Dingwall, L., S. Exon, V. Gaffney, S. Laflin and M. van Leusen (eds.) 1999. Archaeology in the Age of the Internet. CAA97. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 25th Anniversary Conference, University of Birmingham, April 1997 (BAR International Series 750, CD-ROM). Archaeopress, Oxford.

The Effect of Computerisation on Pottery Recording

P. D. Rauxloh and R. P. Symonds

Abstract

This paper discusses the effect which the implementation of a new relational database system has had on the way Roman pottery is recorded, analysed and interpreted. Instead of a technical description, it considers some of the issues which the implementation has raised, and in particular the improvement it provided for the validation of ceramic data. The paper further differentiates itself from a straight technical discussion by being the joint product of the system's developer, and a Roman ceramic specialist who uses it.

1 Introduction

This paper discusses the effect which the implementation of a relational database[1] has had on the way Roman pottery is recorded, analysed and interpreted. It does not offer an exhaustive description of "the system we created to record pottery", but rather considers some of the issues which the implementation of our system has raised, and in particular the improvement it provided for the validation of ceramic data. The paper differentiates itself from a straight technical discussion by being the joint product of the system's developer, (Rauxloh) and a Roman ceramic specialist who uses it (Symonds).

For this reason, the paper consists of two parts. The first concentrates on just two ways in which the new system differed from the old, and considers how these have affected user interaction.

1. It is a relational system
2. It offers improved validation.

The second part presents some reaction from the user's perspective concentrating on how the systems has affected the recording of Roman pottery.

These two sections are preceded by a short introduction to the particular environment in which the system was implemented, in order to illustrate (a) that the constraints imposed by that environment demand a certain formalisation in the way archaeological data is recorded, and (b) that the disciplines imposed by such a formalisation are not synonymous with a stagnating methodology or an appreciation that the nature of archaeological research is a reflexive dialectic one. These matters are taken up in the discussion.

2 The environment

The Museum of London Archaeology Service, (MoLAS) is an archaeological contractor which carries out two major types of work. Firstly, it undertakes desk top assessments, on-site evaluations and full excavations of sites threatened by development and, since the instigation of competitive tendering in the early 1990's, must compete with other contractors for such developer-funded work. Secondly, post-excavation projects are undertaken, which concentrate both on singularly important sites, and on thematic research projects which integrate information from a wide variety of sites in an area.

The latter projects are of crucial importance to the general advance of our knowledge of London and its past, since thematic research projects (mainly funded by English Heritage) are the mechanism by which the importance of individually unspectacular sites is realised. Herein also lies a clear example of why formalised structures for recording information are so important at MoLAS and indeed to the discipline, since whatever claims may be made for the value of the individual perspective in archaeological interpretation, archaeological data has communal value and collateral significance, and a number of years may pass between the excavation of a site and the interpretation of its data within a London-wide research project.

The data systems devised at MoLAS, and more generally those used by archaeological units working in the commercial sector, are necessarily compromised in the application of certain idealised theoretical approaches, since they have the practical problem of maintaining control of very large multi-period data sets and the many individuals that create them. This compromise is manifest in the creation of certain disciplines which by definition constrain individual expression and impose an interpretation.

When this situation is presented in the negative it is called stagnation, yet it is argued that the benefits of fluidity in data systems espoused by some authors (Andresen and Madsen 1996), do not fully reckon with the value and role that stasis plays in the creation of valid data sets. A system optimised for the capture of ever more representative variables of a certain entity rather than one tuned to capture data using an existing set, (like MoLAS's), must concede that its ultimate purpose is the generation of new constraints. That is the purpose of identifying a useful new variable must be to then capture that data in all other instances of the entity where the variable exists to be measured and the research objectives require it. Without such a policy each data set becomes unavoidably contingent on the idiosyncracies of the user; did Dr.X *remember* to test all amphorae for residues? Clearly the identification of new variables is an ongoing and valuable process, yet no less so than the exploitation of such acts. This discussion is continued elsewhere, (Rauxloh In Prep).

MoLAS's relational database system captures all attribute data generated from excavations carried out in the City and Greater London. This data constitutes part of the site archive and as such augments physical paper records, digital spatial coverages photographs and so fourth. In conjunction with consideration of research objectives which the excavation may have caused to modify, this material is then assessed in order to make recommendations about further analytical work. The preparation of data up to this level of 'assessment of potential for analysis', follows guidelines set down by English Heritage, for the management of archaeological projects (English Heritage 1991).

2.1 It is a relational system

A relational database system is no more than a collection of tables that each hold a generic data set, (coins, pottery etc.), and within which exist common fields that enable those tables to be joined together and act as a single table at the time and in the manner of a user's choosing. The common fields in the archaeological tables are sitecode and context, the latter being the term given to the stratigraphically discrete excavated unit. More explicit definitions of relational systems abound, (e.g. Koch and Muller 1993, chapter 1) but our interest here is what this relational environment meant for the archaeologists at MoLAS.

The most immediate benefit of this relational ability is that users in the different areas of expertise, (primarily the finds, field and environmental sections) are able to view and query on line any of the data generated by their colleagues. As the user chooses each form, the sitecode and context number of the record the user had been examining are bound into global variables that are read as query criteria by the called form and executed directly the latter appears. Thus one can observe the coin data for a certain context, go to the Roman pottery form and be automatically presented with the pottery of that date found in the same context.

The new system is an improvement on the older system where such options were practically obstructed by procedural syntax and file privileges, and conceptually hindered by misplaced notions of specialist sovereignty. Users are now able to check on the existence/progress of other specialists data sets which they wish to relate to their own. A product of this is that it made differential section and specialist programming and work rates more apparent, and in its creation caused the notion of data protectionism to be faced and overcome, a notion which was less entrenched than one may have expected.

As well as the forms that capture actual data, the system also has forms which enable users to access and configure the classification code lists which validate entries into data tables through the mechanism of foreign key constraints. If a data field has a foreign key constraint, it means that any entry into that field must appear in the field of the classification table referenced by the constraint. These forms deliver the job of maintaining those lists into the hands of the user for the first time.

The first advantage of this is that new codes can be created and added to the list without recourse to an IT specialist, and since the foreign key linkage is dynamic, the new code is immediately acceptable in the related data table. More subtle advantages include the ability to add further classification variables that enable different views of the data to be generated of interest to different groups and different research goals. For example Roman pottery fabrics can currently be classified and thus viewed, summarised or aggregated by recording their;

1. Early Date.
2. Late date.
3. General ware type : Reduced, Oxidised, Fine etc.
4. Source : Romano British, Imported
5. Origin : Central Gaul, Eastern Gaul etc.

This openness to different coding system encourages different ways of perceiving the ceramic resource, both for the ceramic specialists themselves, and among non-specialists. Previously the complexity of the raw data itself - in this case the Roman fabric codes - has been a bastion of inter-specialist data boundaries, the interested parties literally speaking different languages. By expressing that raw pottery data in more comprehensible terms, the field archaeologist is encouraged to interact with it and drawn in to consider the ceramic material more carefully.

Yet as the right of the specialist to encode their data as they please was granted by this structure, the obligation to exercise that right in a sensible manner necessarily followed. In terms of the system's maintenance, the consequence of abuse would be myriad classifications for the same phenomenon at varying levels of granularity (the infamous 'lumper' and 'splitter'situation), and in general the growth of individualistic schemes of recording that can mature into archives of stunning clarity to single persons.

To aid the sensible use of such a structure, yet not to impede the generation of new ideas, the system both encourages and obliges individuals to discuss things. Users in smaller groups such as the four persons recording Roman pottery, are encouraged to seek consensus with each other before a code is created or updated, since the existence of foreign key constraints and the accessibility of the classification system make any such changes instant and explicit. While the primary motivation for seeking such consensus remains the need to maintain the uniformity of the recording system, it is clearly also encouraged by the fact that they work in a team and value each others opinion. Users in larger groups however - where the risk of misuse is greater - are *required* to reach some level of consensus about any new changes, since only certain individuals in the group are able to alter the group's classification tables.

The increasing ease with which specialist data can be accessed also has consequences for the way in which that data is expressed. Examples include *overall* age and sex fields in the relevant human bone tables to summarise the various osteological indicators pertaining to these values, or the addition of various grouping fields to ceramic fabric references lists as described above.

Finally it was possible to create forms such as a dating sheet, which groups and summarises all datable material in a single place so as to aid its general comprehension by contextualising the separate strands. The dating sheet is populated with summary dating information for a given context by background triggers which fire at the point of saving any datable material. This summary data records the fact that for a given assemblage, there exists for example, coin and worked stone sub-assemblages which may be able to provide dates. The specialists concerned will then visit the sheet and record their considered date for these sub-assemblages, and indeed the overall date of the whole assemblage.

For the specialist, the dating sheet also uses the system's relational ability to provide the facility of calculating default dates for sub-assemblages. These are generated by locating the narrowest date range either directly from those specified by the specialist in their basic data form, (e.g. coins or glass) or indirectly by interrogating the related code lists which hold dates, and deriving the narrowest range from them. The latter operation is at its most complex in the case of Roman pottery where fabrics, forms and decorations codes may all be independently dated and thus require that the narrowest range be first derived for each *type* before that for the entire context may be calculated (see the discussion of soft coding below).

While the system may alert users to potential refinements in the dating of actual artefacts, when for example a type is seen to be frequent appearing in association with markedly later material, the calculation of default dates at the very least makes users confront their own pre-conceptions since on occasions they will be surprised.

2.2 It offers improved validation

Improving validation is a basic requirement of any new system, and for the purposes of this discussion two major types of validation can be identified. These have been christened Hard and Soft coding. The respective connotations of absolutism and of pliability are appropriate, and of relevance to the comments regarding the static and fluid nature of data systems mentioned previously.

Hard coding simple means that level of validation where there is no uncertainty about what constitutes acceptable data. Validation of this type is imposed by basic object definitions, (a 5 digit numeric field does not accept characters or numbers greater than 99999) and by object relationships, (in the relational paradigm foreign key constraints ensure that a referenced key may not be removed). Hard coding also describes validation by acceptable code. Thus all data going into MoLAS's system requires a sitecode, and this code must appear in a master sitecode list. The majority of validation that occurs is of this type.

Soft coding is employed when validating material of a more volatile nature. Like any constraint, it is triggered by contravention of a data object's rules, yet the difference to hard coding is how the machine reacts. This is best illustrated through the example of the validation of Roman ceramic types.

At MoLAS pottery is recorded at the type-within-context level of granularity. A type is defined by its fabric, form and decoration, and a type may only be referenced once in each context. Thus the primary key of the pottery table is derived from a combination of the sitecode,context,fabric,form and decoration. For the Roman period, fabrics, forms and decorations can frequently be independently dated and the reference tables that contain these lists therefore hold a date range. Thus a form may be able to date a type on its own without reference to the fabric in which it has been made.

Soft coding stems from the fact that we are able to contextualise that form, by making reference to the fabric in which it appears, and the decoration that adorns it. Since all three have date ranges, validation exists to ensure that a type, may only consist of such a combination of these three identifiers that *on the basis of current knowledge* of ceramic chronologies could have actually existed.

If the user makes an attempt to pair a fabric with a form where their respective date ranges are not-contemporary[2], the system responds with a message to this effect. At this point the user is able to visit the fabric and/or form classification sheets and modify the dates if this is appropriate.

Both the fabric and form codes are acceptable in the hard coded sense, (i.e. they both exist as valid referable codes), yet their rejection as a pair is based on a softer, movable and in general less absolute entity; a date range. It is exactly by this mechanism that ceramic date ranges have evolved for more than a century. That is pottery being found in situations where it is contextualised by some other datable material, whether it be an historical reference to an occupation date, a securely dated coin hoard or other types of dated pottery. Even objects that appear intrinsically datable, such as *samian* with a makers stamp or coins with a date, rely on sets of relationships such as historical references to ceramic production by X at kiln site Y and archaeological ones between contexts and the use of said kiln, or those between coin minting policies and circulation rates.

The function of soft coding is to acknowledge the highly theoretical nature of date ranges, to alert users to their possible further revision, and to make the process of doing so easier. This is an aid to the specialist and other instances suitable for such validation exist, yet like computers in general, it provides a poor imitation of intellect which has to deal with the fact that most archaeological data is theoretically determined and will only ever be properly 'soft coded' in the head.

3 A users's perspective

The above-described validation system is clearly much more complex than any previous data recording system used at MoLAS, and it represents a considerable advance in the methodology of Roman pottery recording. It cannot validate everything: it does not actually examine the sherds and correct the identifications if they are wrong. Nor can it know if one has typed a valid code which happens to be the wrong code; nor can it indicate the error if one has typed only a fabric code, but forgotten to type in the form and decoration codes, or to fill in other fields. While certain types of validation will always rely on the person doing the inputting, there is scope for further automation in the process, but this system goes quite a lot further than had been previously achieved.

The previous MoLAS system, which was Unix-based, validated only the fabric codes and checked that the number entered for the TAQ was indeed greater than the number entered for the TPQ. In the Oracle system, there are embedded date ranges for all fabrics, forms and decorations, which are drawn upon in order to create an automatic proposed date for a context, and this has led to an unprecedented standardisation in the dating of pottery groups. If one does not remember the date range of a particular type, or if one mis-types the date, this is immediately visible in the date range proposed by the computer. The arrival of date ranges from other sources, such as small finds, coins, and stratigraphy, is still at a relatively early stage (it has taken time to develop the system, and the co-ordination of getting data entered by everyone is sometimes), but the potential for improved communication among all the participants of any given project, and thereby for improved interpretation of the archaeology, is very considerable.

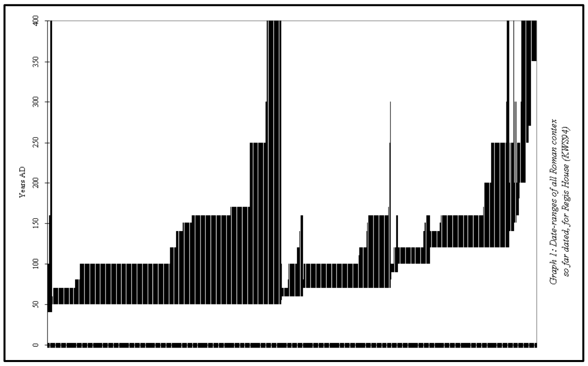

Figure 1. Date ranges for all Roman contexts so far dated for Regis House (KWS94).

Once embedded in the fabric, form and decoration lists, the date ranges can be used to generate date range graphs, either for sites as a whole (see Graph 1), or for contexts within sites (see Graph 2). These are created by transferring the Oracle data directly into Excel, and the greatest advantage of the system in this domain is the rapidity with which such graphs can be generated - this is the first time in the experience of MoLAS Roman pottery specialists that such graphs can be generated with sufficient ease that they can become a normal tool for examining the data at any stage of research.

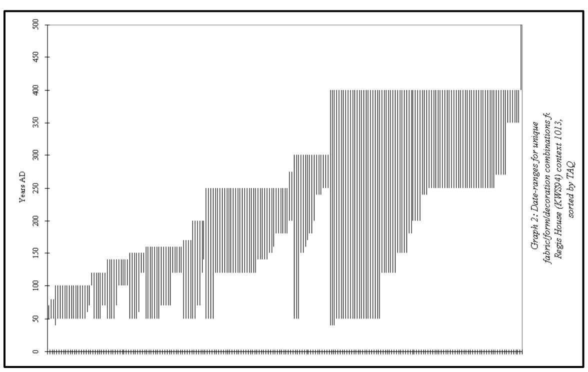

Figure 2. Date ranges for unique fabric/form/decoration combinations for Regis House (KWS94) context 1013 sorted by TAQ.

Pottery data is, however, more than simply dates. A very powerful additional element of the Oracle system is the embedding of grouping attributes for fabrics and forms, such as Ware (type), Source (imported - British or unknown), and Origin (currently with a total of eleven possible regions: Britain, Central Gaul, East Gaul, Gaul, Italy, Mediterranean, North Gaul, Southeastern Gaul (Rhône Valley), South Gaul, Spain and Unknown). With forms, there are nine general form categories: 1 flagons, 2 jars, 3 beakers, 4 bowls, 5 dishes, 6 cups, 7 mortaria, 8 amphorae, and 9 miscellaneous. This list is being expanded in order to group fabric/form combinations in terms of functional categories (cf Greene 1993, 77-80), although this is an area where much more work needs to be done.

These embedded attributes allow for the generation of tables such as the following, which relate to the recent MoLAS site at Borough High Street, near London Bridge (BGH95A)

Fabric Group	Definition	Rows	% rows	Sherds	% sherds
AMPH	amphorae	279	13.3%	761	17.8%
BBTP	black-burnished type	153	7.3%	210	4.9%
FNMP	fine imported	31	1.5%	50	1.2%
FNRB	fine Romano-British	162	7.7%	251	5.9%
FNRD	fine reduced	71	3.4%	100	2.3%
OXID	oxidised	382	18.2%	814	19.0%
OXMR	oxdised mortaria	16	0.8%	21	0.5%
REDU	reduced	494	23.6%	1075	25.2%
SAM	samian	310	14.8%	535	12.5%
TEMP	tempered	199	9.5%	457	10.7%
Total		2097	100.0%	4274	100.0%

Fabric Origin	Definitio n	Rows	% Rows	Sherd s	% Sherds
IMP	imported	649	30.9%	1407	32.9%
RB	Romano-British	1448	69.1%	2867	67.1%
Total		2097	100.0%	4274	100.0%

Fabric Source	Definition	Rows	% Rows	Sherds	% Sherds
BRIT	Romano-British	1449	69.1%	2869	67.1%
CGAUL	Central Gaul	70	3.3%	110	2.6%
EGAUL	East Gaul	48	2.3%	62	1.5%
GAUL	Gaul	3	0.1%	3	0.1%
ITALY	Italy	2	0.1%	3	0.1%
MED	Mediterranean	68	3.2%	106	2.5%
NGAUL	North Gaul	22	1.0%	53	1.2%
SEGAL	Rhône Valley	71	3.4%	221	5.2%
SGAUL	South Gaul	216	10.3%	404	9.5%
SPAIN	Spain	132	6.3%	423	9.9%
UNK	Unknown	16	0.8%	20	0.5%
Total		2097	100.0%	4274	100.0%

Form Group	Definition	Rows	% Rows	Sherds	% Sherds
AMPH	amphorae	278	13.3%	775	18.1%
BEAKER	beakers	117	5.6%	179	4.2%
BOWL	bowls	222	10.6%	381	8.9%
BOWL/DISH	bowls/dishes	67	3.2%	120	2.8%
CUP	cups	53	2.5%	92	2.2%
DISH	dishes	148	7.1%	242	5.7%
FLAGON	flagons	61	2.9%	125	2.9%
FLAGON/AMPH	flagons/amphorae	6	0.3%	26	0.6%
JAR	jars	617	29.4%	1344	31.4%
LID	lids	25	1.2%	31	0.7%
MORT	mortaria	83	4.0%	107	2.5%
MISC	miscellaneous	420	20.0%	852	19.9%
Total		2097	100.0%	4274	100.0%

Tables 1-4. Showing data rapidly generated through the Oracle report mode.

Perhaps even more so than with the date range graphs, tables such as these can be generated through the Oracle report mode in a such a rapid manner that the data can be reviewed on a daily basis, if desired. If a particular series of contexts seem to be exhibiting peculiarities which distinguish them from the rest of the material, tables similar to these can be quickly generated to illustrate the distinctions. There is a considerable degree of flexibility in the report mode, especially since new types of reports can be requested whenever they seem necessary. We have also recently generated similar tables to those above for London as a whole, and for London broken into five zones: the City, the City waterfront, Southwark (English Heritage sites), Southwark (Jubilee Line Extension sites) and No. 1 Poultry. New material can now be compared against the data from all these sites or against London as a whole, to see if there are obvious links.

4 Conclusion

The new system at MoLAS has improved the integrity of archaeological information captured, and has increased accessibility to and communality of the data. It continues to be developed in response to both the new types of output that specialists request, and the different types of archaeology which MoLAS has to deal with such as the more diverse archaeology in the more rural areas of Greater London. Equally, improvements are made to increase efficiency and reduce inputting time such as those currently being made to one of the system's central tables - the field interpretative index - which should provide a marked improvement in post-excavation time and thus cost by making greater use of the hierarchical relationships that exist between interpretative entities. Perhaps the most important imminent improvement to the system is an easy to use graphical query builder or Data Browser which will enable users to phrase their own ad hoc queries, which would not be appropriate to develop into standard reports. This will enable the user to interact more freely with the system without the developer needing to act as the intermediary.

Notes

1 The MoLAS system has been built using the Oracle Workgroup Server running 7.1.3.3.24 version of the RDBMS and with applications developed using the Developer 2000 and Forms 4.5 suite of tools. Network Software is Windows NT 3.5.

2 This validation process takes the view that the date range incorporates time lags caused by differential ceramic longevity, although a detailed discussion of this interesting topic is beyond the scope if this paper.

Bibliography

Andresen, J, and Madsen T,1996 Dynamic classification and description in the IDEA, in P. Moscati (ed.), *III International Symposium on Computing and Archaeology,* Rome, 591-602.

English Heritage 1991. Management of Archaeological Projects, English Heritage.

Greene, K T, 1993. Part 1, The Fortress Coarse Ware, in W. H. Manning, *The Roman Pottery, Report on the Excavations at Usk, Cardiff*, 3-124.

Koch, G, and Muller, R, 1993. Oracle 7 : The Complete Reference, Osbourne McGraw Hill.

Rauxloh, P, In prep. Servicing the Archaeology of Self: Information systems and the hermeneutic spiral.

Contact details

P. D. Rauxloh[1], R. P. Symonds[2]
Museum Of London Archaeology Service
Walker House, 87 Queen Victoria Street
UK
[1]e-mail: *pete@molas.demon.co.uk*
[2]*e-mail: robins@molas.demon.co.uk*