. (2023c) Email to Grace Griffith, 4 August.

Prehistoric Ceramics Research Group [PCRG] (2010) *The Study of Prehistoric Pottery: General policies and guidelines for analysis and publication*. Occasional Papers 1 and 2. Available at: <https://09edd9eb-e245-4890-985c->

e2188bb4bb42.filesusr.com/ugd/93ae8c_49ff5068d62241f0a662881502173bc9.pdf (Accessed: 20 December 2023).

Soden, I. and Ratkai, S. (1998) *Warwickshire Medieval and Post Medieval Pottery Type Series*. Northamptonshire Archaeology for Warwickshire Museum. Unpublished.

Tomber, R. and Dore, J. (1998) *The national Roman fabric reference collection: A handbook*. London: Museum of London Archaeological Service.

Webster, P. (1996) *Roman Samian pottery in Britain*. Practical Handbook in Archaeology 13. York: Council for British Archaeology.