

Assignment 2: The Archaeological Dataset from Chapelfield Road, Guyhirn

In December 2019, an archaeological excavation took place along Chapelfield Road in Guyhirn, Cambridgeshire. The investigation revealed a series of linear features as well as pits, potentially related to salting, and two post-built structures (Fig.1) (Newton 2020; Newton *et al* 2020; Peachey 2020). A total of 132 sherds of pottery, all Roman in date, were recovered during the excavation. The pottery was subsequently assessed following *A Standard for Pottery Studies in Archaeology* (Barclay *et al* 2016). Prior to the insertion of data into any database or dataset, a process of abstraction of information has occurred (Huggett 2018); entailing the selection of specific entities that are to be represented within the data structure, which in this circumstance were largely dictated by the aforementioned standards and guidelines. The data was entered into a Microsoft Excel 'flat-file' spreadsheet, which forms part of the site archive (Appendix 1 and 2).

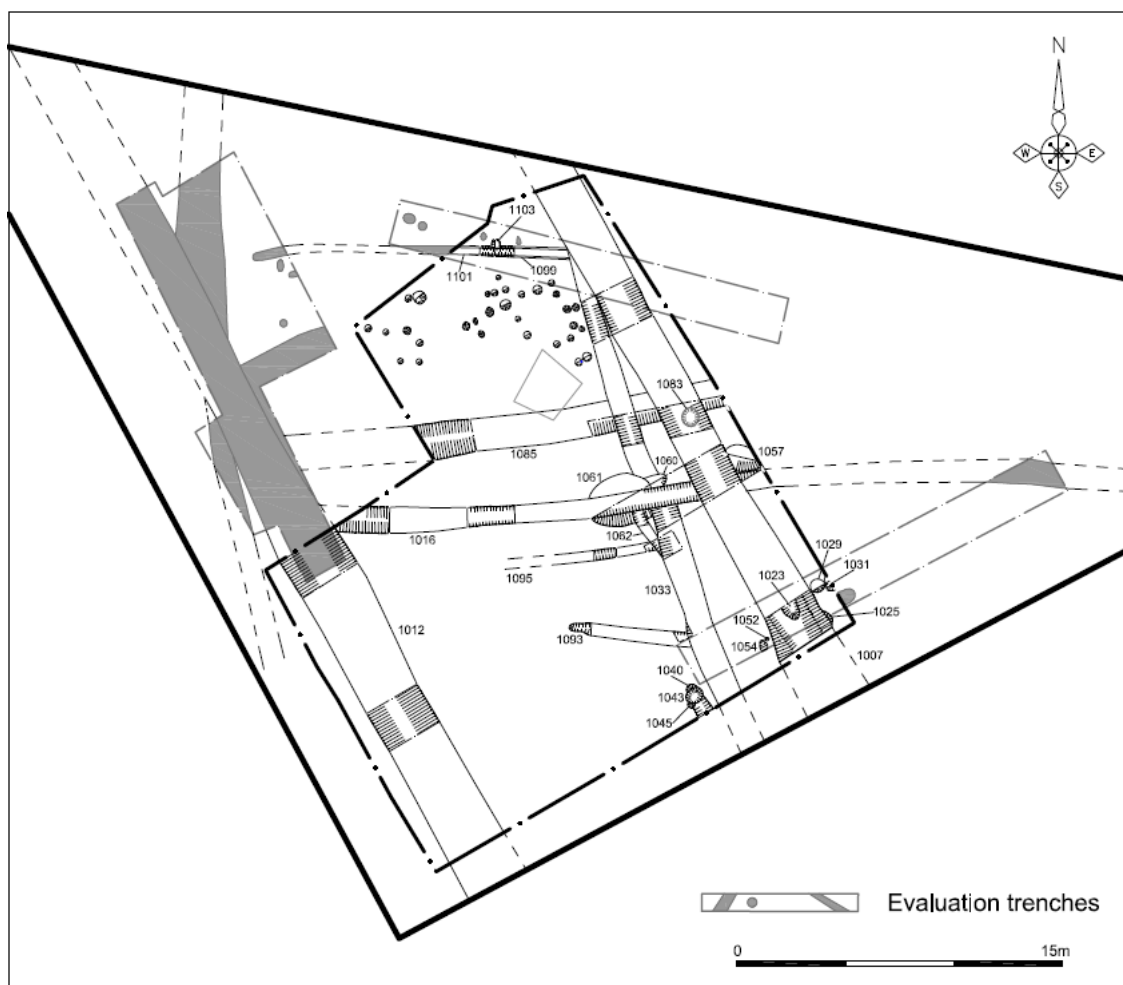


Fig.1: Excavation plan from Chapelfield Road, Guyhirn (Newton *et al* 2020)

The spreadsheet forms the spot-date records, which represents a basic but vital minimum standard of ceramic recording (Marshall and Seeley 2018; Barclay *et al* 2016: 29). As specified within the guidelines by Barclay *et al* (2016: 13), the purpose of the basic pottery analysis is to quantify and characterise assemblages in sufficient detail to facilitate the production of a meaningful report. Not only can this data potentially contribute to broader synthetic analyses of pottery, but can also be used to further our understanding of function, social and economic status, and chronology of an archaeological site or feature. As noted by Gattiglia (2018: 2), databases and datasets "must be as comprehensive as possible and must lead to better analysis, publication, and archiving".

Metadata and Archiving

A metadata statement for the spreadsheet, in line with guidelines by the Archaeological Data Service (ADS 2021), is provided within Appendix 3; this provides information on the resource, allowing the viewer to place the dataset within its wider context and to optimise re-use. While the attached metadata describes the dataset and its structure, a brief overview of the spreadsheet structure is provided below. Long-term survival and access to the data is integral, especially in the context of 'preservation by record', with the data providing ongoing opportunities for research. The pottery dataset, however, which is saved within the '.xls' format, is not recommended for long-term preservation by the Archaeological Data Service (ADS 2021). The flat-file spreadsheet, for the purpose of long-term preservation, should have been saved and deposited within the '.csv' format.

Dataset Structure and Categorisation

The spreadsheet is comprised of two sheets or tables. Both tables use three primary columns to detail the feature, context, and slot from which the pottery was retrieved. These columns, acting as identifiers, are integral to any post-excavation pottery database and subsequent analysis; the archaeological material recovered, in this case pottery, is intrinsically linked to its intra-site location and stratigraphic context.

The primary table details the quantification of the assemblage (Appendix 1). The table quantifies the overall amount of pottery, by sherd count and weight, recovered from each context and describes to which phase or date it belongs as a whole. Subsequent columns within the table treat the fabric types as attributes, with the sherd count and quantity of each fabric type from individual stratigraphic context being recorded. With regards to each diagnostic sherd, the secondary table (Appendix 2) records attributes such as vessel fabric, form, and interpreted date, along with other information on its condition such as abrasion or sooting. In addition to the 'vessel type' column, many of the diagnostic sherds have a reference within the column, 'form comparison'; this allows anyone, if reassessing the pottery assemblage, to re-evaluate the assessment in the future.

Dataset Structure and Resultant Analysis

The spreadsheet, during the reporting stage, was analysed to form an interpretation of the pottery assemblage and narrative of the site as a whole. The spot-date table (Appendix 1) is structured in such a way so that the sum totals of the sherd count and weight can be easily calculated for each fabric type, individual context, and for the entire pottery assemblage. This is done by using the SUM function within Microsoft Excel. Such data, although seemingly basic, is fundamental and informs our interpretations of pottery provenance, trade dynamics, and the intensity of consumption on site. Indeed, using the totals of fabric type within the dataset, we can discern that the pottery principally originated from major industries in and around the Fenland region, which were partially supplemented by wares from the Lower Nene Valley (Peachey 2020).

Largely utilising the diagnostic table (Appendix 2), it was also identified that there was a particular focus on the use of larger vessels that were associated with industrial activity. This information was principally yielded from the 'comments' column, which detailed the key attributes of the sherds. This observed pattern of consumption was decidedly utilitarian and reflected the use of the site for salt-making. Additional analysis of the sum totals noted above supported this characterisation of the assemblage, with a notable paucity of the domestic fine wares being recorded that were widely circulated during this period – perhaps reflecting the limited demands for domestic wares on an industrial site (Peachey 2020).

Qualitative Data: Limitations

Despite the evident potential for the dataset to facilitate meaningful interpretation, the format used and its structural design and classification has considerable limitations. The 'spot date' column within the quantification table as well as the 'vessel date' column within the diagnostic table (Appendix 2), for example, record the interpreted date in text form rather than as a numerical range. This, in addition to increasing the potential for inconsistency during input, decreases the analytical re-use potential. Analysis using the Plotdate technique, for instance, necessitates researchers converting such textual dates into numerical values in order to conduct their analysis – which refines chronological ranges of fabric and form types (Darling and Precious 2015: 8). Furthermore, within the two columns some of the dates recorded are prefixed with '?', denoting an uncertain interpretation – a discussion of such 'fuzzy' values, and other qualitative data, is discussed further below.

As highlighted by Banning (2020: 41), the categorisation and structuring of pottery databases presents a unique challenge. This is primarily because pottery analysis involves specialised 'lexical units' symbolising visual and qualitative characteristics such as decoration, the specifics of abrasion and sooting, and segmented sherd forms. Within archaeology, a major issue with the processing of digital data is the necessity of 'absolute estimations' in place of the subjective interpretations that result from real-world contexts (Gattiglia 2018). Within the diagnostic table (Appendix 2), this problem has been resolved by listing these qualitative details within a 'comments' column. As a result, however, the analytical potential of this data is limited and the field forms more of a catalogue. Questions related to the percentages of sherds with combed decoration, for instance, cannot be easily or quickly answered. This is further complicated by the fact that a pot may exhibit more than one decoration type, or have a 'hybrid' decoration, reflecting the complexity of rules-based categorisation of visual phenomena. Satisfactorily resolving this issue would involve a somewhat excessive process of defining mutually exclusive and unambiguous definitions of, for example, pottery decoration.

Using a relational database, it has been demonstrated that these issues can be resolved by using 'fuzzy' classifications and queries (Niccolluci *et al* 2001; Hermon and Niccolluci 2002; Niccolluci and Hermon 2015;). These classifications and queries are more forgiving and would show artefacts and values that are *similar* to those that were requested within the query. There are nevertheless inherent issues with 'fuzzy' databases, including the expertise required at the design stage and the fact that most databases do not facilitate this function, in addition to spelling errors and the potential for data omission (Banning 2020).

Extracting Meaningful Data

Using the charts and graphs function within Microsoft Excel, a range of data visualisations can be produced, such as a pie-chart that visually represents the proportions of certain fabric types by sherd count within the assemblage (Fig.2). This function, and through analytical operations via the formula function, has the capability to answer research questions within the dataset – be it related to the distribution of pottery by quantity in each feature type, or to the distribution of fabric types across each segment of ditch [1007]. Nonetheless, to extract these data representations from the spreadsheet is a relatively laborious task, often involving the production of additional tables that contain the specific data of interest. Alternatively, if the ceramic data was input into an appropriate arrangement of related tables within a relational database, such operations could be achieved without the necessity of rearranging or creating a new set of spreadsheet tables. Indeed, the utilisation of basic Structured Query Language (SQL) select queries could retrieve all the pertinent records meeting the criteria, in order to answer a predetermined research question (Lock 2003). Such a relational model would have greater analytical flexibility, meaning that the pottery data could be analysed based on

any of its multitude of properties (Keller 2009; Jones and Hurley 2011). As stated aptly by Andresen and Madsen (1996), flat-files such as this one are far too simple of structures to accommodate the immense complexity, diversity, and interrelatedness of archaeological ‘reality’.

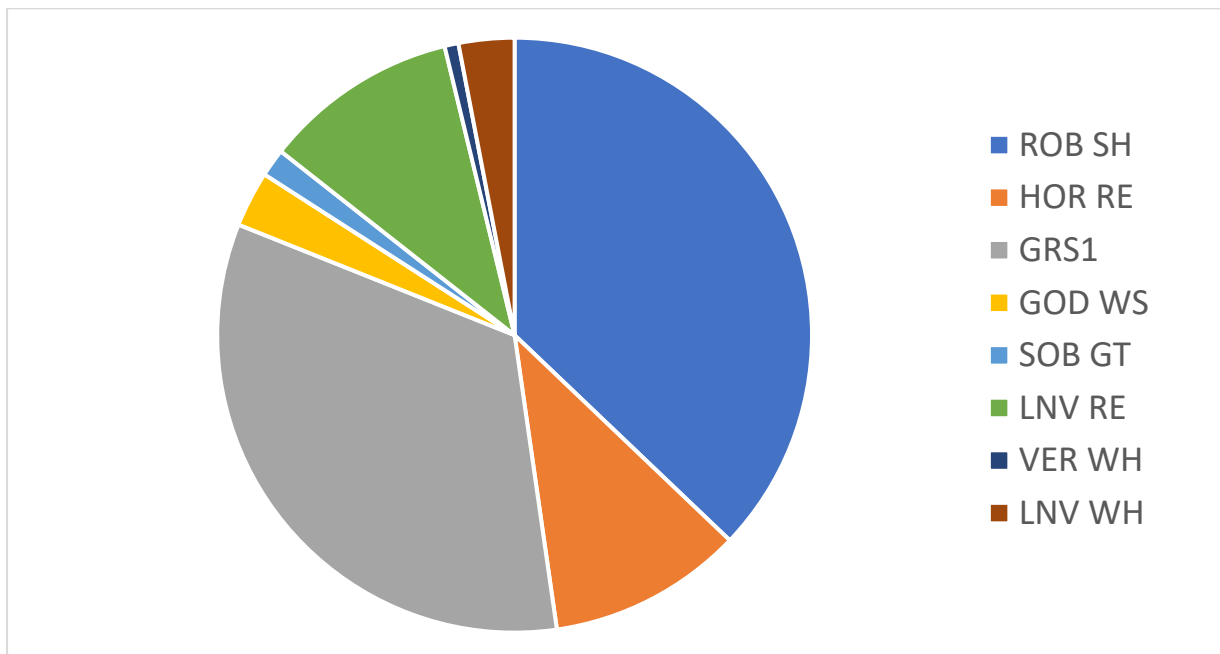


Fig. 2: Quantification of the pottery assemblage by fabric type and sherd count.

Relational Database: Alternative Format

During post-excavation analysis, a series of separate and ‘unrelated’ spreadsheets were produced; digitally storing data associated with the finds recovered, the palaeoenvironmental remains, and the context data recorded on site. Extracting meaningful interpretations from these disparate datasets, like noted above, is a lengthy process. The utilisation of a single relational database for post-excavation would facilitate more flexible and comprehensive analysis. The shortcomings of flat-file databases are apparent when one considers that excavations recover a wide range of *entities*, that have their own *attributes*, many of which have *relationships* between one another. Unlike the flat-files used for the Guyhirn project, relational databases use formal relationships or joins to link discrete tables – reflecting meaningful relationships between real entities (Keller 2009). This would mean that a query could be asked to, for example, show all ditches with a diameter of more than 0.50m that contained saltern briquetage and storage jars. This query could, in theory, be used to identify ditches that may potentially be associated with salt-making and used to direct the brine to working areas.

In addition to the increased research potential of the data when utilising a relational database, this highlights the inefficiency of the flat-file structure. The flat-file structure of the pottery dataset necessitates the input of repeated information, which is described elsewhere, creating extreme redundancy. For example, it would not be necessary to describe the feature type within both the tables specifically used for pottery analysis. Alternatively, as the pottery and contexts are differing entity types with differing files, we can simply make a relation between them within a relational database.

Two entirely redundant columns are also present within the diagnostic table (Appendix 2), those which denote vessel number (V.No.) and illustration reference (?I). These columns contain no data and are wholly unnecessary as none of the elements within the assemblage were given a vessel number, nor an illustration within the resultant report. Although this would have limited impact due to the small

size of the dataset, the presence of such redundant data within a scaled-up version of the spreadsheet would increase the space required for archiving, and potentially increase storage costs.

Conclusion

Despite the limitations of the spreadsheet, the flat-file dataset utilised is notably easy to set up and the software requires relatively little planning or skill (Schloean 2001), meaning it is relatively inexpensive within the context of commercial archaeology. Relational database structures, although superior from an analytical perspective, involve the input of relatively skilled practitioners and require extensive programmes of database design to yield the analytical benefits. Furthermore, to attain the ranging analytical benefits, a singular database for the variety of entities generated during an archaeological excavation would need to be designed and utilised. This endeavour, for such a relatively small excavation, seems somewhat disproportionate. It is also worth considering that if such an investment in training and time was decided, the investment may be better focused on 'emerging' database technologies such as NoSQL graph databases, with its benefits for heavily inter-connected data.

Irrespective of the limitations of the flat-file structure, the spreadsheet fulfils the essential requirements for the creation of a dataset related to the 'spot-date' recording and the analytical assessment of diagnostic material, as specified within the standards and guidelines by Barclay *et al* (2016). Overall, the spot-date and diagnostic tables not only form a long-lasting catalogue of the pottery data retrieved during the excavation, but are also an analytically functional source of archaeological information.

References

- Andersen, J. and Madsen, T. (1996) 'IDEA – the Integrated Database for Excavation Analysis'. in Kamermans, H and Fennema, K (eds.) *Interfacing the Past: Computer Applications and Quantitative Methods in Archaeology*. Leiden: Analecta Praehistorica Leidensia, pp. 3-14.
- Archaeological Data Service (ADS) (2021) *Guidelines for Depositors*. Available at: <https://archaeologydataservice.ac.uk/advice/Downloads.xhtml> (Accessed: 10 November 2021).
- Banning, E. (2020) 'Compilations: Designing and Using Archaeological Databases' in: Banning, E (ed.) *The Archaeologist's Laboratory: The Analysis of Archaeological Evidence*. Cham: Springer.
- Barclay, A. *et al.*, (2016) *A Standard for Pottery Studies in Archaeology*. Prehistoric Ceramics Research Group / Study Group for Roman Pottery / Medieval Pottery Research Group / Historic England.
- Darling, M. and Precious, B. (2015) *A Corpus of Roman Pottery from Lincoln*. Oxford: Oxbow Books.
- Gattiglia, G. (2018) 'Databases in Archaeology' in: López Varela, S and Thomas, J (eds.) *The Encyclopedia of Archaeological Sciences*. Hoboken: John Wiley & Sons Ltd.
- Hermon, S. and Niccolucci, F. (2002) 'Estimating Subjectivity of Typologists and Typological Classification with Fuzzy Logic'. *Archaeologia e Calcolatori*, 13, pp. 217-232.
- Huggett, J. (2018) 'Computer Applications in Archaeology' in López Varela, S and Thomas, J (eds.) *The Encyclopedia of Archaeological Sciences*. Hoboken: John Wiley & Sons Ltd.
- Jones, E. L. and Hurley, D. A. (2011) 'Relational databases and zooarchaeology education'. *The SAA Archaeological Record*, 11(1), pp. 19-21.

- Keller, A. (2009) 'In Defense of the Database'. *The SAA Archaeological Record*, 9(5), pp. 26-32.
- Lock, G. (2003) *Using Computers in Archaeology: Towards Visual Pasts*. New York: Routledge.
- Marshall, M. and Seeley, F. (2018) 'From the Spreadsheet to the Table? Using 'spot-dating' level pottery records from Roman London to explore functional trends among open vessel forms'. *Internet Archaeology*, 50.
- Newton, A. (2020). *Land North of 52 Chapelfield Road, Guyhirn, Cambridgeshire: An Archaeological Excavation Research Archive Report*, Archaeological Solutions report. Unpublished
- Newton, A., Summers, J. and Peachey, A. (2020) *Romano-British Saltmaking in Guyhirn: Archaeological Excavation at Chapelfield Road*, Archaeological Solutions report. Unpublished
- Niccolucci, F., Andrea, A. and Crescioli, M. (2001) 'Archaeological Applications of Fuzzy Databases'. *BAR International Series*, 931, pp. 107-116.
- Niccolucci, F. and Hermon, S. (2015) 'Time, Chronology and Classification'. *Mathematics and Archaeology*, pp. 257-270.
- Peachey, A. (2020) 'The Roman Pottery' in Newton, A (ed.) *Land North of 52 Chapelfield Road, Guyhirn, Cambridgeshire: An Archaeological Excavation Research Archive Report*. Archaeological Solutions, Unpublished. pp. 20-22.
- Schloen, D. J. (2001) 'Archaeological Data Models and Web Publication Using XML'. *Computers and the Humanities*, 33, pp. 123-152.

Appendix 1

						ROB SH		HOR RE		GRS1		GOD WS		SOB GT		LNV RE		VER WH		LNV WH		
F	L	Seg	Description	Spot Date	F	W	F	W	F	W	F	W	F	W	F	W	F	W	F	W		
1007	1009		Fill of Ditch	Roman	1	8					1	8										
1007	1009	B	Fill of Ditch	Roman	2	9					2	9										
1007	1010	A	Fill of Ditch	2nd C AD	8	102	2	56	4	25						1	7	1	14			
1007	1010	B	Fill of Ditch	Roman	2	32	1	17			1	15										
1007	1010	C	Fill of Ditch	Roman	2	20					2	20										
1007	1010	D	Fill of Ditch	Mid 2nd-3rd C AD	7	163					4	106				2	23		1	34		
1007	1011	A	Fill of Ditch	2nd-3rd C AD	11	154	2	75			7	63				2	16					
1007	1011	C	Fill of Ditch	Roman	3	27	2	21			1	6										
1007	1011	D	Fill of Ditch	2nd C AD	49	241	27		2	43	14	150	1	3		4	36		1	9		
1012	1015	A	Fill of Ditch	Roman	1	5									1	5						
1016	1020		Fill of Ditch	?2nd C AD	1	52										1	52					
1016	1021	C	Fill of Ditch	Mid-Late 2nd C AD	5	151					4	67							1	84		
1016	1050		Fill of Ditch	Mid/Late 2nd-Early 3rd C AD	4	121	1	49					3	72								
1016	1050	A	Fill of Ditch	Roman	3	257	3	257														
1033	1034		Fill of Ditch	Roman	1	32					1	32										
1033	1036	A	Fill of Ditch	Roman	2	35			2	35												
1033	1037		Fill of Ditch	Late 2nd-3rd C AD	7	241	1	46	2	24	1	110				3	61					
1033	1037	A	Fill of Ditch	Roman	4	59			1	45	2	10				1	4					
1033	1037	D	Fill of Ditch	Mid 2nd-4th C AD	3	103					1	29			1	53			1	21		
1033	1090	B	Fill of Ditch	Late 1st-4th C AD	3	263			3	263												
1040	1041		Fill of Ditch Terminus	Roman	2	27	2	27														
1054	1055		Fill of Pit	Roman	1	17					1	17										
1060	1064	C	Fill of Pit	Roman	1	16					1	16										
1061	1070		Fill of Pit	(M2? +) Roman	2	240	2	240														
1085	1086	A	Fill of Ditch	Roman	7	151	6	132			1	19										
					132	2526	49	920	14	435	44	677	4	75	2	58	14	199	1	14	4	148

Appendix 2

F	L	Seg	Desc	Fabric	Vessel Type	Form Comparison	d	R.EVE	V.No.	I?	Vessel Date	Comments
1007	1010	A	Fill of Ditch	VER WH	Flagon	\	\	\	\	\	?2nd C AD	broad 4-rib strap handle
1007	1010	D	Fill of Ditch	GRS1	Jar	\	11	0.1	\	\	Roman	everted bead rim
1007	1011	D	Fill of Ditch	LVN WH	Flagon	Perrin 1996: fig.83.90	3	0.25	\	\	L1-L2	slightly everted ring-necked flagon with a small external bead
1012	1015	A	Fill of Ditch	SOB GT	Storage Jar	\	\	\	\	\	Roman	thick body sherd with combed exterior
1016	1020		Fill of Ditch	LVN RE	Jar	?Perrin 1996: fig.85.132	\	\	\	\	?2nd C AD	neck and body sherd of narrow-neck jar; with neck cordon and shoulder band decorated with burnished oblique lines, and a plain burnished body inbetween
1016	1021	C	Fill of Ditch	LVN WH	Flagon	Perrin 1996: fig.86.160	3	1	\	\	M-L2	strongly everted ring-necked flagon with a medium external bead
1016	1050		Fill of Ditch	GOD WS	Jar	Evans 2003: fig.25.10-11)	8	0.1	\	\	M-L2/E3	ovoid jar with a stubby lid-seated rim, with a single groove on top
1016	1050		Fill of Ditch	ROB SH	Storage Jar	\	20	0.05	\	\	Roman	robust everted plain rim of storage jar
1033	1037		Fill of Ditch	LVN RE	Dish	Perrin 1999: Vessels 95-7	9	0.1	\	\	L2-3	overhanging rim with internal groove, imitating samina form 36
1033	1090	B	Fill of Ditch	HOR RE	Jar	Evans 2017: type J10.4	15	0.25	\	\	L1-4	Large jar with strongly everted plain rim and offset at base of neck; heavy soot residue on inside of rim and neck
1061	1070		Fill of Pit	ROB SH	Jar	Perrin 1999: vessel 463	14	0.15	\	\	(?M2/Late?) Roman	Large jar with simple everted plain rim on sinuous concave neck; soot on exterior

Appendix 3

File Name	Title	Description	Creator			Copyright holder			Period of Creation	
			First Name	Last Name	Organisation	First Name	Last Name	Organisation	Start Date	End Date
P8210_ECB6074_Land_North_of_52_Chapelfield_Road,_Guyhirn_EXC_Pot_data.xls	Pottery data for excavations on land north of 52 Chapel Road in Guyhirn, Cambridgeshire.	Basic record of the pottery retrieved during an excavation from land north of 52 Chapel Road in Guyhirn, with quantification and diagnostic tables.	Andrew	Peachey	Archaeological Solutions Ltd	Andrew	Peachey	Archaeological Solutions Ltd	29/09/2017	06/10/2017

Language	Software used	Software version	Supporting documentation file name(s)
English	Microsoft Excel	Microsoft Excel 97-2003	Fabric_Descriptions.pdf
			Comparison_Bibliography.pdf

Sheet Name	Sheet Description	Number of rows	Field name	Field description
Quantification	Quantification of the pottery	28	F	Unique context of feature cut..
			L	Unique context of deposit or layer.
			Seg	Segment/slot of excavated feature.
			Description	Description of feature type (i.e. ditch/pit/posthole).
			Spot_Date	Spot date of the context.
			ROB_SH	Romano-British shell tempered ware, by sherd count and weight (g).
			HOR_RE	Horningsea reduced ware, by sherd count and weight (g).
			GRS1	Sandy grey ware 1, by sherd count and weight (g).
			GOD_WS	Godmanchester white-slipped ware, by sherd count and weight (g).
			SOB_GT	Southern British grog-tempered ware, by sherd count and weight (g).
			LNV_RE	Lower Nene Valley reduced ware, by sherd count and weight (g).
			VER_WH	Verulamium region white ware, by sherd count and weight (g).
			LNV_RH	Lower Nene Valley white/partchment ware, by sherd count and weight (g).
Diagnostic	Key attributes of the diagnostic sherds	12	F	Unique context of feature cut.
			L	Unique context of deposit or layer.
			Seg	Segment/slot of excavated feature.
			Description	Description of feature type (i.e. ditch/pit/posthole).
			Fabric	Pottery fabric code.
			Vessel_Type	Vessel form type (i.e. flagon/jar).
			Form_Comparison	Reference material used for form comparison, with figure and page numbers.
			D	Diameter of the rim.
			R.EVE	Rim estimated vessel equivalent.
			V.No	Vessel number
			I	In text illustration of the sherd
			Vessel_Date	Assigned date of the vessel.
			Comments	Description of diagnostic sherd (i.e. abrasion/sooting/decoration)