Practical Applications of Computer Processing in Winchester

N.G. Steger, with V. Jones,
D.J. Keene, S.V. Keene.

Winchester Research Unit

1. Introduction

All the Winchester Research Unit projects requiring computer analysis are basically similar. They entail processing data pertaining to a large number of items, such as 10,000 small finds or 12,000 sherds of Samian pottery. In these two examples the information necessary for each item must be vetted and cross-checked and a computer record created for each one. The records are then sorted into different arrangements with specified information extracted, reorganized and printed in a variety of formats. These are, as it happens, the characteristic features of a typical commercial (as opposed to scientific) data processing application like inventory or payroll. Consequently the projects could have been programmed using any of the all-purpose programming languages like Cobol or PL/1 or using any generalised information system. We were given the opportunity of using the computers (IBM/370) belonging to the Hampshire County Council and their in-house programming system.

2. Operating Procedures

The Hampshire County Council programming system provides file maintenance and information retrieval facilities in a form which can be used by non-programming personnel as well as by experienced programmers. Coding programs involves filling out a variety of tabular forms, each designed for a specific purpose like record vetting or record selection..

We found the advantages of using the HCC programming system far out weighed the limitations. It is easy and quick to program and extremely easy to test. Debugging time is remarkably low. With minimal instruction the archaeologists themselves could make simple program modifications.

We have a number of projects currently on or near the computer. The Small Finds Analysis is virtually complete although the computer file will be saved indefinitely. The file for the Samian Pottery Analysis is now being created. The Bone Analysis is in the design stage with record formats being prepared. The coarse pottery is about to be converted from PL/I to the HCC system.

The aims of the various projects are similar, being to compile lists of the finds, by kind, by date, by structure and so on, so that a number of analyses can be made (and these where possible by computer as well). We wish to compare the time of introduction, period of full use and time of disuse of the various kinds of finds, to compare the concentrations of types of finds from different properties or structures ar different periods, to see if they suggest use, or reflected known uses, and for finds with site co-ordinates, to look at spatial distribution of categories of object. One of the archaeological problems encountered with the Samian concerns dating. All the sherds had been seen by the specialist and virtually all had been identified.

Samian vessel forms are well catalogued, so the form identifications were easily translated for coding purposes. Dating is more complicated. It depends on a combination of form, fabric, decoration or stamp. Some sherds, decorated or stamped pieces, may be dated very accurately even to within a few years. For other sherds a general period date such as "Antonine" or "Flavian" may have been given and quite often the specialist is only able to date a sherd to within a century. So in the specialists report a sherd date might appear as 145 - 160 A.D., or "Antonine" or "C2". For coding, it was important that all dates should be comparable. A precise definition of what in years was meant by the specialist when he used terms such as "Antonine" or "mid C2" had to be obtained, so that the non-numerical date could be translated into numbers. A single letter code was used to indicate the type of date given by the specialist to allow us to recognise if a sherd had been given a precise or broad date range - i.e. the letter A indicates a precise numerical date, B indicates a precise term e.g. Claudian, C indicates a period such as pre-Flavian, or Flavian, while D indicates a broader date such as first century.

From a programming standpoint the projects are extremely alike and can be most effectively discussed together.

For each project the archaeologists and programmers have designed a format for the punched card input. Each small find or Samian pottery sherd is represented by one 80 column card with a fixed format. The Samian records, for example, begin with a site code followed by year of find, phase, trench, layer, sherd number, form, date, fabric, etc. All the information required for the output report must appear on the cards (unless it is added by the computer like small find super-category or Samian period). Each card is unique. The Small Find find number or Samian sherd number in combination with other fields like site or category, specified as identifier fields in the program, identify each find or sherd without confusion.

We used the following elements of the HCC system. An input program, HCINPUT, creates disc files from, in our case, punched cards. The user describes the format of the individual fields in both the card and disc records and indicates what vetting should occur for each field. A list or range of permitted values can be specified for each field. In addition, values in particular fields can be cross-checked and values in fields altered or filled in on the disc records.

An up-dating programme, HCUPDATE, updates a disc file using insertions, deletions and modifications provided by the user. The same vetting facilities are available as in HCINPUT.

A report generater, ESP, creates reports or lists from any HCC disc file with extremely simple specifications. One form indicates which records should be extracted from the file for a report, according to particular values in specified fields. Another tabular form contains a list of fields in the extracted records to be printed as well as instructions about required subtotals and the order of the report. One line is normally printed for each extracted record.

A wide range of statistical facilities are provided by TAP and TAPEN. Using simple forms, users can request extracted records to be combined into multi-dimensional arrays with row and column totals and percentages. All the standard statistical functions are available, although we have only used a few of them.

The fields on the punched cards are vetted by the appropriate programs and disc files are created for each project. The records in the disc files are stored in an order specified by the user. The Small Finds file is held in order by site, find type, and find number. Individual fields are checked to ensure correct values. Site fields, for example, are checked against a list of valid site codes. Find or sherd numbers are checked to see that they fall within a range of assigned values like 1 - 3,000. Some fields are checked to ensure simply that they contain only alphabetic or numeric data. Any errors are printed with an appropriate message and the error record is not added to the file, so the disc file contains only correct records.

Fields are also checked against one another. Year fields are checked against site codes to make sure that the specified year was one in which excavations occured at that site. On Samian records, the fabric B field (a specific location) is checked against fabric A (a general area like southern or central Gaul) for a valid combination.

Periodically the entire file is printed in the order in which it is held on the disc. This permits the archaeologists to check the records for any errors which have not been picked up by the vetting procedures. Modifications are made to a record merely by punching a card containing the identifying fields, any fields to be altered, and a one-character code to tell HCUPDATE, the updating program, whether to insert, delete or modify.

Once the file is complete, we run the programs producing the reports. One type of report we generate is a list with simple totals. Among others we have generated a Small Finds list of the number of finds for each category within material. Two more list all Small Finds records in order first by category, site, find type and find number and then by site, century, phase, category, find type and find number. Selected records were listed in

century, site, structure, sub-structure, category type and find number order. The Samian specifications are equally varied. For instance, for each site the computer will provide a total listing in phase order of all the Samian, as well as listings of each fabric, vessel form, decoration and potter's stamp. In addition, there will be a total listing by period within each site. Lists can be produced using any sub-set of the computer records in any order which might be useful.

As well as the simple lists, we generated a variety of Small Finds distributions. For iron items, a table gave the number of finds and the percentage for each century by site. Row and column totals and percentages also appeared. Another table did the same for super-category by century within structure and site. Again, any sub-set of records or the entire file can be processed by the computer to create whatever tabulations might be helpful. We have produced many other reports, all of the two types described above. The possibilities are clearly limitless as the computer can manipulate thousands of records with ease.

3. Problems

We were forced to operate under a number of constraints as far as computer programming and operation were concerned. The available staff was very limited. One volunteer programmer/analyst assisted the archaeologists in designing their required record formats and report formats, wrote and tested all necessary programs, organised the final computer runs and checked and distributed output. A member of the HCC Planning Department cooperated with the programmer by answering technical questions, investigating problems connected with the HCC system, setting up and submitting jobs, and acting as liaison with the HCC Computer Centre. A variety of amateur and professional key-punchers had to be called upon to punch all the cards and corrections.

At HCC request, we agreed to operate within the limits of the HCC programming system and not to avail ourselves of its facilities for adding user code in Assembler language to perform special functions. If the HCC system was unable to do what we wanted, we modified our requirements.

Designed for a specific installation, the HCC system has certain limitations for users whose work, like ours, does not conform to HCC standards. A limit exists on the number of values that can be specified for vetting one field as well as on the number of cross-checks and calculations in each program. It is difficult to send parameters to a program if special data is required for each run. The spacing of all reports created by ESP is fixed and cannot be altered. The user specifies only the order of fields as they appear across the page, not spacing between fields and between lines. Each record on the disc file must be unique. The user specifies which fields identify the records on his file and the contents must not be duplicated. For archaeological purposes many Samian pottery sherds are not required to be identified precisely, but our program must assign a unique number to each of the sherds to meet the HCC requirement. Fortunately, the archaeologists were prepared to be flexible about their requirements and we experienced little difficulty working within the system.

We encountered a number of problems arising from both the data processing and archaeological aspects of the work, which delayed our projects. The programmer, whose training for this project consisted of reading the HCC manuals, was initially unfamiliar with the system and could not code as rapidly in the early months as she could later on. In a number of instances, as a result of undocumented system limitations we either had to change the format or our output or try a variety of programming techniques to achieve a desired result. For example, we had to program one Small Finds tabulation individually by site because the statistical package rejected a three-dimensional array of alphanumeric variables. It took several experiments to achieve zero suppression on output as the HCC system does not automatically provide it.

To save money records had been keypunched on a variety of machines by numerous people. Keypunch errors and inconsistencies between IBM and ICL keypunch errors and card codes necessitated a certain amount of repunching and file recreation.

Where punching was done by unit staff involved in the finds projects, accuracy of punching was better, partly we suppose, because it could be done directly from original records, and partly because the puncher understood the data, which minimised nonsensical errors.

The archaeologists, new to this type of data processing, found that the computer records and output did not always turn out as they expected and requested changes to make the input and output more useful. The specifications have never been frozen. The data, prepared, encoded and keypunched by a variety of people, was not as accurate as we anticipated. We had to change many of the vetting instructions, usually to permit greater latitude Some values such as date/form combinations for Samian sherds, ahve been changing so rapidly that we no longer alter the program but check those fields by eye.

Added to these causes of delay, we had the basic problem of a long chain of communication, partially volunteer and therefore not always available, located in several places around Winchester. Merely to implement a programming change involved the archaeologists, volunteer programmer, HCC liaison and the HCC Computer Centre. Consequently wvery step took longer than anticipated and longer than one would normally estimate for such work. Fortunately, most of the projects had or have sufficiently flexible schedules to accomodate delay, although some of the small find analyses could not be produced in time.

A good deal of preliminary work was carried out on a proposed computer analysis in connection with on of the Research Unit's projects of historical investigation. In this case the data would have consisted of biographical evidence for well over 1,000 citizens of mediaeval Winchester. The range of information on each individual was wide and varied and would have

had to have been accommodated on several 80- column cards. It became clear that the time and manpower necessary for creating and verifying the computer file were far greater than were actually available and the scheme was abandoned. The body of data is now being analysed less extensively by other means, but has been arranged in such a way that the computer scheme could readily be taken up again in the future.

There seem to be few archaeological organisations with the resources to provide purpose built schemes. Those who embark on work using other people's systems need to be prepared at the outset to compromise over their requirements and methods, and to abandon if necessary some or all of the work if it is non-productive.

In conclusion it has taken us longer to produce output than expected and the reports do not always look exactly as originally specified. However, the output has been found to be extremely useful and clearly worth the effort, not only for use at present in the unit's work, but also for the future. The lists produced by the computer will form part of the level 3 data to be deposited in the museum. It is the experience of the Winchester Research Unit that at least one type of archaeological analysis can be handled successfully using the basic principles of commercial data processing and an appropriate programming language or information system.