Digging for Data. The Use of Fieldcomputers at IPP Excavations and Surveys

Kamermans, H.; Voorrips, A.; Wijngaarden Bakker L.H. van

DMZ - II.6 : H. Kamermans and A. Voorrips

Digging for Data : the Use of Fieldcomputers at IPP Excavations and Surveys

Summary

There are three main applications of computers in archaeology : Data storage and retrieval, statistical and mathematical analysis, and, using microcomputers, preliminary data entry and processing. In the Albert Egges van Giffen Institute for Prae- and Protohistory both microcomputers as well as the Amsterdam Universities mainframe (SARA) are used in a well balanced interrelationship.

Data processing at the excavation of a Roman Iron age site is presented as an example of an integrated application of microcomputer and mainframe.

Introduction

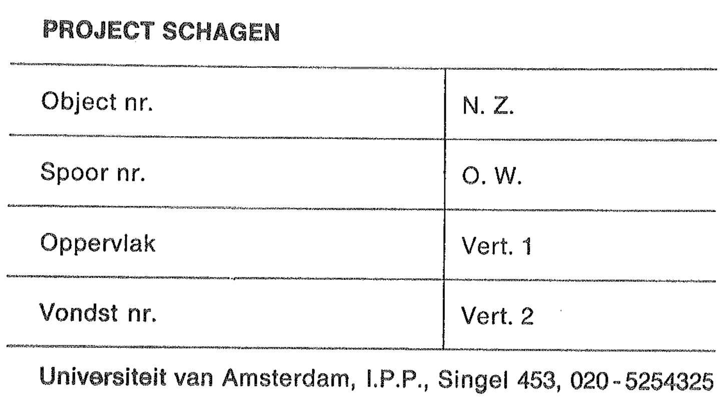
There are three main reasons for using computers in archaeology. Firstly, modern archaeological field research produces an enormous amount of data. Computers are almost indispensable for storage and analysis of these data. Secondly, the development of statistical and mathematical methods for the analysis of archaeological data leans heavily on the availability of computers and analytical computer programs. Thirdly, archaeologists working abroad often cannot take their finds out of the countries where the fieldwork takes place. Materials must be coded and analysed in a primary fashion during the fieldwork campaigns. Small computers are feasible for these purposes. Moreover the coded data can be taken home and used for more substantial analysis after fieldwork has been completed.

The Albert Egges van Giffen Institute for Prae- and Protohistorie (IPP) has used computers in archaeological research since 1971 (Groenman-van Waateringe, this volume). From 1979 onwards emphasis has been put on the use of computers in fieldwork situations. The IPP has an alphanumeric and a graphics terminal, connected to the University of Amsterdam's main frame computer, and four microcomputers. All microcomputers can be linked to the main frame as well. Most of them are intensively used in the field, both during excavation and survey projects.

The major application of a microcomputer in the field is to enter, temporarily store, update and retrieve data. Monitoring the « off line » data entry and some preliminary sorting and printing is done mainly for correcting the data. Then, when feasible, the microcomputer can be linked to the big Amsterdam host system, and the « clean » data are quickly transmitted for storage to a database residing in the host system. From this database, further information can be retrieved, be it for administrative purposes or for analytical ones. As the storage media of our microcomputers (5 ¼″ floppy disk units) have a restricted capacity and since the Amsterdam host system provides reasonable low cost services, we are not planning to shift the balance in favour of microcomputers. Thus we consider our microcomputers valuable additions to the possibilities offered by a big host system. It must be noted that, besides the applications of microcomputers dealt with in this paper, the use of microcomputers as word processing systems has become extremely popular, and for good reason. The use of a microcomputer with a good word processing package often will mean a substantial decrease of publication costs and a substantial increase in publication speed.

General field procedures

The excavation techniques used at the IPP are the same as on almost every dig in Holland. A bachhoe digs and levels trenches. Layers and plans are drawn, finds are collected and put into bags, samples are taken. The *first data flow* consists of the information which is irretrievably lost as the excavation proceeds : the spatial location of finds and features and the relations between the observed phenomena. A duplicate label is filled out with findnumber, objectnumber (pit and level), feature number, x, y, z-coordinates and, if samples are taken, sample type (Fig. 1). One label goes in the bag with finds to the field lab, the other goes to the computer.


PROJECT SCHAGEN

Object nr.	N. Z.
Spoor nr.	O. W.
Oppervlak	Vert. 1
Vondst nr.	Vert. 2

Universiteit van Amsterdam, I.P.P., Singel 453, 020-5254325

Fig. 1. Find label for the IPP project Schagen (photo IPP).

The *second data flow* consists of further information that is normally collected during the excavation campaign itself, and which, together with the first flow, provides the excavator with the necessary feedback to regulate the excavation process. In the field, lab finds are washed, numbered, weighed, counted and assigned to categories like pottery, bone, wood, etc. This information is written on forms and these go to the computer.

The field computer is sitting somewhere safe and warm within reach of the excavation. Someone is there all day to enter the data from first and second flow into the microcomputer. A data base program creates a menu on the screen (Fig. 2) such that everybody who can type can enter the data. In the evening the data is sorted and printed. The printout is checked and mistakes can be corrected. The data from the field and from the lab are easily linked by their corresponding find nummer. The field supervisors have at their disposal sorted and corrected data such as : Where did we find *Terra Sigillata* ; where were our phosphate samples taken ; give me a list of all finds from feature 453, etc.

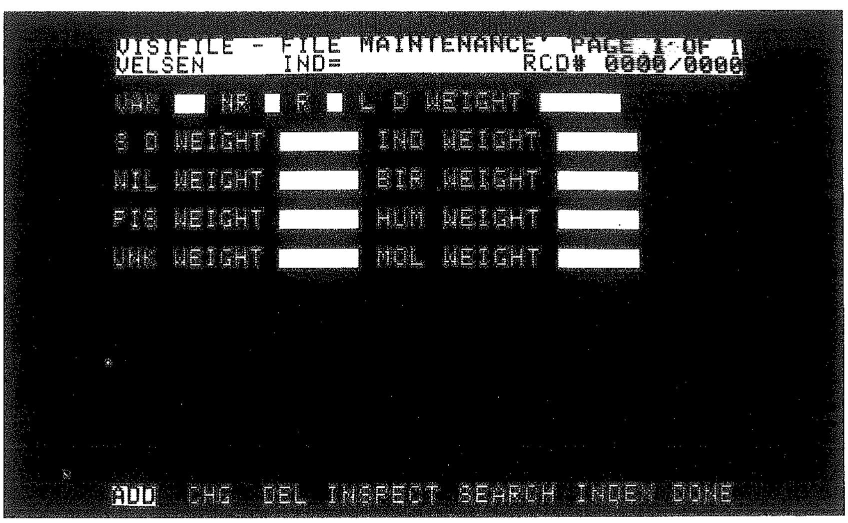


Fig. 2. Screen menu for entering data for the IPP project Velsen (photo IPP).

The *third data flow* contains the information specific to a particular find category and/or to a particular research method. Pottery is analyzed in a different way from pollen samples, or charcoal, and so on. Especially in excavations abroad when the finds are not allowed to leave the country, these data are collected during the field season. In this case information is entered into the field computer. In the end, all the corrected data will either stay stored on floppy disks until after the campaign or be transferred by telephone to the host computer system in Amsterdam.

2. *An example of an application*

Schagen Muggenburg (North-Holland) is a Roman Iron Age site dating to the 2/3rd century A.D. which was excavated in 1983 and 1984 (Therkorn, 1984). At this excavation we used the relational database management system (DBMS) dBASE II (Radliff, 1981) as a data entry program. It is easy to structure and easy to query by non-experienced users. The database structure was defined by the computer department of the IPP, but the data entry and retrieval was done by various members of the excavation crew. With dBASE II data can be entered with a maximum of 32 fields and 1000 characters per record, which was sufficient for the first data flow. The second data flow was divided over three different files. Search, sort and print routines were programmed such that casual users should have no trouble in working with the local databases. After correction the data were transferred by telephone to the host computer in Amsterdam.

On the host computer databases had been defined under a DBMS called SIR (Scientific Information Retrieval, Robinson *et al.*, 1980). In SIR a hierarchical database with possibilities for networking can be created. The structure for the Schagen database and the retrievals were written by the computer department of the IPP. The first data flow formed the highest level in the SIR database. The second data flow formed the second level, and so on, down to the information about individual artifacts. The advantages of SIR over dBASE II are enormous. Its capacity is nearly unlimited, and it has extensive search and selection possibilities. The program has also some statistical functions like frequencies and crosstabulations, and it can automatically create SPSS (Nie *et al.*, 1975) and BMDP (Dixon & Brown, 1979) files for further statistical analysis. A report generator helps with the creation of reports (van Wijngaarden-Bakker, this volume ; Lange, this volume). In addition to our SIR databases we use a number of graphical packages like DISSPLA (Integrated software systems corporation, 1981) and SURFACE II (Sampson, 1978) to produce distribution maps, contour maps and block diagrams.

3. *Problems*

The computer processing of archaeological data has its general and specific problems. General problems to be dealt with are the lack of data checking procedures during the processes of data entry, data correction and data transmission to a host system. In Voorrips (1984) these problems are considered in more detail.

The use of dBASE II creates some specific problems. The fact that, for the Schagen research, the second flow had to be divided over three files had to do with the limit of 32 fields per record. dBASE II can only work with two files at the same time but, with a little programing, data entry can be performed without anybody noticing that three files are in use. The new dBASE package, dBASE III, (Ashton-Tate, 1984), can handle ten files at the same time, but only runs on 16-bit machines. Another major drawback is the fact that sorted files have to be stored on the same disk as the original data. This means in the case of Schagen that only 500 field records could be stored on one floppy disk and that only 500 field records could be sorted or searched at the same time. These limitations can be solved with a hard disk for the microcomputer. In conclusion, a 16-bit microcomputer with a hard disk, and dBASE III as its DBMS seems, at the moment, a good solution for entering and retrieving data during an archaeological excavation.

SIR, our DBMS on the main frame, has some disadvantanges as well. The SIR query language is complex and not very userfriendly. This means that it is almost impossible for the excavation staff to query the SIR database without help. For the more often occurring questions retrievals, prewritten by the IPP computer department, are needed. For new or complicated queries *ad hoc* help from the computer department is necessary. There is a new version of SIR called SIR/SQL+, however. In this version it will be possible to query the database as if it were a relational («flat») database. A retrieval of 17 lines of original SIR query is than replaced by 4 simple commands. It would certainly make query by casual users much easier. Nevertheless, it is clear that SIR does not provide the best possible solution for archaeological database management. Both its hierarchical structure in which at best some pseudo-networking can be accomplished, as well as its difficult basic query language make us look forward to the implementation of a well designed relational DBMS, preferably on the host system, but if necessary even on a medium sized « own » system.

4. The future

The development of hardware during the last few years is remarkable. The software however had evolved less dramatically. Maybe we must put our hope on the progress in artificial intelligence. A number of general subsystems are already being designed and it will be only a matter of time until a complete artificial intelligence « toolkit » will be available for managing databases, including archaeological databases (Cerri, in press). Then it will be possible for every archaeologist to convert the data stored in a DBMS into usable information.

Acknowledgement

The authors wish to thank Gertrudis A.M. Offenberg for the title of this paper.

REFERENCES

ASHTON-TATE, 1984, *dBASE III User Manual*, Los Angeles.

CERRI, S.A., in press, « *Intelligent* » *interactive teaching*. Pisa.

DIXON, W.J., 1979, *BMDP-79. Biomedical Computer Programs P-Series*, Berkeley.

GROENMAN-VAN WAATERINGE, W., *Introduction to the Workshop on data management of archaeozoological assemblages*, in *this volume*.

INTEGRATED SOFTWARE SYSTEMS CORPORATION, 1981, *DISSPLA Display Integrated Software System and Plotting Language, User's Manual*, San Diego.

LANGE G., *Networking in SIR database management System. An archaeobotanical example*, in *this volume*.

NIE, N.H., C.H. HULL, J.G. JENKINS, K. STEINBRENNER, D.H. BENT, 1975, *SPSS. Statistical Package for the Social Sciences. Second edition*, New York.

RADLIFF, W., 1981, *dBASE II Assembly-Language Relational Database Management System. User's Manual*, Los Angeles.

ROBINSON, B.N., G.D. ANDERSON, E. COHEN, W.F. GAZDZIK, L.C. KARPEL, A.H. MILLER, J.R. STEIN, 1980, SIR Inc. *SIR user's manual version 2*, Evanston SIR Inc.

SAMPSON, R.J., 1978, *SURFACE II graphics system* (*revision one*), Kansas.

THERKORN, L.L., 1984, *Schagen : Muggenburg*, in *Holland*, 16, p. 215-216.

VOORRIPS, A., 1984, *Data catchment analysis and computer carrying capacity*, in R. MARTLEW (ed.), *Information Systems in Archaeology*, London.

WIJNGAARDEN-BAKKER, L.H. VAN, *Scientific information retrieval of zooarchaeological data*, in *this volume*, p. 115-128.

H. KAMERMANS
Instituut voor Praehistorie
Universiteit van Leiden
Reuvensplaats 4
2311 BE LEIDEN, the Netherlands

A. VOORRIPS
Albert Egges van Giffen Instituut voor Prae- en Protohistorie (IPP)
Universiteit van Amsterdam
Singel 453
1012 WP AMSTERDAM, the Netherlands