# The Romano-British Pottery Assemblage from Wroxeter Problems in Assembling a Database

S.J. Pierpoint
Wroxeter Roman City

A.M.C. Rochfort
University of Birmingham
(now of the University of Southampton)

This paper examines two initial problems faced by achaeologists in assembling a dataset -- cataloguing the data and transferring it to a mainframe computer. The obvious alternative of using an on-site micro was not considered suitable because of the quantities of data and lack of finance for a suitable system.

The Roman city of Wroxeter is one of the four largest of such in Britain, covering an area of some 200 acres (85 Ha.). Today almost the entire city is in the care of the Department of the Environment and thus protected from the depredations of the plough. Over the decades, the city has been the target for many archaeological excavations and today it is being excavated by two separate teams, working during the season almost within conversation distance. As on most Roman city excavations, work has been concentrated on the main public buildings. For the last twenty years Philip Barker, heading one of the teams, has been excavating the great Baths Basilica at the centre of the city using the latest techniques of excavation. The Basilica has been dug (in part) many times before, but Philip Barker's painstaking approach has revealed a plethora of phases of site use, never before recognised (Barker 1981, fig.4). Occupation of the site can now be traced, at least from the first construction of the huge church-like Basilica, in the mid-second century, througn various stages of re-flooring and then collapse (and probably demolition) of much of the Basilica in the fourth century. Then, long after the famous proclamation of Honorius in 410 A.D. that the British should look to their own defence, the city, far from declining, embarked on new phases of building. The old Basilica site was levelled with rubble and a classical style mansion or administrative building was constructed in wood on the site, with an accompanying "shopping precinct". At about the same time, large timber houses and out-buildings were also flourishing. The site was probably not abandoned until the 6th century A.D.

During the four centuries or so of the site's use, pottery was used, broken and discarded in and around the various buildings (We are still a little doubtful if contemporary pottery was used in the latest stages of the site's occupation.) The various excavations on the basilican site have probably produced evidence of at least 15,000 different pottery vessels. One major initial task is the cataloguing and analysis of the 120,000 sherds from the latest excavation. The very numbers themselves are intimidating.

Roman pottery analysis is becoming an increasingly important aspect of post-excavation work on sites like Wroxeter. It can provide vital information about a site's occupation and use as well as dating evidence. The main aims of the particular work at Wroxeter were as follows:

i) To produce a catalogue of all sherds recovered, identifying particularly the types of vessels and fabrics present. This will permit future reference and research;

ii) To produce basic quantifications of wares and fabrics on the site. This is important for a number of reasons including considerations of trade from different production centres;

iii) To aid the building up of the site chronology, and complimentary to this, to elucidate details of different wares and vessel forms. This is often seen as the primary task of the pottery analyst;

iv) To establish patterns of discard and residuality. Very little of the pottery at Wroxeter was in the place where it had been initially discarded, such as rubbish pits. Much of it had washed into drains, been trampled into floors or simply been included with other building rubbish. One difficulty was to decide how best to use this mass of pottery, including such large quantities of residual wares, and to decide which sherds were actually useful for dating particular contexts. For example, it would be useful to date the use of the different Basilica floor levels, but this is just the sort of area people will keep clear of refuse.

Before we can seriously consider any of these primary tasks it is first important to consider the nature and quality of the data. Computers are only useful if the data is good, on the old principle "garbage in -- garbage out". At Wroxeter the main difficulties were with on-site recovery. At the Basilica, recovery is quite good. All the soil is sieved through 3mm dry screens. But dry-sieving is limited in

the results it can achieve. Last year a brief sample wet-sieving programme was mounted using the same 3mm sieves. This technique essentially means that practically all sherds larger that 3mm are recovered. Looking at sherds of greater than this size, the experiment demonstrated quite clearly two important facts:

The biasing effects of normal recovery methods

| ware | black burnish | Severn Valley | Samian ware | calcite gritted |
|---|---|---|---|---|
| wet sieve no. of sherds | 530 | 517 | 41 | 266 |
| without w.s. | 188 | 176 | 35 | 45 |
| % without w.s. | 26% | 25% | 46% | 14.5% |

1. At Wroxeter dry-sieving recovers only ca. 50% of all the pottery sherds by weight. The sampling bias is not random either. Larger proportions of the darker wares were being lost. For example the loss rate for calcite gritted wares was twice the average and that for the shiny surfaced Samian considerably better than average. Thus counts of weight of sherds at Wroxeter would considerably bias the representation of different wares. However it was recognised that dry-sieving did recover some 95% of rim sherds and it was realised that a useful count of sherds could be effected on rims.
2. The quality of recovery on site varied sharply. One major factor affecting the recovery was the weather. The loss rate on site on wet days was twice that on dry days. Thus once again a bias is creeping into the data.

The effects of the weather on pottery recovery

| | wet days | | | dry days | | |
|---|---|---|---|---|---|---|
| | no. of sherds | weight of shds | mean wt | no. of sherds | weight of shds | mean wt |
| w/o wet sieve | 101 | 1027 gm | 10.2 gm | 489 | 2553 gm | 5.3 |
| wet sieve | 596 | 1320 gm | 2.2 gm | 1075 | 1740 gm | 1.6 |
| total | 697 | 2347 gm | 3.4 gm | 1564 | 4293 gm | 2.7 |
| % wet sieve recovery | 85.5% | 56.2% | --- | 69.0% | 40.6% | --- |

To my mind (SJP) a sampling strategy involving wet sieving would produce a truer representation of material on site. This might be particularly important if we are interested in the extent of trade from particular manufacturing centres. In other words such a strategy would maximise data quality. On the other hand such a strategy can create as many problems as it solves unless it is carefully planned.

It is interesting that there are still two distinct viewpoints in archaeology as to the handling of Roman pottery data. One stresses the value of quality in data and maintains that sampling strategies are an essential component of any such analysis. The other stresses quantity and the need to catalogue everything regardless of quality. The only drawback of the first approach for Wroxeter is simply the difficulty of constructing a sampling programme that will not throw away a great deal of information. Rare fabrics, crucial for dating, could easily be sampled out and lost. On the other hand it is simply impractical to wet-sieve everything. A pragmatic solution is best, and at Wroxeter we have gone some way towards a successful solution.

Because all contexts are at least dry-sieved losses of sherds are minimised (recovery rates are ca. 30% better than by trowelling alone) and we know that counts of rim sherds will be fairly representative of actual quantities in the archaeological contexts. The wet-sieving programme means we have a measure of the quality of the recovery. However, I (SJP) would like to use this paper to create a discussion on the handling of large pottery data sets. What are the advantages of

using such large quantities of material (80,000 + cases at Wroxeter)? What should our priorities be? Recovering more data and cataloguing all of it, facilitated by large computer systems, is no answer.

Our immediate aims at Wroxeter were to process the material we had in the light of our knowledge of data limitations. All the material was to be catalogued and analysed. It was decided at a very early stage that to fulfil our aims we were going to need a mainframe computer, and because Birmingham University was already involved with the project through Philip Barker, the DEC 2060 there was the obvious candidate. On the other hand the pottery could not really be moved, and our first problem was how to transfer the catalogue from Wroxeter, fifty miles from Birmingham, onto the DEC 2060. Various alternatives were considered, but only two seemed likely: we could catalogue the material onto paper and have the resulting catalogue professionally typed for the computer; or we could catalogue the material on a microcomputer and then transfer the tapes (we could not afford disc systems) to the DEC. For reasons of time and money the latter course was adopted. Our solution would cost only one half to two thirds the other. Preliminary enquiries indicated that professional typing was likely to be very expensive. With a good cataloguing program on micro this work can be considerably quicker than using paper forms. In tests, cataloguing using the micro proved to be 50% quicker than using paper records and also eliminated problems of bad hand-writing.

Wroxeter therefore duly acquired its micro, a 16K Commodore PET with tape drive. Most of the memory space is used up by the BASIC cataloguing program written by SJP and built for speed and accuracy. The program verified all records as they were entered, and would query values it had not encountered before. The verification side is extremely important. Unfortunately it was discovered that there were very occasional errors created in data transfer and a second program had to be run to check the data again, once it was on the DEC.

Because many archaeological sites are remote from computers there will be many cases with the problems we faced in data transfer if they are using mainframe computers. From my (SJP) experience the system we used, particularly without a disc drive, was still wasteful in terms of time and money, both of which are vital concerns in archaeology's impoverished state. About 25% of cataloguing time was spent in transferring data from cassette tapes to disc and then to the DEC. This was

particularly frustrating when the cataloguing program proved so efficient. The tape system was also found wanting -- the time taken by winding backwards and forwards on tapes was a real annoyance. Copying onto two tapes or copying from one to another to allow duplicate copies also used much valuable time.

The PET 16K itself proved perfectly adequate for the task. Aside from occasional irregularities in the Wroxeter electricity supply and the regular wearing out of the PET return key (pressed two million times) there were no serious problems. The PET BASIC was straightforward. The only slight difficulty was the habit the PET had of jumping into program mode when mistyping at speed, and thus closing files.

The data, in the form of 51-character strings, had to be transferred to a directory on the DEC before processing could start. Because the DEC has many users and might not always be able to attend to the PET, and because the PET had no start/stop transmission facility, it was decided to transfer the records to an intermediate system with a faster rate of character transfer than the 300 baud the PET could offer, which could communicate with the DEC (and so withhold data if the DEC were unable to receive it). To this end, a procedure was set up to read the tapes on the PET, and transfer them to an IMS 5000. From an IMS to the DEC was already a well-documented routine (Ramsay) and that part of the transfer was accomplished at 1200 baud.

The number of records alone -- some 80,000 -- made it imperative that the method used to analyse would not be swamped by the sheer quantity of the data. With more that 20 fields per record, obviously manual methods were out of the question. A database -- in this case RAPPORT by Logica -- was therefore implemented to allow researchers convenient access to the data. When the data had been transferred to files on the DEC, it was then read into the database, which could be interrogated by the archaeologists.

RAPPORT is a relational database system, which effectively means that the data records must all be of the same format and length. (You can tell if you got it right if the data, when printed out, is a rectangle.) In this case, the records were encoded and indecipherable to anyone unacquainted with the code. An example of a record before it is entered in the database is strikingly uninformative :

R2--R1----4P--B2-3---000201------RD--13000100007---

but this record contains information about fabric, shape, vital statistics, context and decoration. When a number of these records are inserted into the database, information about distribution of types of vessel, fabrics, percentages of rim fragment and like concerns can be generated and displayed in convenient formats. For this project, convenient formats were deemed to be histograms and tables, as being the best and most easily-understood forms of output.

RAPPORT offers an Interactive Query Language facility which is both simple to use and very powerful. It easily performs the sort of enquiries which researchers need, such as counting, averaging, retrieving all instances of a certain value in a field, and so on. The database can also be accessed by programs in an enhanced FORTRAN (called FORTRAN+) which are compiled and executed outside the Interactive Query Language, and can be either interactive or batch. Ordinary FORTRAN programs can use the database at one remove, by using data files generated by one of the previously described methods directly accessing the database, and held in the user's directory.

The archaeologists who will be using the database will require both interactive and external programs for their research. The former will mostly be used for exploratory searches, and the latter largely for statistical purposes. They need lists of values present for any particular field, with totals of sherds involved, means and percentages of values, histograms showing frequency, and averages, ranges and standard deviations. These are the initial requirements merely to find out what is included in the collection -- later research will be more detailed.

Programming in FORTRAN+ and ordinary FORTRAN, for really involved database manipulation, will probably be beyond the range of most archaeologist-researchers. However, in the cases of large transactions, or complicated retrievals, programs in these languages will have to be attempted.

It is not impossible -- because we have both done it so we know -- that archaeologists can become happy computer users. It is a bit like preaching to the converted, as many others have done the same, but it is generally felt that more could be attempted and more accomplished if the great majority of archaeologists

could be persuaded. Archaeologists should not be frightened of computers or intimidated by large quantities of data, but as archaeologists they should really be concerned with high data quality and with designing systems and program sets to answer their particular problems and suit their particular needs. At Wroxeter we have achieved a practical solution to our initial problems, working all the time to this end, within the limits of time and money.

REFERENCES

Barker, P. (ed) 1981 "Wroxeter Roman City : Excavations 1966 - 1980", DoE Publication

Logica 1979 "RAPPORT Designing and Using a Database"

Ramsay,P. 1981 "A Suite of Programs for File Transfer Between Computers", University of Birmingham

Rochfort, A. 1982 "Design of a Database for Wroxeter Roman City Pottery", University of Birmingham MSc project