PART III: DATA ORGANIZATION AND DATA BASE MANAGEMENT

4

Introductory Remarks on Data Base Management

CHRISTOPHER CARR

DATA BASE MANAGEMENT AS ANALYSIS

The term *data base management* often brings to mind the image of complex programming of a computer to *store* data. However, it involves much more than that. For projects that encompass large amounts of data of diverse classes, data base management (DBM) involves the careful consideration of the particular variables, kinds of observations, and kinds of relationships among them—of the infinite ones that are possible—that have potential analytic value within the paradigm of the researcher and that should be preserved for later analysis. This is true whether the information to be preserved is stored with computer technology or with written records, photographs, maps, or other storage techniques (Chenhall, 1982, p. 6). In the perspective of the previous two chapters, then, DBM is *the* initial step of analysis. It is concerned with defining the total data structure which has potential relevance to a broad problem domain, and from which certain aspects will later be selected deductively or inductively for analysis in order to investigate particular questions. To the extent that this initial step can considerably limit the kinds of variability and relationships that can be investigated later, as well as the quantitative means for doing so, it is of fundamental importance.

Once the data base manager has decided what variables, kinds of observations, and kinds of relationships should be preserved, the task more often associated with DBM remains: determining a means for storing such information within the physical constraints of computer technology so that it is preserved completely and accurately, and can be retrieved and added to with accuracy. The logic that is required for success in this aspect of DBM is analogous to that required in selecting and applying a statistical technique to data. Just as a statistical technique must be logically concordant with and sensitive to the relevant aspects of a data set that reflect the phenomenon of interest and its nature, so, too, the means by which data are stored on a computer must be logically concordant with and allow for the preservation of all information that is of potential interest and that reflects the phenomena of potential interest and their natures. In both cases, method must not *constrain* the form of the data as they are processed.

DBM, then, has two aspects. One is concerned with problem definition, specifying the phenomena of potential interest and their nature, and *modeling* the data entities and relationships that reflect those phenomena. The second is concerned with *preserving* that model. The segregation and stepwise nature of these two phases of DBM, as well as the top-down, theory-to-system approach to the design of DBM systems that this implies, is emphasized in the following chapter by Parker, Limp, and Farley. The authors suggest that the successful design of a DBM system requires the stepwise development of two models of the data: 1) a *conceptual model* which stipulates the entities, attributes, and relationships of potential interest, and 2) a *physical model* which specifies how that information is to be preserved. The conceptual model is the product of the data modeling phase of DBM, whereas the physical model is the logically equivalent transformation of the conceptual model, which is achieved through the use of principles of DBM (e.g., the process of normalizing a hierarchy of relationships).

The conceptual process that is advocated by Parker et al. for designing a DBM system differs to some degree from that described by Chenhall (1982). Both sets of researchers emphasize the necessity of carefully considering, prior to the design of a system, its purpose and the questions it is to answer. However, Chenhall organizes his operations of system development around the *research activities* in which the system is to participate; different files are established up-front for different activities. The data categories and relationships to be stored within each file are determined only after the initial inventory of files is specified (Chenhall, 1982, p. 5). In contrast, Parker et al. organize their operations of system development around the *nature of the data* that the system is to preserve; data modeling is fully completed before the consideration of files. The latter approach seems preferable in that the nature of the data base can play a somewhat more important role in the programmer's consideration of how to structure the system so that it can store and retrieve the desired information with the least storage and search costs.

The viewpoint that the structure of a DBM system should be logically concordant with the nature of the data to be preserved has naturally led Parker et al. to consider the general nature of archaeological data, and thus, the general structure required of an archaeological DBM system. Defining archaeological data in the broadest sense as the kind that is generated by multiyear, regional research and CRM programs, as opposed to site-specific data, alone (Brown et al., 1982; Gaines, 1974), Parker et al. suggest that archaeological data bases have a number of characteristics that pose management problems. In particular, they

1) are very large;
2) are complex, having multiple, overlapping relationships of both the object-clustering and attribute-clustering kinds;
3) involve diverse kinds of entities of different social scales (e.g., regions, sites, proveniences, artifacts) which are described by diverse kinds of attributes (e.g., soil type, edge angle) on different measurement scales (nominal, ordinal, interval);
4) define a matrix that has a high percentage of empty or zero cells;
5) are continually added to; and
6) are accessed in multiple, diverse subsets to investigate diverse hypotheses.

These characteristics, in combination, cause problems in storing, retrieving, and updating data in a manner that is accurate and complete, yet also parsimonious and economical. To overcome these management problems, Parker et al. advise the use of a *relational* DBM system, as opposed to one structured with pointers. Basic concepts of the relational approach are illustrated with the Arkansas Archeological Survey's AMASDA-DELOS system. Thus, the chapter by Parker et al. evaluates quantitative methodology for its logical concordance with data structure, as do many other chapters in this book.

In addition to the several themes mentioned previously, a number of other important points are made by Parker et al.

1) A physical data base structure that has logical concordance with the data to be preserved need not be a *natural* structure—a mirror representation of the data that involves empty cells and path dependencies. It can be an unnatural structure (e.g., a set of relational files) that has been transformed from the natural one to make storage and search economical.

2) Development of a conceptual model of the data that is to be preserved has *heuristic* advantages. It requires the researcher to consider relationships between variables and kinds of entities that might not otherwise be considered. (See also Chenhall, 1982, p. 2; Keene, chapter 10).

3) The common motivation behind the development of DBM systems and archaeological typologies is discussed. The limitations of the latter, in terms of data preservation and retrieval, are enumerated.

4) A DBM system can serve both inventory and research purposes. This is illustrated by the AMASDA-DELOS system. It is capable, for example, of inventorying the locations and dates of excavation of sites, yet also can retrieve ecological data that is useful in settlement modeling studies. (For a contrasting opinion on function segregation, see Chenhall, 1982, p. 2).

Finally, it is necessary to mention the importance of a DBM system of the encompassing kind that is discussed by Parker et al. The value of such a system extends beyond its capability of simply *preserving* information of potential relevance. It also extends beyond its capability of allowing the researcher to simply *extract*, in a purely deductive fashion, portions of a data base that are to be used in quantitative analysis when solving specific problems (Chenhall, 1982, p. 8; Brown et al., 1982, pp. 76-78). When used to its full potential, a DBM system can also serve as a *data screening device* within the context of an exploratory data analysis or constrained exploratory data analysis framework (see Carr, chapter 2). It allows the researcher to inventory the kinds of network relationships that do exist between variable states within the data set—of all those relationships that could be present in the data, as specified by the conceptual and physical models. In addition, it allows the researcher to determine the commonness of the network relationships that do occur between variable states. In this way, the researcher is provided with a summary of the nominal scale structure of the data base, which would be impractical to establish from a listing of the raw data, itself. Such a summary can be used to increase the efficiency with which more elaborate data screening procedures (e.g., crossplots, correlation analysis) are applied to the data at later stages of analysis. It can also be used to ensure that no significant relationships are overlooked (Farley, 1983).

REFERENCES

Brown, J.A., Clayton, S., Wendt, T., & Werner, B. (1982). The Koster project information retrieval application. In S.W. Gaines (Ed.), *Data bank applications in archaeology* (pp. 67-79). Tucson: University of Arizona Press.

Chenhall, R.G. (1982). Computerized data bank management. In S.W. Gaines (Ed.), *Data bank applications in archaeology* (pp. 1-8). Tucson: University of Arizona.

Farley, J.A. (1983). *Data base management, systematics, and their relationship to the analytical process in archaeology.* Unpublished master's thesis, University of Arkansas, Fayetteville.

Gaines, S.W. (1974). Computer-aided decision-making procedures for archaeological field problems. *American Antiquity 39*(3), 454-462.

For Concordance in Archaeological Analysis

BRIDGING DATA STRUCTURE, QUANTITATIVE TECHNIQUE, AND THEORY

Christopher Carr
GENERAL EDITOR

WESTPORT PUBLISHERS, INC.
in cooperation with the Institute for Quantitative Archaeology,
University of Arkansas

LIBRARY OF CONGRESS CATALOGING IN PUBLICATION DATA

Main entry under title:

For concordance in archaeological analysis.

Includes bibliographies and index.
1. Archaeology—Statistical methods—Addresses, essays, lectures.
I. Carr, Christopher, 1952- . II. University of Arkansas, Fayetteville. Institute of Quantitative Archaeology.
CC81.F66 1985 930.1'028'5 85-8861
ISBN 0-933701-00-4

Printed in the United States of America