MASCA JOURNAL

MUSEUM APPLIED SCIENCE CENTER FOR ARCHAEOLOGY
THE UNIVERSITY MUSEUM • THE UNIVERSITY OF PENNSYLVANIA
33RD & SPRUCE STS. • PHILADELPHIA, PA. 19104 • 243-4060 (Area Code 215)

Volume 1 Number 2 June 1979

COMPUTERIZED DATA BASE MANAGEMENT SYSTEMS IN ARCHAEOLOGICAL RESEARCH: A SELGEM CASE STUDY

W. SCHAUFFLER

South and Southeast Asia Section The University Museum University of Pennsylvania Pa. 19104

Introduction

Computerized data base management systems (DBMS), once used almost exclusively by large museums for administrative functions and collection management, are now being adapted by archaeologists for specific research projects. Our use of a DBMS for field records and artifactual materials from excavations in Thailand is but one example of this tactic for which there are only a few well-documented precedents (*e.g.*, Bennett 1974, and Bohrer and Adams 1977, for ceramics and botanical materials, respectively from the American Southwest).

We will not argue the relative merits or shortcomings of particular DBMS used by archaeologists (among them, GRIPHOS, TAXIR, GIPSY, MARK IV and SELGEM: see Chenhall 1975) or debate the methodological implications of using a DBMS in the first place (Doran and Hodson 1975). Rather we will outline our use of SELGEM as a research tool and discuss how it was adapted to the needs, resources and objectives of our Thai project.

DBMS selection and implementation

In 1974, the Thai Fine Arts Department and the University Museum initiated a joint long-term investigation of the prehistory of northeast Thailand (Gorman and Charoenwongsa 1976; Schauffler 1976). The need for computerized data processing was dictated by the sheer quantity of disparate field records and artifactual materials (engendering variable provenience information and idiosyncratic observations as well as standardized attributes). The choice of a particular DBMS was influenced by several other factors. We needed a system that was sufficiently comprehensive to be useful during initial sorting and cataloguing, subsequent statistical description, final publication and eventual archival storage of the data. We could not afford to make extensive software modifications nor did we want a system dependent on prohibitively expensive or exotic computer hardware or software that might restrict future use of the system or the data base. We recognized that the system would be managed and used by archaeologists (not computer programmers), supported in the main only by volunteers and work-study students. These considerations led us to select the SELGEM system (an acronym derived from SELF GEnerating Master), the operational and documentation features of which were prepared by Reginald Creighton and his colleagues at the Smithsonian Institution and available for just the cost of duplication. (For supplementary documentation see also Neuner 1976.)

Primarily a non-interactive batch system with versions operating in a number of computer environments, SELGEM was certainly the most economical option for our project. Fifteen program modules covering four broad functional classes (data entry, editing and restructuring; file resequencing; data retrieval; report writing) offered a flexibility and redundancy that were ideal here.

The system was installed for approximately $700 with a minimum of professional help primarily to work out procedures to execute the SELGEM program modules in accordance with the protocols of the IBM 370/168 OS/VS2(MVS) system. By far the majority of our costs arose from testing these procedures and evaluating the capabilities of the system prior to the entry of real data (*cf.* Humphrey and Clausen 1977). Depending on how our use of SELGEM evolves in the future, however, we anticipate that increased programming support will be necessary as the data files become larger and more complex.

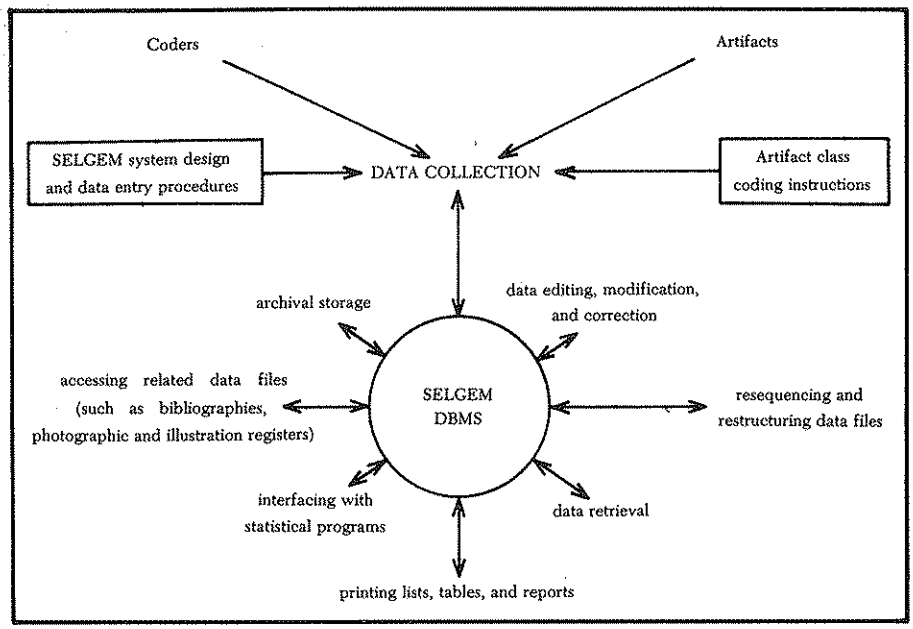
Data acquisition and entry

While it is incorrect to regard data entered into a DBMS as "fossilized" it is true that DBMS usage calls for a more critical attitude towards the data, keener insights into the structure and inter-relationships among the data and an unambiguous idea of what will be done with that data. With SELGEM any particular data entry (or set of entries) can be changed or deleted; data can be re-ordered according to specified criteria; and observations arising from study of the data can be entered into the data base at any time (Fig. 1).

Apart from the reconstruction of pottery, the greatest amount of human labor in our research project has gone into the coding of data. Coding formats and instructions have been developed for each artifact class: bangles, pellets, spindle whorls, beads, crucibles, animal figurines, rim sherds, pottery, and many others. Two levels of structure, one required by the SELGEM system and a second specified in the coding formats and instructions, have served to systematize the work of the volunteers, not only to focus their observations but to insure the accuracy and consistency of their work. With a number of volunteers working simultaneously on several classes of artifacts, a structured approach to data collection is imperative.

For most variables, the appropriate attribute state is coded in a fixed format although a more elastic format is used in those instances where the attribute states are not mutually exclusive. Additional comments and observations recorded where appropriate. Virtually all variables, regardless of the scale of measurement, are coded numerically to utilize the

Fig. 1:
Functional overview of data collection and management using SELGEM. In our research program we use the modular SELGEM programs individually or in various combinations as needed.



retrieval potential of the SELGEM system, facilitate statistical tabulations, and reduce data entry costs. In this way, data are entered only once and subsequent selection and transformations of the data are accomplished by means of SELGEM or other computer programs.

Though it is difficult to estimate costs here accurately, we can, as an approximation, itemize the figures for the entry of one of our early data files as follows. Of $318 spent in preparation of 1660 records (in practice, bags of sherds), yielding a total of 13560 lines of data, 76% of our costs were in key punching, a little under 9% were in creation of the data file using SELGEM programs, 6% went into correcting that file and 9% was used in printing out the file for correction and use. Consequently the average cost of creating each record was about 19 cents; of each line (*i.e.*, each artifact variable), about 2 cents. Subsequent refinement of our entry procedures now suggests we can significantly reduce unit costs, though it must be remembered that our estimates do not include any allowance for staff salaries, system implementation, maintenance charges or any incidental expenses that might have arisen.

Using the system and the data base

The SELGEM program modules can be used individually, in various combinations or interfaced with other programs. We first used the SELGEM to list the bags associated with

Fig. 2
Example of output from a SELGEM report writing program. In this case the format has been designed to fit on a 5 x 8 inch catalog card. Most of the original data was coded numerically on less than two computer cards.

```
THAI FINE ARTS DEPARTMENT/UNIVERSITY MUSEUM RESEARCH PROGRAM

 ARTIFACT RECORD INFORMATION
   CLAY PELLET, SERIAL NUMBER 00127350, FROM BAN CHIANG BCES
   SMALL FIND NUMBER:  0356                BAG NUMBER:  0968
   CONDITION OF ARTIFACT:  MORE THAN 1/2 OF ARTIFACT EXTANT

 PROVENIENCE DATA
   LAYER NUMBER:  13                       SQUARE:  D4
   QUADRANT:  SOUTHWEST
   BURIAL NUMBER:  008
   ARCHAEOLOGICAL ASSOCIATION AND LOCATION:
     BURIAL RELATED--BENEATH  THE LEFT  UPPER LEG

 ANALYSIS AND OBSERVATIONS
   CODER:  ARCH RILEY                      NUMBER OF PIECES:  001
   CONDITION OF SURFACE & EDGES:  PRISTINE
   WEIGHT OF PELLET:  07.4 GRAMS    DIA. OF PELLET:  1.9 CM.
   POTENTIAL USE RELATED DAMAGE:  NO EVIDENCE FOR DAMAGE
```

each of approximately 125 burials so that these bags could be pulled and the pottery reconstructed, one burial at a time. Working with burial plans and uncomputerized bag logs, it was not unusual to spend five hours or more to list and check the bags associated with just one moderately complex burial. Using the SELGEM retrieval program, the same task was accomplished in less than twenty minutes by coding a set of indicators to test particular data parameters, a series of expressional tests on these indicators, and specific action instructions. A typical retrieval run would usually cost less than $2.00.

The retrieval program can also be used to select particular records (*i.e.*, artifacts or bags) as a function of specified criteria for subsequent tabulation by such statistical packages as SAS or SPSS. Here, variable comments would be excluded in favor of fixed format numerically coded attributes. The powerful SELGEM retrieval program can augment the more limited data selection capabilities of these other programs and permit more precise manipulation of the data base. Given the vagaries of archaeological data, this capability would be useful, for example, in including or excluding from further analysis, artifacts or variables as a function of a complex set of conditional criteria.

The reordering of data files is simplified by the SELGEM system. Only a few simple control cards were required, for example, to sort the computerized bag logs into order by square and, within squares, by layer. This facility is ideal, indeed essential, for a multi-disciplinary, multi-national research project where a number of specialists, working independently, need access to original field records.

There are a number of options for printing reports with the SELGEM system. The most sophisticated program allows the user to design the printing format and translate the numerically coded nominal variables into intelligible text. Output from this program can be processed with other programs to prepare camera ready copy for publication and reports or to produce inventory cards for a card catalog. If one were to cross index illustrations and photographs with the artifactual data, the time and effort needed to prepare, standardize, and correct artifact descriptions for publications could be significantly reduced (**Fig. 2**).

For archival purposes, the data base and the SELGEM system could be stored and updated periodically to insure compatibility with changing computer hardware and software. We hope to use SELGEM to compile a definitive bibliography of Southeast Asian prehistory as an adjunct of our research program. While other functions for SELGEM have been suggested, it is quite obvious that any DBMS is a double-edged sword capable of vitiating the research strategy of any unwary archaeologist. Defective data will invariably generate meretricious interpretations.

Moreover, a DBMS is not a panacea for the logistical problems of an archaeological research project. Indeed, one must be willing to pay for the benefits derived from SELGEM in a variety of ways, not the least of which is the fact that scholars will be better able to scrutinize (and criticize) the research methods as well as the conclusions. A decision to adopt computerized data processing techniques requires an irrevocable commitment of continued programming support, computer time, and competent workers through the various phases of the research program.

Conclusions

Our experience suggests that that under the direction of an archaeologist, the SELGEM system can be a dynamic and economical research tool. It can be implemented outside of an institutional context and adapted to the evolving configuration of the particular research program. A DBMS such as SELGEM can be used to structure a more critical perspective on archaeological data and the insights derived from the data as well as to facilitate their dissemination and preservation.

Acknowledgments

Dr. Chester Gorman and Pisit Charoenwongsa are the co-directors of the Northeast Thailand Archaeological Project. Financial support for the implementation and use of SELGEM was provided by the National Science Foundation (Grant number BNS77-14425). We gratefully acknowledge the prodigious efforts of the many work-study students and volunteers who have worked with us over the last three years.

References

Bennett, M. A., 1974: *Basic Ceramic Analysis. San Juan Valley Archaeological Project no. 1*, Eastern New Mexico University Contrib. Anthrop. 6, no. 1.

Bohrer, V. L., Adams, K.R., 1977: *Ethnobotanical Techniques and Approaches at Salmon Ruin, New Mexico. San Juan Valley Archaeological Project Technical Series no. 2.* Eastern New Mexico University Contrib. Anthrop. 8, no. 1.

Chenhall, R.G., 1975: *Museum Cataloging in the Computer Age* (AASLH, Nashville).

Creighton, R., n.d.: *SELGEM Basic Programmer's Documentation.* (Office of Computer Services, Smithsonian Institution, Washington), photocopied manuscript.

Doran, J.E., Hodson, F.R., 1975: *Mathematics and Computers in Archaeology,* 332 (Harvard Univ. Press: Cambridge, Mass.).

Gorman, C.F., Pisit, C., 1976: *Expedition 18* (4), 14.

Humphrey, P.S., Clausen, A.C., 1977: *Automated Cataloging for Museum Collections: A Model for Decision and a Guide to Implementation.* (Lawrence, Kansas: Association of Systematics Collections).

Neuner, A.M., 1976: *SELGEM Manual.* (Lawrence, Kansas: Association of Systematics Collections).

Schauffler, W., 1976: *Expedition 18* (4), 27.