Knowledge Base or Database? Computer Applications in Ethnology

SENRI ETHNOLOGICAL STUDIES 20 1987

Knowledge Base or Database? Computer Applications in Ethnology

LOUIS D. BURNARD
Oxford University Computing Service

A museum, its contents, their description and the totality of informed opinion about them, may be regarded as a single integrated information system. This paper proposes the use of information system modelling techniques as a necessary first step in the computerization of ethnological collections. A summary description of the widely-used ANSI/SPARC architecture and the Entity-Attribute-Relationship technique for conceptual schema design is followed by detailed accounts of three projects at Oxford University. The Beazley Archive database integrates descriptive, iconographic, cataloguing and attribution data concerning several thousand Greek vases in a single system. The Lexicon of Greek Personal Names provides an epigraphical index to all attested Greek names from the earliest times to the mid 7th century. Both these projects were converted from conventional flat files to their present form, in which specialised content addressing hardware is used to support a simple relational query interface for non-specialist users. The third project is the evolution of a complete conceptual model capable of supporting queries about all significant aspects of the conservation, accession, cataloguing and description of museum objects, currently being undertaken in collaboration with the Ashmolean Museum at Oxford. Because much semantic information is included in the relationships modelled, a high degree of expert knowledge may be built into the implemented database, sufficient, it is claimed, for this to be better referred to as a knowledge base.

MUSEUM OR INFORMATION SYSTEM?

One crucial difference between a museum and (say) a junk shop or a furniture repository is that the former may be regarded as a collection of information rather than simply of objects. Information is instantiated by the objects (just as the objects are given meaning by the information they convey) as much as it is represented in the catalogues and other indexes which describe them. Information is also present in the sum total of the expert knowledge concerning the objects and their descriptions instantiated by the museum staff's learning and the whole body of research on which it can draw. As new objects are added or catalogued and as opinions about existing objects and their interrelations shift and change, this information is subject to continual expansion and review; it is, however, rarely or never discarded.

If we regard the whole of a museum as an integrated system in which information is stored and processed, the task of modelling it within a computer appears at first sight rather daunting. Nevertheless it is no different from the task of modelling the complexities of any large organization, in some respects simpler. We should hope therefore that the tools and methodologies which industry commerce and government have evolved to deal with the problems of representing their versions of reality should be equally applicable to the problem of representing the informational contents and structure of a large museum.

The process of computerizing such large information-processing organizations as government agencies or commercial companies is now fairly well understood. Briefly summarized, it is necessary to:

(a) describe the meaning of the complex information structures which we perceive to exist in the real world and their interdependence: this, following ANSI/SPARC [ANSI 1975; TSCHIRITZIS and KLUG 1977], we call the *conceptual schema*;

(b) for this single conceptual schema, describe how its structure is to appear to one or more users of the computer held system which models or implements it: each of these descriptions corresponds with an ANSI/SPARC *external schema*; and

(c) determine the underlying physical structures necessary to support the union of all external schemata used for a particular conceptual schema: this corresponds with the ANSI/SPARC *internal schema.*

Three aspects of this architecture seem relevant to our present purposes:

(1) The central role of the conceptual schema, without which as a unifying principle the other two schemata are a mere collection of *ad hoc* rules and procedures.

(2) The independence of the conceptual schema from all implementation considerations, and the implied feasibility of a plurality of these.

(3) The fact that the conceptual schema is intended to capture the meaning of information rather than its representation.

The bulk of this paper describes our experience in using this methodology in such seemingly unusual application areas as ethnology or museum documentation. We begin therefore by attempting to justify the use of this method and by giving a brief outline of its chief characteristics.

THE CONCEPTUAL SCHEMA

End users of any type of computer system inevitably form some conceptual model of its structure, often based on rather hazy notions of its physical implementation. It is in our view essential that the conceptual view should not be constrained by such considerations, but rather should aim to reflect only the information needs of the users of the system; this, following [GRIETHUYSEN 1982], we call the *universe of discourse.*

This goal is however surprisingly difficult to attain; particularly in a research environment, a little learning can be worse than complete ignorance. The user whose experience of database has been obtained from currently available microcomputers often has the greatest difficulty in thinking of his informational requirements in non-physical terms. Even those users who have not succumbed to the lure of the floppy disk tend to see their information in terms of the particular data structure in which it has previously been physically instantiated (index cards, ledgers, textual description, etc.). Even when computerization has been embarked upon specifically in order to overcome some informational restriction inherent in the physical constraints of a manual system, it is not unusual to find users requesting that those very constraints should be eternally perpetuated in the very system which is supposed to do away with them. Inflexible and archaic manual cataloguing procedures are replaced by equally inflexible and archaic (but automated!) procedures, often with the added bonus of arcane mystification introduced "because of the computer."

Computerization is the best opportunity any information processing organization (whether a great museum or a humble research project) may ever have to look at itself, what it does and what it would like to do. The importance of the conceptual schema to this process is that it offers a neutral mechanism by which the information needs of the organization can be expressed, quite independently of both existing manual systems and proposed computer ones.

This is not of course to advocate a *new broom* policy. One reasonable requirement of the conceptual schema is that it should begin by describing all existing views of the universe of discourse, although it is often the case that during this process many inconsistencies and unnecessary constraints are brought to light. Only if the conceptual schema is defined reasonably correctly and reasonably completely can we be reasonably confident that systems implementing it will be capable of meeting any future requirement, as yet unanticipated. This cautious optimism is firmly grounded in the very abstraction of the conceptual schema from current requirements. The modelling process requires us to identify an idealised Platonic structure for our universe of discourse from which not only our current requirements but also all conceivable future ones may be derived. This extensibility is an important objective: if history tells us anything, it is that tools intended for one purpose are invariably used for another.

From the point of view of the end user, the process of conceptual schema design has a number of other benefits which (in our experience) greatly offset an initial impatience with its seemingly purposeless abstraction.

In the first place, the end user of a new system is involved from the start in the design process, and moreover on an equal (if not superior) footing with the system designer. Consequently he perceives the system as one over which he has control and the nature of which he has largely determined. The neutrality of the conceptual modelling method employed (which should appear to favour neither the end user's perceptions nor those of the computer system designer) is clearly of great importance here. Whatever approach is chosen (see the following section), its formalism must be comprehensible to both.

Secondly, the independence of the conceptual model from any one computer implementation has many obvious benefits. It reduces the system implementor's problems considerably by providing an implementation-independent specification against which any particular combination of hardware and software can be validated. The process of mapping the formalism of an agreed conceptual schema onto the formalism of a particular piece of software is well defined and can be repeated as systems change without perturbing the end user's view of the system. This can be particularly important in the research environment, where funding of hardware and software is frequently an eccentric and irrational business in which choices are made at least as often on political or emotional as on pragmatic grounds.

Finally, more than one end user has remarked on the usefulness *per se* of designing a conceptual schema, even if no computer implementation of it ever actually materialises. Particularly in research, the discipline of re-assessing the conceptual framework within which information is processed, if only in order to give it conscious expression to another person, can be of the greatest value.

Representing the Conceptual Schema

Several different formalisms have been proposed as means of representing the conceptual schema; a very useful ISO report [GRIETHUYSEN 1982] distinguishes three main approaches: the Entity-Attribute-Relationship (EAR) approach, the binary relationship approach and the interpreted predicate logic approach. Most current practitioners have chosen to adopt the first approach, or a variant of it, if only because it is best supported by currently available software. The binary relationship approach has, however, recently witnessed a resurgence of interest, with the availability of specialised *triple-processing* hardware [FROST 1982; JOHNSON 1984], while the predicate logic approach is clearly of relevance in the development of expert or rule based systems using Prolog and similar languages [KOWALSKI 1984].

The EAR approach is not without its critics: it is not best suited to situations where highly volatile information is involved, nor is its methodology fully defined in a mathematical sense [KENT 1979]. Nevertheless, systems employing variants on it have come to dominate the business of computing over the last five or six years. We have also found it to be applicable in cases where other advantages of the database approach are perhaps of less importance.

Standard text books such as [DATE 1981] naturally stress such elements of the database approach as shareability, support for multiple concurrent update, integrity, security or high-volume high-performance retrieval of data. These are often matters of little concern to the small research projects which are the bread and butter of university computing services. For us the chief benefits of the database approach are quite different: characteristically, as I have argued above and elsewhere [BURNARD 1980], lying in the simplicity and power of the EAR approach as a means of creating a conceptual schema. This can be of benefit even to a single user retrieval-only flat file system!

In the remainder of this paper I shall limit discussion largely to our use of the EAR approach in conceptual modelling. The external schemata (implemented conceptual models) discussed are based for the most part on ICL's data management products, in particular IDMS, a Codasyl standard network database management system [ICL 1983a], and Querymaster, its relational query language interface [ICL 1983b].

Designing a Conceptual Schema (Figs. 1, 2)

The process of constructing a conceptual schema is essentially one of categorization and examination of the constraints determining what is meaningful within the universe of discourse. The method is simply to identify the *entities* of interest within an information system, their constituent *attributes* and the *relationships* between them. An attribute (such as "age") is said to have *value* (say, 28) which may be drawn from a particular *domain* (say, "positive integers between 1 and 100"). It is customary to represent the schema graphically by a network in which the nodes represent entities and the arcs relationships.

In the approach we have adopted attributes are bound to particular entities, and are not therefore normally included in the graphical representation of the model. If PERSON is an entity and OBJECT is an entity, although both may have an attribute AGE, these are considered to be different attributes, and may be drawn from different domains. Where this is not the case, it is probable that the attribute common to the two entities would more profitably be regarded as an entity in its own right, associated with the other two by some relationship. If PERSON has an attribute COUNTRY-OF-BIRTH and OBJECT an attribute COUNTRY-OF-ORIGIN, it might be useful to identify COUNTRY as an entity in its own right, with relationships BORN-IN associating it with PERSON and MADE-IN associating it with OBJECT.

A relationship expresses some perceived association between entities. More strictly, distinguishing between entity types (*e.g.*, PERSON, TOWN) and entity occurrences (*e.g.*, "Belinda", "Oxford"), we may say that where particular entity occurrences are meaningfully associated (other than by virtue of being of the same entity type) and the association can be given a name, then an occurrence of some relationship-type exists. For example, we may express the fact that Belinda lives in Oxford as an instance of the relationship-type LIVES-IN occurring between entity types PERSON and TOWN.

The two aspects of a relationship which are conventionally represented in a conceptual schema are its *cardinality*, that is, whether one or more than one (represented by the crows foot symbol) instance of an entity participates at each end of it, and its *optionality*, that is, whether participation in the relationship is a necessary condition for the existence of an entity (solid arc) or not (broken arc).

In Fig. 1, the relationship LIVES-IN has 1 : n cardinality as many people live in one town but a person can only live in one town; consequently there is a crows foot at the PERSON end of the relationship arc. As a person cannot exist without living in

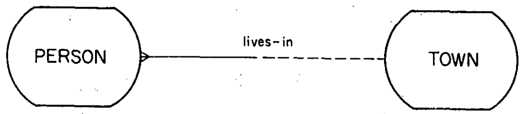


Figure 1. Two Aspects of Relationship: Cardinality and Optionality

some town, the end of the arc nearest PERSON is solid; as a town can (in this model at least) exist with no people living in it, the TOWN end of the arc is drawn with a broken line.

Note that this is an example of the LIVES-IN relationship; if (as is more probable) we were concerned with the HAS-LIVED-IN relationship, its cardinality would be different: Belinda might have lived in many different towns. Some approaches, *e.g.* [CHEN 1976], permit the definition of attributes associated with relationships (*e.g.*, the date Belinda came to live in Oxford); in our experience again it is simpler to regard this as a new entity type (say, RESIDENCE).

As a further simplification, only binary (that is, involving two entity types) relationships are used, n-ary relationships being better represented by additional entities. Thus, while we could express the fact that Belinda attributed the "Last Supper" to Giotto as an instance of a ternary relationship ATTRIBUTION defined between the entities PERSON, PAINTING and ARTIST, it is more convenient to regard this as an instance of an entity type ATTRIBUTION to which a PERSON is related by an instance of the ATTRIBUTOR relationship, an ARTIST by the relationship ATTRIBUTEE and a PAINTING by an instance of the relationship ATTRIBUTED to (see Fig. 2). We do not, however, exclude reflexive relationships, (for example, the relationship COPY-OF between one instance of a PAINTING and another), although in practice, these also often indicate hidden entities.

Note incidentally that in this model an attribution can exist without being related to some person as ATTRIBUTOR.

Several full tutorial accounts of this method are to be found (*e.g.*, [ROBINSON 1981; TSCHIRITZIS and LOCHOVSKY 1982]) and a good summary of its chief characteristics is given in Appendix D of the reference [GRIETHUYSEN 1982].

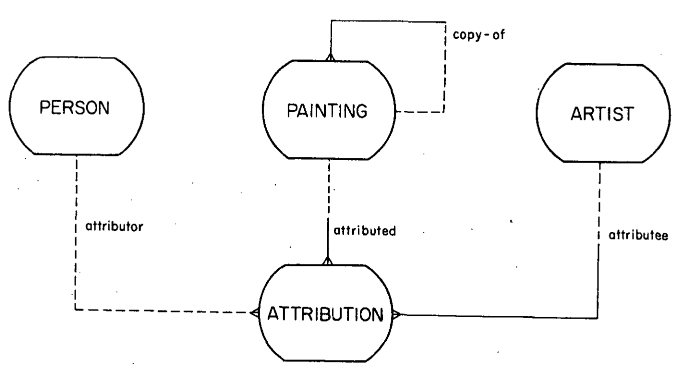


Figure 2. Simplification of N-ary Relationships by Introducing Additional Entities

DATABASE APPLICATIONS AT OXFORD UNIVERSITY COMPUTING SERVICE (OUCS)

At OUCS we have been using this method for the past five years. The first system built with its aid was an IDMS database used to record the breeding patterns of a large population of wild birds just outside Oxford [McCLEERY and PERRINS 1985]. This was followed by a system which supports the administration of a national clinical trials service [UK-TIA 1979]. Both of these systems are still in use, having survived one change of mainframe and about five upgrades of operating system without needing any change in their original design. They are in some respects unlike the bulk of database applications at OUCS, which have been drawn increasingly, over the years, from projects in the Humanities, perhaps reflecting an overall change in the user population at our centre.

Unlike their scientific colleagues, British researchers in the humanities typically do not cooperate, except on really large projects, and even then collaborators will usually insist on more autonomy than would be normal in a scientific project of comparable size. This characteristic would appear to contradict one of the fundamental tenets of the database approach: that data should be shared. On the other hand, the information handled in a typical humanities research project is often far greater both in volume and in complexity than that of a scientific project. Indeed, the complexities of some projects are such that it can be quite difficult to persuade researchers that they can be properly understood by another human being, let alone a computer.

These considerations both indicate the usefulness of a preliminary conceptual analysis, for even the smallest (in terms of personnel) project. The project described in [BURNARD 1980], for instance, involved a single doctoral student and is not atypical of several others in which we have found this approach to be useful.[1]

Recently we have embarked on a series of larger projects which have a common ethnological interest and it is these which I propose now to describe. Two of these (the Beazley Archive and the Greek Lexicon) have been to some extent computerized *ab initio*, but have now encountered limitations inherent in their original design, while the third (the Ashmolean Project) is a *green fields* project in which the first serious mistakes have yet to be made. The first of these projects is now in production mode, while the load-up of the second has recently begun, testing of its prototype having concluded successfully in June, 1984. The third, which is also the most ambitious, is still at the design stage.

The Beazley Archive

The Beazley Archive, which is housed in the Ashmolean Museum at Oxford, is one of the largest collections of information about Greek vases in the world. The

1) Here and elsewhere, the word "we" should be taken to imply a reference to my colleague Paul Salotti, without whose active and critical collaboration many of the projects described would never have seen the light of day.

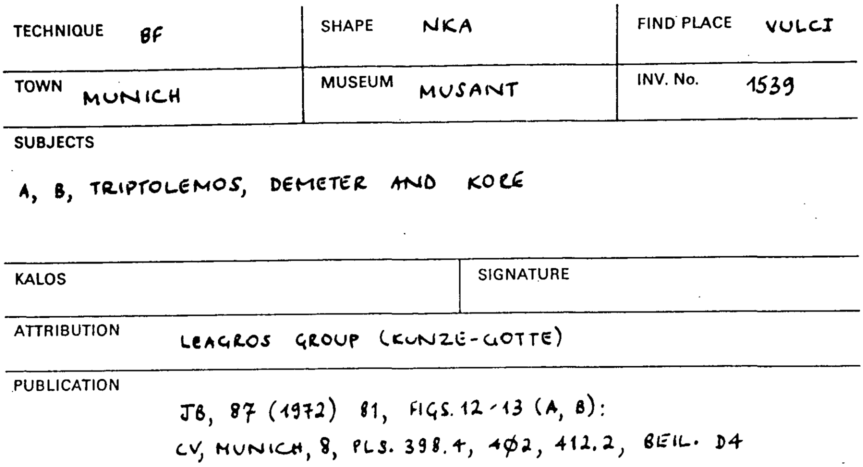
core of its collection consists of photographs, drawings and notes bequeathed to the university by Sir John Beazley, but it has been greatly augmented by other scholars. In 1979, under the direction of the Beazley Archivist, Dr. Donna Kurtz and the Lincoln Professor of Classical Archaeology, John Boardman, the Archive began systematically collecting published citations, both to Greek vases which had previously been described in one of Beazley's publications [BEAZLEY 1956; 1963; 1971] and to those which had not. The former category were simply collected for publication in book form as the Beazley Addenda [BURN and GLYNN 1983] while the latter were held in computer files, to which access by scholars throughout the world was offered by the Archive's staff. At present about 8,000 vases fall under this second category; a further 30,000 are described in the Beazley publications which will also, eventually, be added into the system.

Information held about each vase or fragment comprises a coded description of its form, provenance, cataloguing history, iconography, details of any attributions made for it, and, of course, the bibliographical citations which were the original *raison d'être* of the project. An earlier paper [GLYNN 1983] describes the system as originally implemented. It is perhaps typical of many other projects where a manual card-based system has been automated more or less unchanged.

The software used in this initial project was a well known package called Famulus [PROWSE 1983] which is widely used at many British universities and museums [BARTLE and COOK 1982] and is typical of many conventional file-based systems. Data are categorised by a tag, just as in a manual system it might be by a labelled position on a record card. Data entered in one field cannot, however, be related to data entered in another. Software of this type usually (though not inevitably) supports only sequential access to the underlying physical records, and so retrieval of information involves either a search through the whole file or the construction and maintenance of separate indexes for each category of information thought to be of interest. In short, the only entity modelled by such systems is the record card itself, not the information which that record card was originally intended to capture.

In the case of the Beazley Archive, this had led to a number of problems, chiefly having to do with consistency. If we examine the sample record given by Glynn [GLYNN 1983], reproduced in Fig. 3, it is evident that several fields (that is, attributes in our terminology) contain values drawn from domains which are quite large. TECHNIQUE is fairly simple: a manageably small set of codes for this and for SHAPE had been defined. Even in the SUBJECTS field it had proved feasible, if difficult, for the Archive to keep to a controlled vocabulary, given its expert knowledge of the subject. However, with fields such as MUSEUM and PUBLICATION it was proving quite difficult to maintain consistency in the abbreviations used to refer to named items occurring perhaps very infrequently in the whole database, without frequent recourse to hand-made checklists which were inevitably out of date.

A further oddity of card-driven systems like this one is that some of the fields actually contain different attributes, drawn from different domains. Thus, ATTRI-



TECHNIQUE BF	SHAPE NKA	FIND PLACE VULCI
TOWN MUNICH	MUSEUM MUSANT	INV. No. 1539

SUBJECTS

A, B, TRIPTOLEMOS, DEMETER AND KORE

KALOS	SIGNATURE

ATTRIBUTION LEAGROS GROUP (KUNZE-GOTTE)

PUBLICATION JB, 87 (1972) 81, FIGS. 12-13 (A, B):
CV, MUNICH, 8, PLS. 398.4, 402, 412.2, BEIL. D4

```
IDEN 99999
TECH BF
SHAP NKA
FNPL VULCI
LOCN MUNICH; MUSANT; 1539
SUBJ A,B, TRIPTOLEMEOS, DEMETER AND KORE
ATTR LEAGROS GROUP (KUNZE-GOTTE)
PUBL JB, 87 (1972) 81, FIGS. 12-13 (A, B):
     CV, MUNICH, 8, PLS.398.4, 402, 412.2, BEIL. D4
```

Figure 3. Sample Record Card in the Beazley Archive

BUTION contains both the name or description of an artist or potter and the name of a scholar, but only human intelligence can tell which is which, by interpreting the parentheses. Similarly, the beginning of the SUBJECTS field includes a code for the part of the vase decorated, drawn from quite a different domain from that of the rest of the field. Finally, in some fields (but not others) it is possible to have a number of values one after another, each with quite a complex structure. Again, only human intelligence can determine where one such sub-field stops and the next begins.

These practices are, of course, all perfectly acceptable in an information system such as a card index or a book which is to be processed by human intelligence. Human beings understand parentheses and semicolons and know that "A, B" is unlikely to refer to a picture on a vase, just as they know that "ARV" is an abbreviation for some publication while "xxiv" is not. To the computer, however, in the absence of more precise information, everything within one field is equally meaningful (or meaningless), and the whole of each field must thus be processed as a homogeneous string of characters. If the nature of the data to be processed is not accurately discriminated, then any search for particular items will involve the scanning and rejection of much irrelevant material. The intelligence which a human being would bring to bear on the search (*e.g.* that if you are interested in SUBJECTS on side "A" and the current subject field does not begin with "A", then there is no need to look any further at it) is not necessarily available.

The immediate causes of the Archive's decision to redesign their system, however, related more to functional limitations of the Famulus package. For example, when recording the LOCN field for a new entry it was essential to check that no entry already existed with the same value for that field. This would indicate that the vase in question had already been catalogued and that the new citations should therefore simply be tacked onto the PUBL field of an existing record. This checking could be done, however, only by maintaining an up-to-date sorted index of all LOCN fields and checking each new candidate against it by hand. The expense and nuisance of keeping this index up-to-date was increasing sharply as the database grew bigger, and the index to it therefore more necessary. The simple logistical problems of maintaining a file of 8,000 Famulus records in the absence of specialised software or staff were also beginning to cause grave problems.

In October, 1983 we therefore embarked on the process of converting the archive's file of Famulus citations into an IDMS database. A conceptual model was designed in close consultation with staff who had worked on the original system and would be using the new one; a pilot system was tested during the early months of 1984, and the first full system went live in May of the same year.

In defining the conceptual schema and the IDMS (external) schema derived from it, the opportunity was taken to include in the structure as many semantically significant relationships as possible. These not only simplified the construction of a good interactive update program, but, as we shall see, also had considerable impact on the retrieval program used to interrogate the resulting database.

Referring again to Fig. 3, we can categorize the information on the record card as concerning the following real world entities:

A vase (TECHNIQUE, SHAPE)
Places (FIND PLACE, TOWN)
A museum or collection (MUSEUM)
A catalogue number within a museum (INV. NO.)
A description of the decoration on a vase (SUBJECTS)
Various types of inscription on a vase (KALOS, SIGNATURE)
An artist or potter (ATTRIBUTION)
A scholar (ATTRIBUTION)
A periodical or other publication (PUBLICATION)
A reference within a periodical (PUBLICATION)

We can further identify relationships amongst these entities, as follows:

(1) Many vases may be associated with the same place as FIND PLACE (but each vase may have only one FIND PLACE—or none at all).
(2) A vase may have many catalogue numbers in different museums. Likewise there may be many references to it in different publications.
(3) A scholar may have made attributions of many different vases, possibly to many different artists. The same vase may have been attributed to many different artists.

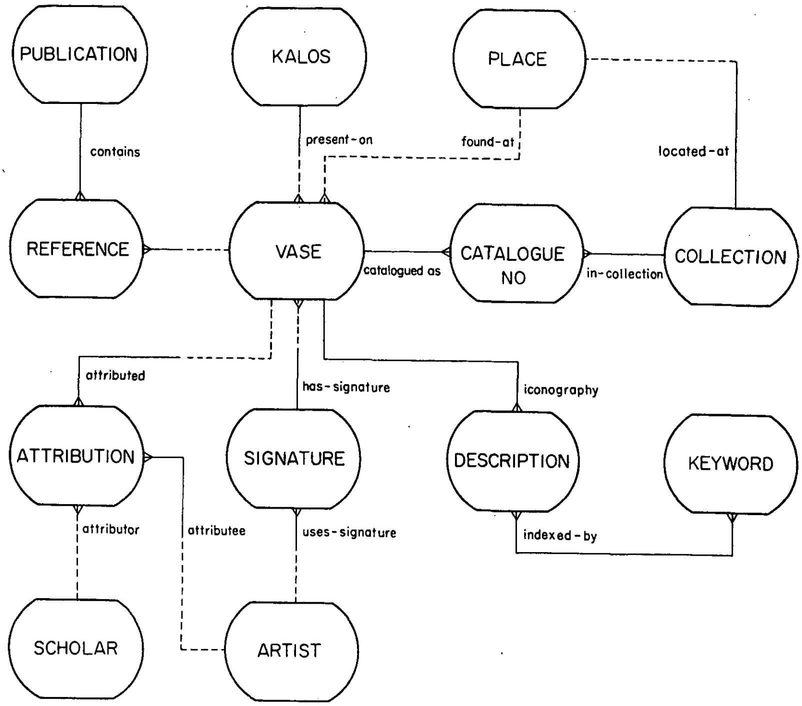


Figure 4. Beazley Conceptual Schema

(4) An artist may be associated with a vase, either by virtue of an attribution made by some scholar or by means of a signature on the vase.

(5) A particular key word may appear in many descriptions and each description may contain more than one key word.

(6) The same KALOS (a stylised inscription) may appear on different vases.

These and other relationships are summarized in the conceptual schema represented in Fig. 4.

Our hypothesis, that any meaningful question about the universe of discourse represented by the given data can be expressed (and hence answered) as a traversal of this network, has not yet been falsified.

The second stage in the construction of the database was the definition of external and internal schemata appropriate to this agreed conceptual model. As in the other projects we have implemented, one external schema was an ANSC standard network schema [ANSC 1984b] as implemented by ICL's version of the widely used IDMS package [ICL 1983a]. A second external schema uses a relational query processor Querymaster [ICL 1983b] to provide a simpler user interface. Full details of the implementation are beyond the scope of this paper, but the following general points may be of interest.

IDMS implements storage structures which are very similar to those of the EAR conceptual schema: entities can be mapped directly to IDMS *records*, attributes to *data-items* and relationships to *sets* as a first approximation. Although the concept of domain is missing, it can be implemented by means of an additional record and set type. Thus in our example, the acronyms used to prefix a catalogue number must correspond with (*i.e.*, be derived from the same domain as) the mnemonic key of some existing COLLECTION record. IDMS offers facilities to enforce this rule, and also to allow it to be overridden when storing the first catalogue number from a previously unknown collection. The controlled vocabulary used for the iconographic descriptions in the database is similarly constrained by the system.

Information concerning each distinct entity occurrence is stored only once, thus both enforcing consistency and providing a built-in index facility which is kept up-to-date automatically. For the sets used to represent relationships, pointer chains, in combination with the keyed access facilities of IDMS, usually allow performance of certain access paths to be optimised if necessary. (Again, the details of this are beyond my present scope.)

The third stage in the construction of the database was the thankless task of reprocessing the existing data files so that individual constituents of the various fields could be identified and reconstituted according to the storage structures required by the new external schema. At the same time, the opportunity was taken to correct various errors and inconsistencies in the spelling of names or the style of bibliographic citations. This stage, largely complete by the end of 1983, required a number of special purpose programs and much patience.

During the load-up of the backlogged data, an interactive program was written (using Fortran77 and a version of the ANSC Data Manipulation Language) to enable Archive staff to amend and add to the database. While commonplace in a conventional DP environment, the interactive mode of working came as a pleasant surprise to users accustomed to the tedium of batch-oriented systems like Famulus, in which the smallest alteration to one record might necessitate an overnight batch job to completely revise the whole file.

The sample dialogue given in Fig. 5 shows how the data for the record given in Fig. 3 would be entered using this program.

Note that much of the data (*e.g.*, museum codes, publication codes, etc.) can be automatically validated by the program on entry. When corrections are needed, they involve a minimal disturbance of existing data. All record management, including provision of automatic journalizing and security dumps, is carried out by IDMS.

For *ad hoc* queries of the database, we did not feel that the network external schema was satisfactory, largely because of its inherent procedurality; we therefore wished to provide as a second external schema something which would offer a full relational view of the conceptual schema.

The relational view of data is perhaps even better known than the network view. [DATE 1981] includes a classic statement of it and ANSC is currently debating a proposed standard Relational Data Language to accompany its standard Network Data Language, based loosely on the well known IBM product SQL [ANSC 1984a].

```
 WHICH VASE? 99999
No vase exists with that ID
 Did you intend to create a new vase record?Y
 Technique       BF
 Subsidiary tech
 Shape code      NKA
 Condition
 Any (more) cataloguings to add for this vase?Y
 Serial no.      0
 Collection      MUNA
 Catnumber       1111
 Any (more) cataloguings to add for this vase?N
 Any (more) attributions to record for this vase?Y
 Artist ID?LEAGR
 Scholar name?KUNZE-GOTTE
 Artist qualifier
 Details
 Any (more) attributions to record for this vase?N
 Any (more) publications for this vase?Y
 Publication ID?Jb
 Reference detail?87(1972) 81, FIGS 12-13 (a,B)
 Any (more) publications for this vase?Y
 Publication ID?CV
 Reference detail?Munich, 8, pls 398.4, 402, 412.1, Beil. D4
 Any (more) publications for this vase?n
 Any (more) decorations to add for this vase?y
 Location code?A,B
Enter description line by line ending with ****
 TRIPTOLEMOS, DEMETER AND KORE
 ****
 Any (more) decorations to add for this vase?N
 Any index entries to add for this vase?N
Any Kalos, signature or provenance data should be entered as modifications
 ENTER COMMAND:MP
 Provenance?VULCI
Old provenance:Unknown
New provenance:VULCI
 ENTER COMMAND:DV
Details of vase 99999    at 13.55.52 on 08 DEC 1986
-------------------------------------------------------
  Technique : BF                    Shape  : NKA
  Provenance : VULCI
Cataloguing history
-------------------
0 MUNA1111          Munich                  Antikensammlungen

Publication record
------------------
 Jb       87(1972) 81, FIGS 12-13 (a,B)
CV       Munich, 8, pls 398.4, 402, 412.1, Beil. D4

Attributed toGR  LEAGROS GROUP (   ) by KUNZE-GOTTE

Decorated area : A,B
TRIPTOLEMOS, DEMETER AND KORE
=============== End of vase details ===============
 ENTER COMMAND:XV
```

Figure 5. Sample Dialogue for Entering the Record Card in Fig. 3

In this type of external view, the entities of the EAR model are seen as *tables*, made up of *columns* corresponding to attributes and rows corresponding to entity occurrences. Where columns of two tables contain values drawn from the same domain, it is possible to *join* rows from the two tables wherever the values of the common attributes are equal, and thus provide a relationship between the two tables or entities.

The relationships of the EAR model, regarded in the relational external schema as dependency rules, determine how rows from different tables may be combined. Because in our model the relationships were chosen to represent semantically significant connections between data items, any system using these relationships appears to exhibit a degree of intelligence. It becomes possible, in fact, for the user to access a database in terms of the original EAR model. The benefits of this capability need not be elaborated.

Determining an internal schema capable of supporting both relational and network schemata is non-trivial. The decision to use IDMS as the first external schema had, however, effectively determined the nature of the internal schema. IDMS storage structures permit the optimisation of some relationships (those implemented by pointer chains) at the expense of others (those requiring a physical scan); a relational interface must, however, be capable of supporting all relationships equally well. In general, relational database systems have had to trade off relational completeness against performance, or inflict on the user such decisions as whether or not to maintain an index, which have little to do with his universe of discourse. In our case, these problems did not arise, largely because the ICL Content Addressing File Store (CAFS) was available [MALLER 1979].

Unique to ICL equipment, CAFS is a simple component added to a normal disk file controller, which allows effective sequential searching by the disk drive hardware rather than by the central processor. A search for up to 16 key values can be carried out in parallel, at a speed limited primarily by the reading speed of the physical device; at our installation, it is up to 1 Mb per second. Because retrieval is performed before records are passed back to the central processor, the processor time required for complex searches of large databases (notably joins on non-keyed items) drops dramatically. This is particularly true where part of the access path involves a sequential search through the database for particular values which are not indexed. In our experience, which is not unusual, a query which does not use a key but does use CAFS will normally take somewhere between 1 and 10% of the processor time needed for the same query without using CAFS. In elapsed time, comparable or better improvements are also obtained for users sitting at terminals [ICL 1984].

The software interface initially provided for this device was a relational query processor called Querymaster [ICL 1983b], which could be used with IDMS databases as well as with conventional ISAM files. Querymaster is unlike most relational query processors in the extent to which it is non-procedural. That is, to select columns from different tables, the user usually needs to specify only the names of the columns required and any conditions on which values are required. Querymaster itself will work out how the required relational joins are to be made. For example, in response to the query

LIST VASE-NUMBER WHERE COLLECTION-ID="XYZ"

Querymaster needs to display rows from the VASE-NUMBER column (in the VASE table) which can be joined to rows in the CATNO table which can themselves be joined to rows in the COLLECTION table. These joins can be made using the IDMS set types corresponding with the relationships shown in Fig. 4. The joined rows can, at the same time, be scanned rapidly by CAFS to pick out only those rows in which the COLLECTION-ID column has a value of "XYZ". Similarly, the query

LIST ARTIST-NAME, SCHOLAR-NAME, VASE-NUMBER...

requests columns from three tables (ARTIST, SCHOLAR and VASE). It is unambiguous, because there is only one way in which the three corresponding entities can be related (via an ATTRIBUTION). In the case of an ambiguous query, such as

LIST ARTIST-NAME, VASE-NUMBER...

Querymaster has to decide whether to list ARTISTS related to VASES by SIGNATURE or by ATTRIBUTION. It will normally choose the path involving fewest record retrievals, following a variety of implementation-specific heuristics; if the path lengths involved are identical (as they are in this case), its choice will be arbitrary, but consistent. The syntax also permits the user to specify which relationship is to be used if the default is inappropriate. It is interesting to compare this method with more procedural relational systems such as SQL, where the user must always specify the route to be followed when tables are related together.

One additional feature of the Querymaster/CAFS combination worthy of note is its support for text items. Many system designers, faced with the semantic complexity of this type of data, have taken the line of least resistance by treating all data as pure text, possibly subdivided by topic, but with no other structure. As should be apparent, we have always resisted this temptation, but this decision does not mean that text items should never appear. In the Beazley Archive, the iconographic descriptions should clearly be treated as text: to express all the complexities of the various mythological and iconographic elements which might figure in vase decorations would clearly be wasted effort.

In the conceptual model, an entity KEYWORD is related to the entity DESCRIPTION. At an earlier stage of design, a recursive relationship on KEYWORD was proposed to give a thesaurus capability, but was in the event considered unnecessary. As implemented in the IDMS schema, this relationship became a useful term index, permitting the retention of full textual descriptions in a format suitable for display or output, but indexed by rigorously controlled (and normalised) key words.

With the availability of CAFS searching facilities on the textual description itself, however, it became apparent that the key word index was to some extent redundant. The CAFS text-searching facilities can perform very fast string searches for individual words or parts of words within a string. Approximate string matching, which is also possible, allows the user (to quote recent ICL publicity) to "find a noodle in a hatrack."[2] Of more significance to us is the ease with which complex searches can be expressed. For example,

LIST VASE WHERE TEXT-WORD="ACHILLES", "TORTOISE"

will list all the vases whose descriptions include references to either Achilles or the Tortoise.

LIST VASE WHERE TEXT-WORD="ACHILLES" AND TEXT-WORD="TORTOISE"

will recover only those vases on which both Achilles and the Tortoise are depicted. To achieve the second case using the key word index, the user would need to form the intersection of the two lists himself; when access is made via the text item, this operation is unnecessary.

The Lexicon of Greek Personal Names

Since 1972 an ambitious research project under the direction of Mr. P. M. Fraser (All Souls College, Oxford) has been accumulating records of all Greek personal names attested from the earliest historical times up to the mid 7th century AD. The intention is to replace the Lexicon of Pape-Benseler [PAPE and BENSELER 1911], now invalidated by the enormously increased epigraphical evidence produced by a century of archaeological excavations. This project, which is funded by the British Academy, is now approaching the end of its first phase, in that publication of the first volume of the new Lexicon by Oxford University Press is due in mid 1985 [FRASER 1976; ASHPLANT 1983].

From the earliest stage, the computer was used as an aid in managing the vast amounts of data involved in the project. Like the Beazley Archive, the Greek Lexicon began by using Famulus, in their case as a means of capturing and manipulating conventional citation slips rather than museum catalogue cards.

One added complication was that the data necessarily included large amounts of Greek, which had to be transliterated using special coding for accents, breathings, reconstructed letters, etc. This, in fact, proved to be a comparatively simple task once a sufficient body of expertise in using the scheme had been accumulated. The acquisition of specialised terminals capable of both displaying and manipulating Greek characters, which would be the first requirement of any comparable project starting today, was not even envisioned as a possibility ten years ago. Such is the pace of change in computer technology.

Lexicon staff continued to amass and record citation slips using the Famulus package for many years, more or less independently, with little need to merge the resulting files. As this need became apparent, however, special purpose checking programs were developed to assist in the process of integration and validation, but these rapidly proved to be inadequate for the sheer bulk of data involved—10,000 different names associated with 60,000 different people and a total of 90,000 references

2) Quoted in *Punch*, April 11, 1984.

are estimated for the first volume alone. Accordingly, in the summer of 1983, the Lexicon embarked on a conversion process similar to that of the Beazley Archive, as described above.

The original Famulus files had used four fields only, containing the name of a person, a date (either of a person or of an inscription), a place name and a reference. This last field also contained large amounts of miscellaneous details about the person, such as variant spellings of the name, career details sufficient to distinguish him from anyone else, editorial apparatus, etc.

In discussion with Lexicon staff it became apparent that the conceptual model underlying the system in fact distinguished quite sharply between a person and a name. The same name, or orthographic variation on it, might refer to many different people, while the same person might be referred to in different contexts by different names. Although the primary purpose of the project was to produce in book form a catalogue of names, these names were only of importance as they related to identifiable people. The same name for two different people would generate two separate entries in the printed Lexicon, while of all the names associated with a given person only one would be chosen as the primary name to which all other entries would refer.

This many-to-many relationship between NAME and PERSON suggested the existence of an intermediate NAMELINK entity, with such attributes as the particular reconstruction or accentuation of the name and its role (primary, *trianomina* etc.) for the person. An occurrence of this link entity exists for each name used for a particular person.

Perhaps because the first volume of the Lexicon was to deal with material from the Aegean islands and Cyrenaica, geo-political considerations had considerable impact on the design of the conceptual schema. In it a PERSON can be related to one or more particular geographic entities (TOWNS or DEMES, hierarchically grouped within ISLANDS or analogous larger geographical units) at different times. Each occasion on which a person is attested as residing at a particular place is considered to be an occurrence of the RESIDENCE entity. The POLIS (that is, political or tribal grouping) with which that residence is associated will depend on the date associated with the RESIDENCE.

For example, an inhabitant of the island of Telos at a date subsequent to the annexation of that island by Rhodes would be related to the POLIS "Rhodian Telioi," whereas a person whose residence predated the annexation would have no such connection.

In the original files, the dates of various events had been recorded in a wide variety of styles intended to capture the varying degrees of exactness which could be attached to them. Although appropriate to a printed volume, such forms as "c. 3 BC", "3–7 AD", "4th c.", or "521 BC" obviously could not be presented to the computer as coming from the same domain. It was therefore agreed that for each such date both an earliest possible and a latest possible date would be calculated during the conversion process, both being stored as integers in the range 1 (for 1000 BC) to 2000 (for 1000 AD). The text forms of the date are additionally retained for printing out. The earliest date for a person is additionally used to identify a YEAR entity with which he is associated, as a coarse grouping similar to the coarse grouping by location provided by the RESIDENCE entity.

The original data also included fairly sparse information about other entities, notably family relationships between people, their occupations and careers, and occasional references to the scholars responsible for the more tenuous attributions. All but the first of these were regarded as of too little significance to be included in the final model, but the existence of information about family groupings was considered to be of sufficient interest to warrant definition of a reflexive RELATED-TO relationship in the conceptual schema. This enables the system to support queries such as "List all those people whose fathers are known to the system or who share a common parent". The remaining miscellaneous material is retained simply as textual commentary, logically associated with the appropriate entity occurrences. The complete conceptual model is reproduced in Fig. 6.

Implementation of this system followed more or less the same pattern as that of the Beazley Archive as described above; it is not therefore further described here. One important difference in the uses to which the system is to be put, however, is that the printed volumes of the Lexicon, when they finally appear, will be produced directly from the database. A retrieval program currently (July, 1984) under test will traverse the database in the appropriate sequence, inserting typesetting codes and thus produce an input file for the OUCS Lasercomp typesetting system from which camera-ready copy of the highest quality can be generated. A sample page is given in Fig. 7.

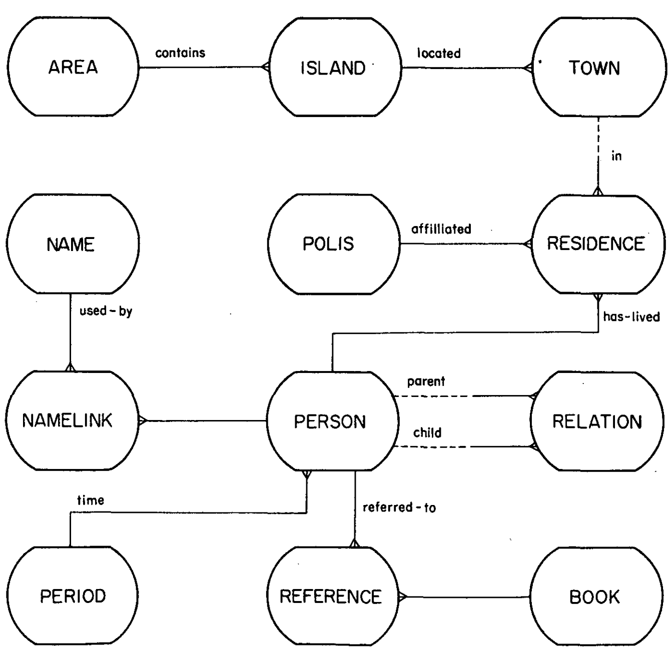


Figure 6. Lexicon Conceptual Schema

Retrieval from IDMS; program as at August 4th 1984

Ἡλιόδωρος 1 Ἡράκλεια

Ἡλιόδωρος
LESBOS:
—MYTILENE: (1) c.245–236 BC *IG* IX 1^2 (1) 25, 25 (s. *Ξένων*)

Ἥλιος
CRETE:
—GORTYN: (1) ?iv AD *IC* 4 p. 366 no. 356
RHODES:
—LINDOS: (2) c.170 BC *IG* XII (1) 819, 23 (f. (nat.) *Τελεσίδας*)
TENOS: (3) imp.? *IG* XII (5) 987 (s. *Σαρπήδων*)

Ἦλις
KOS: (1) ?i AD PH 250, 11 (*Γ. Ἰούλ. Ἡ.*)

Ἡλοθάλης
KOS: (1) vi BC *RE* Supplbd. 4 (f. *Ἐπίχαρμος*)

Ἤμειρα
CRETE:
—RHIZENIA: (1) f.iii BC *IC* 1 p. 300 no. 17 (d. *Δερδωπίδας*)

Ἡμέρα
THASOS: (1) i/ii AD *Ét. Thas.* 5 p. 157 no. 321 (d. *Ἀσκληπιάς, Ζεῦξις*)

Ἵμερος
RHODES:
—KAMIROS: (1) c.258 BC *TCam* 8, 7; 23, 13 (s. *Ἄγριος*)
——Loxidai: (2) c.12 BC ib. 3 *Δc*, 52 (s. *Ἱεροκλῆς*); (3) f.i AD ib. 4 *b*, 5 (f. *Ἴσανδρος*); (4) ◇ ib. l. 27 (s. *Ἴσανδρος*)
—LINDOS: (5) c.180–170 BC *ILind* 172, 1, 12 (s. *Παυσανίας*); (6) 86 BC ib. 293 *c*, 32 (s. *Νικασίμαχος*); (7) 55 BC ib. 324, 15 (f. *Ἵμερος*); (8) ◇ ib. l. 15 (s. (nat.) *Ἱεροκλῆς*, s. *Ἵμερος*); (9) 27 BC ib. 378 *b*, 70 (I f. *Ἵμερος* II); (10) ◇ ib. l. 70 (II s. *Ἵμερος* I)
——Argeioi: (11) 22 AD ib. 419 III, 150 (f. *Παντακλῆς*)

Ἥμων
CHIOS: (1) iv/iii BC D.L. iv 34

Ἡνατίων
CRETE:
—BIANNOS: (1) iii BC *AAA* 6 (1973) p. 112 (s. *Εὐρύμαχος*)
—GORTYN: (2) ii/i BC *IC* 4 p. 312 no. 260 (s. *Ἑρμίας*)

Ἡνίοχος
IMBROS: (1) iv/iii BC *IG* XII (8) 85 b, 41
TENOS: (2) 200–168 BC *IG* XII (5) 919–20 a (f. *Κλεινομάχη*); (3) ◇ ib. 921 (s. *Ἰατροκλῆς, Ναύσιον*); (4) ii BC ib. 911, 10; (5) ◇ *SEG* XIV 553 II, 34

Ἠπίη
THASOS: (1) iv BC *Ét. Thas.* 3 p. 301 no. 83 (d. *Φίλων*); (2) i BC–i AD? *SEG* XVIII 343, 1, 14, 19 etc. (d. *Διονύσιος*)

Ἤπιος
SAMOS: (1) c.380 BC Barron p. 208 no. 145
TENOS: (2) ii BC *SEG* XIV 553 II, 44

Ἠπίς
CHIOS: (1) ?iv/iii BC Zolotas p. 250 no. *KA*

Ἡραγόρας
CHIOS: (1) iv BC *SEG* XXII 511, 1 (f. *Εὐήνωρ*); (2) m.iv BC ib. 508 A, 3–6 (f. *Ἀναγόρας, Ἀναξίδημος*); (3) iii BC *FD* III (3) 226, 10, [12] (f. *—χος*)
KALYMNOS: (4) ii–i BC *GVI* 946 (f. *Ξενοκλῆς*)
——Pothaia: (5) iii BC *TCal* 216 a (I f. *Ἡραγόρας* II, f. *Δαμόξενος*); (6) ◇ ib. loc. cit.; 216 c-f (II s. *Ἡραγόρας* I, f. *Ἡραγόρας* III, f. *Ἀβροστράτη, Δαμόξενος*); (7) ◇ ib. 216 f (III s. *Ἡραγόρας* II, s. *Φίλτιον*); (8) c.200 BC ib. 85, 10 (s. *Δαμόνικος*)
KOS: (9) c.366–300 BC *BMC Caria* p. 196 no. 32 (*Ἡραγ[όρας]*); (10) ii–i BC *GVI* 946, 8 (f. *Ξενοκλῆς*); (11) c.82 BC *IG* XII (8) 260, 16 (s. *Πραξίφαντος*)
RHODES: (12) ii–i BC *ASAA* 2 (1916) p. 136 no. 2 a, 7 (*AE* 1915, p. 128 no. 1 A, 7) (*[Ἡρ]αγόρας*); (13) m.i BC *BMC Caria* p. 254 no. 264 (=?= *Ἡραγόρας*); (14) m.ii BC *RE* 110; Nilsson 223; *BUST* 17 (1963) pp. 90–1 figs. 5–6; *EAD* XXVII sub E 37 + p. 290; (15) i BC *IG* XII (8) 186 a, 12 (f. *Θεύδωρος*); (16) m.i BC *Coll. Gillette* no. 198 (=?= *Ἡραγόρας*); (17) c.68 BC *IG* XII (1) 46, 276 (s. *Παυσανίας*)
—LINDOS: (18) c.180–170 BC *ILind* 172, 4 (f. *Ἄνδρων, Εὐφράνωρ*); (19) c.170 BC *IG* XII (1) 819, 22; (20) 65 BC *ILind* 308, 33; (21) 43 BC ib. 346, 51 (s. *Λυσίμαχος*)
——Klasioi: (22) 27 BC ib. 378 *b*, 125, 178 (I f. *Ἡραγόρας* II, f. *Ἡραγόρας* II); (23) ◇ ib. l. 125, 178 (II s. *Ἡραγόρας* I, s. *Ἡραγόρας* I)
SAMOS: (24) f.iv BC *IG* II^2 6417, 5 (s. *Ἡρόδοτος*)
THASOS: (25) a.340 BC *Ét. Thas.* 4 nos. 696–8; (26) f.iii BC *IG* XII (8) 286, 11 (s. *Λάμπων*); (27) s.iii BC ib. 293, 33; 302, 10 (s. *Ἀριστείδης*, f. *Ἀριστείδης*); (28) m.ii BC ib. 300, 14 (s. *Φείδιππος*); (29) imp.? ib. 627 (s. *Μέστος*); (30) ◇ ib. 492 (f. *—ουσα*); (31) ◇ *BCH* 91 (1967) p. 602 no. 53 (s. *Ἀριστόδημος*); (32) ◇ *Ét. Thas.* 5 p. 133 no. 252 (f. *Καλλιτύχη*); (33) i AD ib. p. 114 no. 218, 3 (s. *Εὔφριλλος*); (34) f.i AD ib. p. 107 no. 204, 11 (f. *Ἀρχέλεως*); (35) ii AD *IG* XII (8) 589, 1 (s. *Νικάδης*); (36) ◇ ib. l. 6 (s. *Ἀρχέλεως*); (37) ?s.ii AD *Ét. Thas.* 5 p. 144 no. 291 (s. *Δολᾶς*); (38) ?iii AD ib. p. 112 no. 212, 2

Ἡραγόρης
SAMOS: (1) vi BC *AM* 87 (1972) p. 131 no. XVII (*DGE* 715 4) (s. *Πυθαγόρης*); (2) v BC *AM* 31 (1906) p. 416 n. 1 (I f. *Ἡραγόρης* II, f. *Ἡραγόρης* II); (3) ◇ ib. loc. cit. (*[Ἡρ]αγόρης*: II s. *Ἡραγόρης* I, s. *Ἡραγόρης* I); (4) v/iv BC *AD* 11 (1927–8) *παραρτ.* p. 32 no. 6 (f. *Πασίγνωτος*); (5) m.iv BC *SGDI* 5719 (I f. *Ἡραγόρης* II); (6) ◇ ib. loc. cit. (II s. *Ἡραγόρης* I)
THASOS: (7) ?c.490 BC *Ét. Thas.* 3 p. 36 no. 6; (8) iv BC *IG* XII Suppl. p. 161 no. 396 (f. *Φανόκριτος*); (9) f.iv BC ib. XII (8) 279, 8 (f. *Δηϊθράσης*); (10) iv/iii BC *Ét. Thas.* 3 p. 272 no. 34, 39; *IG* XII (8) 297, 7; XII Suppl. p. 165 no. 429 (s. *Μενέμαχος*); (11) f.iii BC ib. p. 161 no. 391, 9 (s. *Σώπολις*); (12) i AD *BCH* 86 (1962) p. 594 no. 15 (f. *Νικαία*)

Ἡραθλίων
THERA: (1) vi/v BC *SEG* XXV 935 (s. *Ἡγησίλοχος*, f. *Κυλίφακτος*)

Ἡραιάς
CRETE:
—ARKADES: (1) imp.? *IC* 1 p. 25 no. 47 (m. *Αὔγη, Ἀπολλώνιος*)

Ἡραιεύς
LESBOS:
—MYTILENE: (1) iv BC *IG* IV 1^2 121, 120

Ἡραῖος
LESBOS:
—ERESOS: (1) 306–301 BC *IG* XII (2) 526 d, 20
—MYTILENE: (2) c.300 BC *IG* XI (4) 594 (s. *Ζωΐος*); (3) ii/i BC ib. XII Suppl. p. 25 no. 77 (s. *Ἀπελλῆς*)

Ἡραῖος
KOS: (1) ii BC *BCH* 86 (1962) p. 275 no. 4, 9 (s. *Ἀλέξανδρος*)

Ἡραις
AMORGOS:
—MINOA: (1) imp.? *GVI* 297
THASOS: (2) i BC–i AD? *IG* XII (8) 430 (d. *Φείδιππος*)

Ἡραισκος
CHIOS: (1) i BC *IEK* 74 (f. *Παράσιος*)

Ἡραισχος
LESBOS:
—MYTILENE: (1) imp.? *IG* XII Suppl. p. 26 no. 94 (f. *Μακαρία*)

Ἡρακλᾶς
LESBOS:
—MYTILENE: (1) imp.? *IG* XII (2) 481, 1–2 (f. *Ἐπαφρόδιτος*)

Ἡρακλᾶς
AMORGOS:
—MINOA: (1) imp.? *IG* XII (7) 339 (f. *Γέμελλος*)
KOS: (2) imp.? PH 245 (s. *Ἐπιθύμητος*)
THASOS: (3) imp.? *BCH* 91 (1967) p. 604 no. 56 (f. *Ἐλπίς*)

Ἡρακλέα
CRETE:
—SETAIA: (1) ii–iii AD *IC* 3 p. 165 no. 2

Ἡράκλεια
SAMOS: (1) ii/i BC *GP* 3620 (d. *Μελίτη, Ξοῦθος*); (2) imp.? *AM* 25 (1900) p. 211 no. 124 (-λησ-: d. *Ἄλτης*)

Figure 7. Retrieval from Lexicon

Once publication is complete the database will be maintained and, finances permitting, updated as new name variants or additional references are discovered. Whether or not a bureau service similar to that planned by the Beazley Archive will be made available to interested Hellenists remains to be determined.

It is gratifying to be able to quote the following vindication of our approach to this type of data management problem, which appears as the conclusion of an internal report by one of the Lexicon's staff:

> The conversion of the Lexicon data to an integrated system has been a very useful and salutary experience, chiefly in defining domains for the different records and forcing decisions on the exact meaning of certain pieces of information. It should extend the usefulness of the data beyond the printed publication. [RAHTZ 1984]

MUSEUM DOCUMENTATION IN THE UK

Perhaps surprisingly, there is no common policy on computerization amongst large museums in the UK. This contrasts with the situation in other countries: in the USA, for example, a number of systems tailored specifically for museum use exist, and there is even a museum computer network which publishes its own informal newsletter (*Spectra* published by Museum Computer Network Inc.). In France, the Louvre has developed a very successful system based on the Honeywell text-searching package MISTRAL [GUICHARD 1981].

In the UK, a British Library-funded report [PORTER, LIGHT and ROBERTS 1977] proposed a unified approach to the computerization of museum catalogues as long ago as 1977, and a Museums Documentation Association (MDA) was set up at the same time. This organization currently provides a number of invaluable services to member museums, ranging from standard record cards to a complete computer package, called GOS [PORTER 1982]. But it has been unable to achieve much in the way of standardisation, despite a number of relatively successful applications of GOS, *e.g.* [NEUFELD 1981; PRICE 1984]. Another British Library-funded report [BARTLE and COOK 1982], which surveyed existing computer applications in British Archive Services in January 1983, found at least eighteen different and largely incompatible systems in use.

The one characteristic shared by all these systems (including, though to a lesser extent, GOS and its derivatives) is an inability to represent structural information of the type I have been describing. The general rule is to use text-based systems such as Famulus which, as discussed earlier, effectively model a representation of the real world rather than the real world itself.

In the case of GOS, although relationships between attributes of records may be defined, they may only be structured hierarchically. For example, the attribute PROVENANCE might have subcategories COUNTRY and PLACENAME, the latter being further subdivided into REGION and TOWN, etc. These subcategories may not, however, participate in more than one hierarchy. Consequently, although attributes such as COUNTRY may be common to many entities, they cannot be subcategories within more than one GOS category. Instead, indexes must be declared and maintained independently of the record structure. The indexing facilities provided are often very sophisticated (including such useful features as fuzzy matching and the definition of synonyms or related terms) but the expense of maintaining indexes independently of a database can be high. Furthermore, in our view, the absence of any implementation-independent data description component greatly reduces the usability of such systems.

The Ashmolean Project

In 1983 the Ashmolean Museum celebrated its 300th anniversary, with much pageantry and no little justifiable pride in its status as one of the oldest established and most significant of university museums. Its four major departments—Western Art, Antiquities, Eastern Art, and Coins and Medals—now have an international reputation which I need not labour here. No doubt it was purely coincidental that the Department of Antiquities should decide to begin its fourth century by investigating the feasibility of computerization. This initiative came not from the Accessions Department of the Museum, nor from its central governing body (which is a logical rather than a physical object), but from Conservation Laboratory staff, who had already tried both GOS and Famulus but found them inadequate for their purposes.

For a variety of reasons, we shared their lack of enthusiasm for the existing software options. In April 1984, we embarked on a project to define a conceptual schema, both as a convenient means of evaluating the capabilities of existing software and as the basis for a prototype system which could be implemented with locally available software such as IDMS. Since discussion of this conceptual schema is still continuing,[3] it would be unwise to assume that the version of the model presented here is in any sense definitive.

The Ashmolean Model

The fundamental entity in our model is the OBJECT. This may, of course, be any sort of museum object (a painting, a statue, a piece of furniture, etc.) which is treated as a unit for purposes of storage, conservation or cataloguing. These three activities are also a simple way of subdividing our model into three submodels, each centred on the OBJECT entity. These submodels represent information about

(a) an object's history as a museum piece: its previous owners, cataloguing details, etc.

(b) an object's conservation: any physical changes or observations of it made while in the care of the Museum.

(c) an object's description: the sum of opinions as to its original purpose, origins, etc.

---

3) It is a pleasure to acknowledge here the stimulating effect of our lengthy discussions with Gwyn Miles and Dr. Helen Whitehouse, of the Ashmolean Museum.

Although these three submodels are discussed independently below, it should be remembered that a primary design objective was to integrate them into a single information system. Although they may seem independent, the information produced by or processed within one submodel is invariably relevant to the others. For example, an analysis carried out to conserve an object may produce new data authenticating (or invalidating) the attribution of that object to a particular time or place. Equally, the success of a particular treatment may be crucially dependent on a knowledge of the object's previous environment or history, which is often regarded as an integral part of its description. This is particularly true of the Ashmolean, which still retains such curiosities as "Guy Fawkes' Lantern" and "Pocahontas Mantle." We have tried to represent in the full model all of these interrelationships as thoroughly as possible; it is only for clarity that I discuss the three submodels independently.

THE OBJECT HISTORY SUBMODEL (Fig. 8)

The entities in this submodel are all concerned with events or perceptions concerning the object's existence as a museum object. Once the basic problem of establishing the object's identity had been overcome, such concepts as its ownership, published references to it, and the place where it is currently stored are simple enough to model.

The problem has two aspects: sometimes the same physical object is given a number of different names and sometimes the same name refers to different physical objects. It is commonplace for an object to be given different names at different times, particularly if its ownership changes; it is also not uncommon for the physical constitution of a given named object to change over the years. For example, a statue described in 1756 as having a head may since have lost it. The head in question may have been fixed to a different (formerly headless) statue in the early 19th century. Clearly, there are three distinct entities involved here (the two torsos and the head), yet one entity (the 1756 catalogue number) identifies two of them at one time and one only at another. And what was the other torso called before it gained its head?

Our solution to both classes of problem is to identify two distinct entities: the *object* and the *object-name*. Relating the two is an intermediate entity, the *event*, which may also be related to an instance of a PERSON entity. The OBJECT-NAME entity has such attributes as catalogue number, title or description, while the OBJECT entity has such attributes as physical description, dimensions, current location, etc. For simplicity we assign each OBJECT a unique identifier quite independent of any catalogue number it may have. The PERSON entity represents any agency outside the museum; its attributes include name, address, etc. The EVENT entity has attributes date and event-type, the latter indicating what has caused this particular mapping between the logical OBJECT-NAME and the physical OBJECT. Typical event-types might be "accession" or "constitution."

The use of this mapping may be demonstrated by the following not untypical scenario.

(1) In 1802, the Duke of Blankshire bequeaths to the Museum a large statue known as "The Blankshire Venus." Not to be outdone, his neighbour Earl Asterisk bequeaths the famous Asterisk Marbles, some 200 assorted and uncatalogued pieces.

At this stage, we have the following entities:

OBJECT-NAMES:	(1)	1802.1000, The Blankshire Venus
	(2)	1802.2000, The Asterisk Marbles
PERSONS:	(1)	Duke of Blankshire
	(2)	Earl Asterisk
OBJECT:	(1)	No. 1000 Wt. 550 Kg, Ht. 2.4 m, etc.
	(2–201)	Nos. AM1–AM200
EVENTS:	(1)	1802 accession. Related to OBJECT (1), PERSON (1) and to OBJECT-NAME (1).
	(2–201)	1802 accession. Each related to OBJECT-NAME (2) PERSON (2), and a different one of the 200 OBJECTS Nos. AM1–AM200.

To identify all OBJECTS acquired from Earl Asterisk, we select all EVENTS of type "accession" related to PERSON (2), each of which will also be associated with both an OBJECT and an OBJECT-NAME (possibly the same one). To identify all catalogue-numbers (*i.e.* OBJECT-NAMES) associated with OBJECT(x) we select all EVENTS of type "accession" or "recataloguing" associated with OBJECT(x), each of which will be associated with some OBJECT-NAME.

(2) The celebrated Prof. Marmi-Perduti having gained wide acceptance for his theory that the left arm of the Blankshire Venus is in fact spurious, the Museum is prevailed upon in 1840 to remove it from the statue.

This gives rise to two new OBJECTS, say Nos. 1001 (the now armless Venus) and 1002 (the spare, possibly spurious, arm). It also gives rise to two new EVENT entities (with date 1840, type "constitution"), one associating OBJECT (1002) with a new OBJECT-NAME ("Arm, formerly attached to the Blankshire Venus"), and the other associating the existing OBJECT-NAME (1) with OBJECT (1001), the now armless Venus. A third EVENT (type "former-constitution") is also created, associating OBJECT (1002) with OBJECT-NAME (1000). In the absence of any restriction on the date applicable to the information, any request for information about OBJECT-NAME (1) must now be interpreted as a request for information about OBJECTS 1000, 1001 and 1002.

(3) The equally celebrated savant M. Marbres-Perdus has meanwhile been browsing through the Asterisk Marbles and discovered (he assures the Museum) the true original left arm of the Blankshire Venus in the form of OBJECT No. AM42. Amidst much popular rejoicing, in 1846 the new arm is fixed in place.

This also creates a new OBJECT (say No. 1010) and two new EVENTS (date 1846, type "constitution"). One of the EVENTS is associated with the existing OBJECT 1001' and the other with the new OBJECT 1010. Both EVENTS are associated with OBJECT-

NAME (1). The date attribute of the existing EVENT record associating OBJECT AM42 with OBJECT-NAME (2) is changed to indicate that the relationship is no longer current.

Note that OBJECT AM42 may not be removed, even though the corresponding entity in the real world has ceased to exist, being now a part of the Blankshire Venus. In our model, no entity occurrence can be removed as long as some other entity occurrence is related to it; therefore, as long as the information in one or more EVENTS is required, the OBJECTS associated with those EVENTS must also be retained.

Without this structure it would be quite difficult to answer such questions as "What catalogue numbers have been given this object during its history?" or "Which objects were known by this catalogue number at a particular time?" Of course a text-based system could achieve this simply by indexing the individual catalogue numbers wherever they occured, but it might prove difficult to make much sense of the resulting multiple references.

Once the basic disticntion between OBJECT and OBJECT-NAME had been established, the rest of this submodel presented few problems, the only matter controversy being whether the remaining entities should be directly related to OBJECTS or indirectly via an EVENT record (Fig. 8).

STORE and LOAN are both directly associated with an OBJECT, being both concerned with the physical whereabouts or origins of an object. A STORE is any location (shelf, cupboard, gallery, etc.) within the museum where OBJECTS may or should be stored. The current design does not support information about the former locations of objects within the Museum: if this were required, the STORE-to-OBJECT relationship would have to go via EVENT. A LOAN relates a particular OBJECT to a PERSON (or

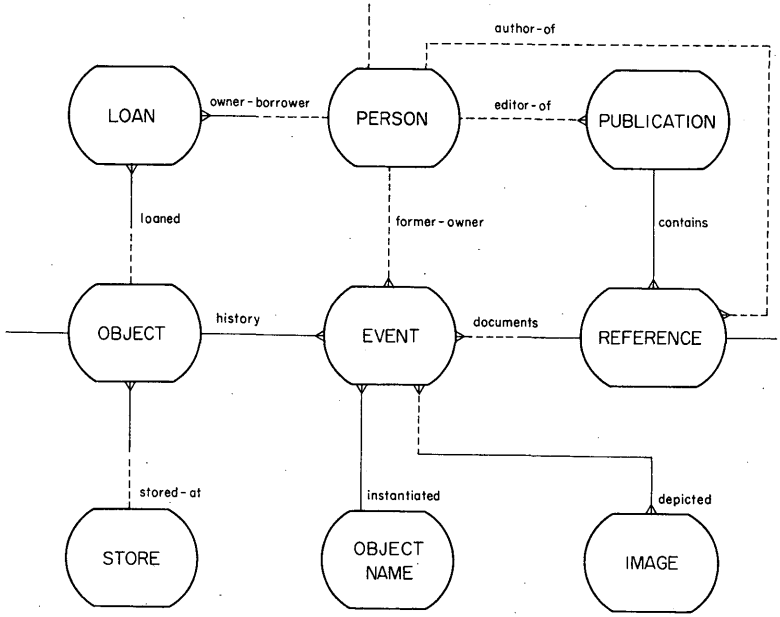


Figure 8. Ashmolean Object History Submodel

INSTITUTION) which has lent it to or borrowed it from the Museum. Although previous LOANS probably are of interest, because physical objects are the subject of the transaction, LOANS are directly related to OBJECTS. There is a case to be made for treating a LOAN as a special category of EVENT.

An IMAGE is any representation of an OBJECT, such as a drawing, a photograph or a cast. Although this is clearly related to a particular physical object, we have chosen to link it indirectly via the EVENT entity. Although an IMAGE necessarily shows the appearance of an OBJECT at some point in time, it is useful to be able to know the OBJECT-NAME to associate with the IMAGE. A drawing of "The Blankshire Venus" could represent one of several different objects, depending on its date. Relating it to the EVENT record nearest the date of the drawing automatically relates it to both the proper OBJECT and the proper OBJECT-NAME.

A REFERENCE is a published account of, or reference to, some OBJECT or group of OBJECTS. There is no reason to suppose that the group of OBJECTS referred to will necessarily correspond with those associated with any existing OBJECT-NAME, although this probably will normally be the case. No doubt several scholars, following M. Marbres-Perdus, have published accounts of hypothetical reconstructions in which one or more of the Asterisk Marbles participate; clearly these are references to groups of OBJECTS for which the museum has no OBJECT-NAME (other than that given by the REFERENCE itself). Rather than give every perception of one or more OBJECTS an OBJECT-NAME in its own right, we simply create an EVENT for every OBJECT referred to. More typically, however, when a REFERENCE is to some existing OBJECT-NAME, it must be associated with all the EVENT records appropriate to its date for the same reasons as those given above in connection with IMAGES.

A REFERENCE may be associated with two different PERSONS, one as an author of the reference itself, and one as editor or publisher of the PUBLICATION in which the reference appears. As with the Beazley and Lexicon systems, the distinction between PUBLICATION and REFERENCE is made largely for bibliographical convenience. No provision is currently made in our model for multiple authorship.

THE OBJECT CONSERVATION SUBMODEL (Fig. 9)

This submodel is concerned with all aspects of the physical treatment of the objects in the Museum's care. Three types of operation are distinguished: EXAMINATION, ANALYSIS and TREATMENT, each corresponding with an entity in Fig. 9.

An EXAMINATION may be a simple visual inspection or a complex process of disassembly and measurement. Its distinction from an ANALYSIS is that the latter, usually carried out on a small sample taken from the OBJECT, may be performed by some outside agency. Resulting from either (or both) is a MATERIAL-DESCRIPTION, which summarizes the material constituents of some or all of the OBJECT. As this DESCRIPTION, like others, will be key word indexed in any resulting implementation, we do not include as entities here the key words, which are fairly self-evident terms such as "bronze," "annealed," etc.

As the result of an ANALYSIS (or for some other reason), an OBJECT may undergo

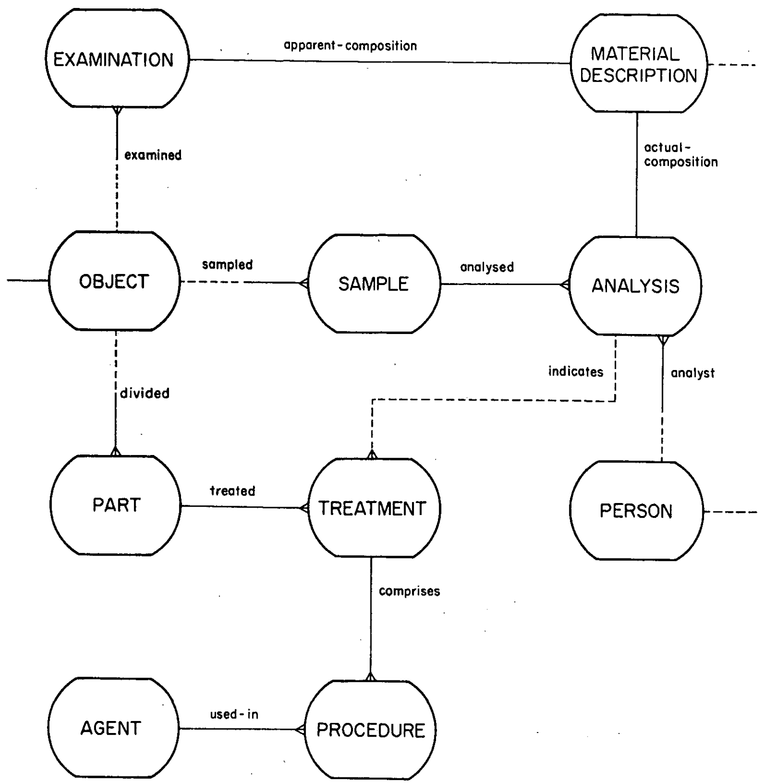


Figure 9. Ashmolean Object Conservation Submodel

some kind of TREATMENT, which comprises a number of PROCEDURES (*e.g.*, "washing," "polishing") in each of which a specific AGENT (*e.g.*, "methylated spirit," "pumice") may be involved. A particular TREATMENT is carried out on a specific PART (not necessarily the whole) of an OBJECT.

The Conservation Department did not enter the saga of the Blankshire Venus until 1850 when it was proposed that the statue might be loaned to the Great Exhibition. At this point an EXAMINATION was carried out and a MATERIAL-DESCRIPTION added to the database indicating simply that the piece was made of marble. It was decided that the whole statue should be thoroughly cleaned (a TREATMENT, involving two PROCEDURES, the first "poulticing" with AGENT "magnesium silicate," the second "polishing," with AGENT "talc").

In 1934, however, an X-ray scan of the statue carried out to determine how safely it could be moved, revealed that it contained iron dowels. For our purposes, the X-ray scan is considered as a type of ANALYSIS, the iron dowels being added as a new MATERIAL-DESCRIPTION for the OBJECT.

Without this type of structure, it is difficult to see how a straightforward text-based system can efficiently support such queries as "Which objects submitted for analysis have we not yet received material-descriptions for?", "Has this object already undergone some kind of anti-corrosion treatment?", "Have objects made of bronze usually reacted favourably to treatment with alkaline Rochelle Salts?" or "Which copper objects in the museum show evidence of annealing?"

Moreover, because OBJECT features in the Object History Submodel as well, it is possible to obtain information from both submodels. Thus queries such as "Which objects formerly belonging to Earl Asterisk are made of alabaster?" or "Give the current locations (STORE) of all objects treated with benzotriazole between 1970 and 1980" are simple traversals of the network. As we shall see, information from the third submodel can also be directly linked to information in this one, so that, for example, the attribution of an object to some time or place may be directly related to the chemical or physical analysis which gave rise to it.

THE OBJECT DESCRIPTION SUBMODEL (Fig. 10)

This submodel is concerned with those unchanging attributes of a particular OBJECT which are used to describe it, in order to place it in some cultural context, and thus identify its function or provenance. "Provenance" here is not to be taken as "the place where an OBJECT was before it entered the museum"; this we regard as a previous ownership, belonging to the Object History Submodel. Neither is the term used to refer to a particular site where objects have been discovered (usually by excavation); this we call a FINDSPOT.

OBJECTS may be associated with a FINDSPOT, and hence an excavator. We could have chosen to regard the excavation as an EVENT in the OBJECT's history, but have included it in this submodel, partly because of the self-evident relationship between the PLACE entity and the FINDSPOT entity, but mainly because its association with a site is generally regarded as an important descriptive element.

Every OBJECT has associated with it a single DESCRIPTION, which summarizes the current state of opinions concerning it, in the format one might expect to see in a catalogue. As opinions change, parts of this entity may be changed; it is used as a kind of blackboard. Independently of this description, a large number of ATTRIBUTIONS may be associated with the OBJECT, each of which (like the EVENT entities) is used to relate it to one or more of the other entities. These ATTRIBUTION entities are not removed, even when discredited.

As discussed above, an ATTRIBUTION may be regarded as a triple, linking an attributor, an attributee and an attributed OBJECT. In our current model, the attributor may be a PERSON (directly, or by authorship of some REFERENCE), an INSCRIPTION (*e.g.*, the artist's signature to a painting) or the MATERIAL-DESCRIPTION resulting from some ANALYSIS (*e.g.*, a radio-carbon dating). There may even be no attributor at all.

The attributee of an ATTRIBUTION may be one or more of a *person* (an artist or school, a manufacturer or workshop), a *time* (that is, some point between two distinct dates), a *period* (that is, a named range of times), a *place* (a geographic location) or a *culture* (some named political or ethnographic grouping).

When first acquired, the Blankshire Venus was described as a Greek original of the 5th century. An ATTRIBUTION entity was duly created, associating it with the

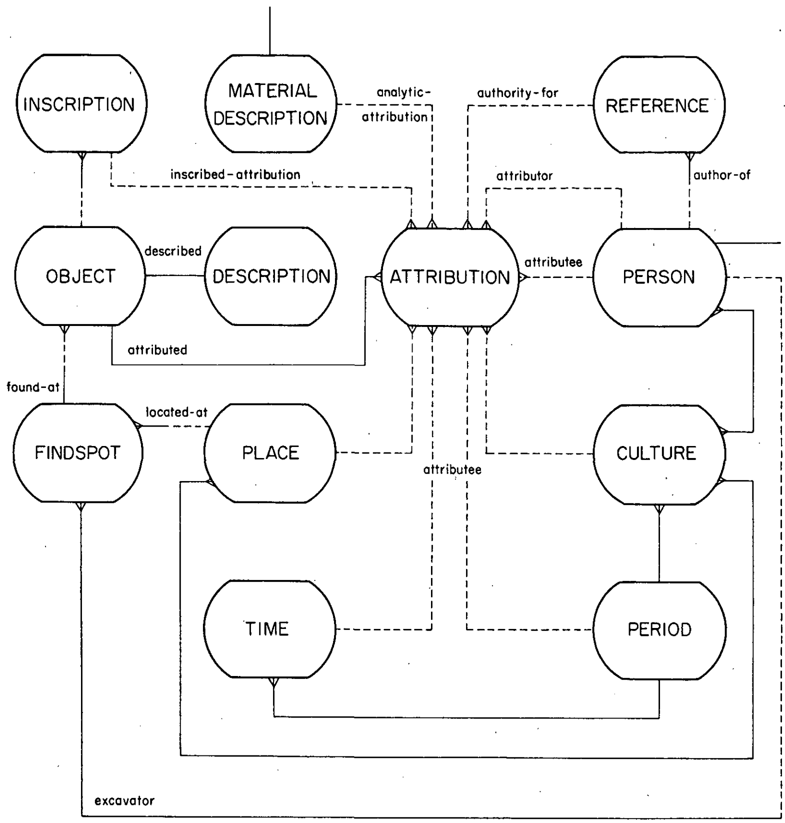


Figure 10. Ashmolean Description Submodel

PERIOD "classical," the PLACE "Greece," the TIME "−499 to −400" and the CULTURE "Attic." In 1878, however, Dr. Verlorenemarmore put forward the theory that the Venus was actually a second-century Roman copy of a lost Greek original. This gives rise to a new ATTRIBUTION, associating it with PLACE "Rome," PERIOD "100 to 199" and also with the PERSON responsible for the attribution. Note that no attributions can be made concerning the putative original in our model.

As with EVENTS, ATTRIBUTIONS are always retained, and perhaps include some sort of status indicator (an attribute with values drawn from a domain ranging from "currently accepted" to "barely credible," "unconfirmed," etc.). This indicator will naturally change as new evidence about an attribution accumulates. In 1962, for example, a lead isotype analysis of a sample taken from the Venus, revealing trace elements of a type associated with marble from the former Roman quarries at Aphrodisias in Turkey, lent considerable weight to Dr. Verlorenemarmore's hypothesis. A new ATTRIBUTION entity was also created, associated with both the MATERIAL-DESCRIPTION derived from the ANALYSIS and the PLACE "Turkey." No assumption can be made about any association between this PLACE and the Roman period, however, and consequently this ATTRIBUTION is not connected to any other of the attributee entities.

Although these attributee entities are all clearly interrelated in some sense, the connections are not easy to define. Even the relationship between a PERIOD and a number of TIMES is less simple than may appear: the PERIOD "Bronze Age," for example, implies three quite different TIMES in an African, a European and an Asian context. As a further level of complexity, there is no reason to assume that Professor X's "African Bronze Age" is coterminous with Professor Y's; indeed, if there were, much scholarship would grind to a halt.

At an earlier stage of the design process we had hoped that the CULTURE entity might be a helpful way of identifying particular combinations of TIME, PERIOD, and PLACE. For example, an ATTRIBUTION to "Minoan" would be all that was necessary to associate an OBJECT with the PLACE "Crete," PERIOD "Bronze Age" and TIME "2000 to 1500 BC." However, as Fig. 10 shows, a particular CULTURE may be related to many PLACES and (though restricted to one period) many different TIMES. To define a different CULTURE for each particular combination of TIME, PLACE and PERIOD would have resulted in a very large number of entity occurrences only dubiously present in the real world.

However, as more information is added to the system, so it should become easier to identify such patterns as do exist. When an attribution is associated explicitly with only a PERSON, the system will be able either to inspect other attributions for the same PERSON, to suggest candidate TIME or PLACE connections. Eventually it should be possible to use the relationships in the bottom half of Fig. 10, once a database of standard mappings between CULTURE and PERIOD, PLACE and TIME has been built up. The mapping data needed to support these interrelationships are implicit in the totality of existing attributions. One of the most useful aspects of the implementation of this system may well prove to be the definition of these relationships.

We do not include in this model purely art-historical or interpretive categorizations of OBJECTS (*e.g.*, "modernist," "Neo-Gothic," "Hellenistic"). Such terms arguably occupy one extreme of the same dimension along which terms like "Minoan" or "Scythian" may also be plotted, which we have chosen to regard as CULTURES. The notion of *school* (as in "Pre-Raphaelite School" or "School of Praxiteles") seems to us to be about half-way along this continuum, and is therefore the point at which we cease to regard such terms as grounds for an attribution to some entity which has or had objective existence in the world. Key word indexing on the contents of the DESCRIPTION entity should be adequate to support this type of query, although this and related implementation problems have yet to be confronted in earnest. It seems probable that some sort of hierarchic coding of iconographic interpretations of many objects will be desirable, similar to that used in some current British and German catalogues, such as those of the Marburg Collection (see, for example, [*Marburger* 1985]).

The complete conceptual model, as of August 1, 1984, is shown in Fig. 11. It is likely that the first external schema for the system will be an IDMS database front-ended by Querymaster using CAFS, as described above. Alternatively, less sophisticated software could be used to implement the system piecemeal on a number

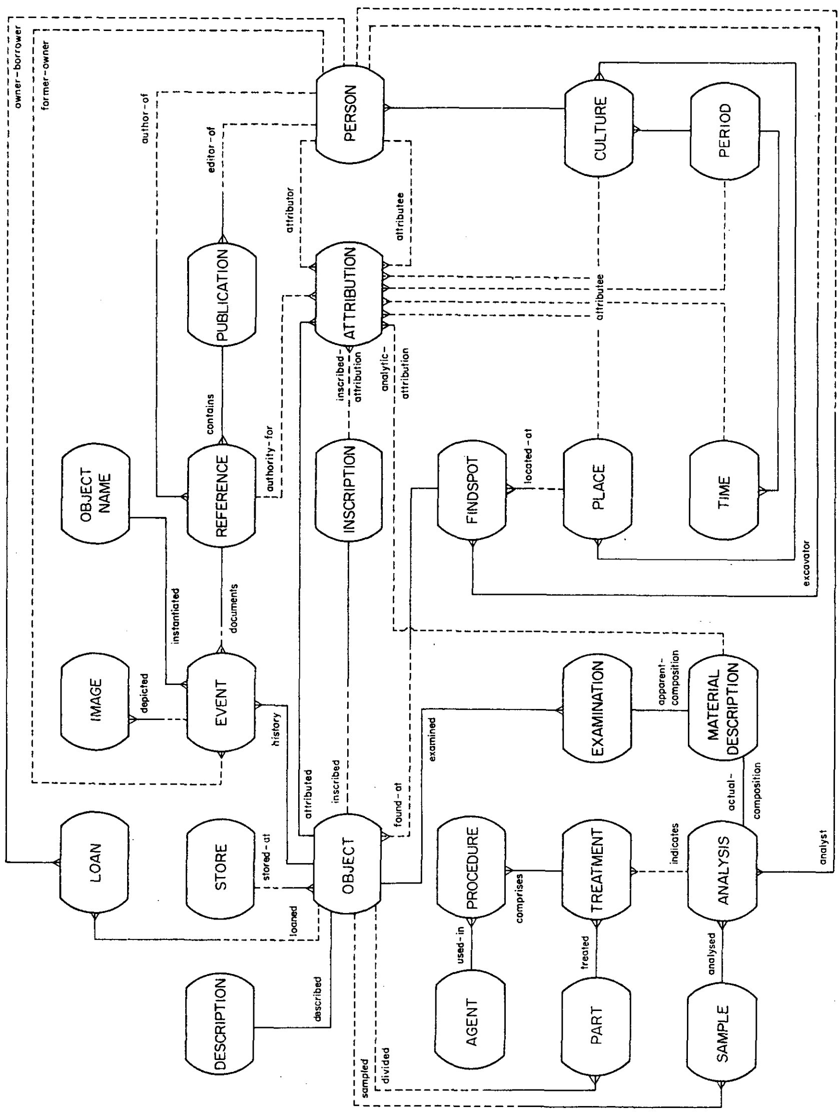


Figure 11. Ashmolean Conceptual Schema

of microcomputers. We are reasonably confident that whatever software or hardware solution the Museum eventually adopts, our approach has successfully helped them to define the problem.

CONCLUSIONS

The information systems of the past were distributed across three different categories of data management system: the human mind, the accumulation of human records and artifacts, and the interaction between the two, which gives meaning to both. We are familiar with methods of exploiting the computer's immense powers of symbolic manipulation for the second of these categories. Knowledge engineers and workers in artificial intelligence are beginning to show us ways of exploiting them for the first. I have tried to argue in this paper that tools for exploring the third category already exist. It may be easier than we think to transform a database into a knowledge base.

BIBLIOGRAPHY

ANSC

1984a *A draft proposed relational database language.* ANSC (X3H2-84-24), ANSI.

1984b *A proposed network database language.* ANSC (X3H2-84-48), ANSI.

ANSI-X3-SPARC Study Group on Database Management Systems

1975 Interim report 75-02-08. *ACM SIGMOD Newsletter FDT*, Vol. 7, No. 2.

Ashplant, T.

1983 The Lexicon of Greek Proper Names. In L. Cignoni and C. Peters (eds.), *Computers in literary and linguistic research: Proc. of the 7th International ALLC Symposium*, Pisa: Giardini.

Bartle, R. and M. Cook

1982 Computer applications in Archive: a survey. *British Library R & D Report 5749*, Archives Unit, University of Liverpool.

Beazley, J.

1956 *Attic Black Figure Vase Painters.*

1963 *Attic Red Figure Vase Painters.*

1971 *Paralipomena.*

Burn, L. and R. Glynn

1983 *Beazley Addenda.* Oxford.

Burnard, L. D.

1980 An application of Codasyl techniques to research in the humanities. In J. Raben and G. Marks (eds.), *Databases in the humanities and social sciences*, Amsterdam: North-Holland.

Chen, P. P.

1976 The entity-relationship model: towards a unified view of data. *ACM TODS*, Vol. 1, No. 1.

Date, C. J.

1981 *An introduction to database systems*, 3rd. ed., Reading, Mass.: Addison Wesley.

FRASER, P. M.
1976 *A new Lexicon of Greek Proper Names in Tribute to an Antiquary: essays presented to Marc Fitch by some of his friends.* London: Leopards Head Press.

FROST, R. A.
1982 Binary relationship storage structures. *Computer Journal*, Vol. 25, No. 3.

GLYNN, R.
1983 The Beazley Archive Project. In F. Lisserrague (ed.), *Image et ceramique grecque*, Marseilles.

GRIETHUYSEN, J. J. VAN (ed.)
1982 *Concepts and terminology for the conceptual schema and the information base.* ISO (TC97/SC5/N695).

GUICHARD, S.
1981 Catalogue informatisé des antiquités Egyptiennes du Musée du Louvre. *Göttinger Miszellen Beitreage zür ägyptologischen Diskussion*, Vol. 51, Göttingen.

ICL TECHNICAL PUBLICATIONS
1983a *IDMS R00153, R00154, R00155, R00156.*
1983b *Using Querymaster R00260.*

ICL COMPUTER USERS ASSOCIATION CAFS SPECIAL INTEREST GROUP
1984 *Exploiting CAFS-ISP.* Working Party Report.

JOHNSON, R. G.
1984 Integrating data and metadata to enhance the user interface. In J. Longstaff (ed.), *Proc. of the Third British National Conference on Databases*, Cambridge University Press.

KENT W.
1979 Limitations of record-oriented information models. *ACM TODS*, Vol. 4, No. 1.

KOWALSKI, R.
1984 Logic as a database language. In J. Longstaff (ed.), *Proc. of the Third British National Conference on Databases,* Cambridge University Press.

MALLER, V. A. J.
1979 The Content Addressing File Store. *ICL Technical Journal*, Vol. 1, No. 3.

Marburger Index
1985 *Marburger Index.* München, Saur.

MCCLEERY, R. H. and C. PERRINS
1985 Territory size reproductive success and population dynamics in the Great Tit (Parus Major). In R. M. Sibley and R. Smith (eds.), *Behavioral ecology: Ecological consequences of adaptive behavior*, Oxford: Blackwell.

NEUFELD, S. D.
1981 The MDA systems and services: a user's view. *MDA Occasional Paper*, No. 6, Duxford.

PAPE, W. and G. E. BENSELER
1911 *Wörterbuch der griechischen eigen namen.* Braunschweig.

PORTER, M. F., R. B. LIGHT and D. A. ROBERTS
1977 A unified approach to the computerization of museum catalogues. *British Library R & D Report*, No. 5338HC, London.

PORTER, M. F.
1982 GOS: a package for making catalogues. *Inf. Technol. R & D*, Vol. 1, pp. 113–129.

PRICE, D.

1984 Computer based storage and retrieval of palaeontological data at the Sedgwick Museum... *Palaeontology*, Vol. 27, Part 2, pp. 393–405.

PROWSE, S.

1983 *Software for producing library catalogues.* Oxford: Elsevier.

RAHTZ, S. P. Q.

1984 *LGPN Report on Database Conversion.* (unpublished report, Oxford).

ROBINSON, H.

1981 *Database analysis and design.* London: Chartwell-Bratt.

TSCHIRITZIS, D. C. and A. KLUG (eds.)

1977 *The ANSI-X3-SPARC DBMS Framework: Report of a study group on database management systems*, New York: AFIPS Press.

TSCHIRITZIS, D. C. and F. H. LOCHOVSKY

1982 *Data models.* Prentice-Hall.

UK-TIA STUDY GROUP

1979 Design and protocol of the UK TIA aspirin study. In G. Tognoni and S. Garallini (eds.), *Drug treatment and prevention in cerebro-vascular disorders*, Amsterdam: Elsevier.