THE BEAZLEY ARCHIVE: MAKING A HUMANITIES DATABASE ACCESSIBLE TO THE WORLD

JONATHAN MOFFETT
ASHMOLEAN MUSEUM, OXFORD

Archives, galleries, libraries and museums are the principal storehouses of the raw materials that fuel humanities research. The scholarship embodied in this research depends on the artefacts, books, documents, manuscripts, pictures, sculptures, etc. which are kept in such institutions. It is thus incumbent on these resource centres to make available the information they hold. This paper describes one such attempt with an archive held by the University of Oxford.

Take an early twentieth-century scholar who wanted to research some aspect of the humanities, for instance, ancient Greek civilization as reflected in the decoration of a particular type of manufactured vessel. In order to conduct such research the scholar had to travel the globe, visit many museums and galleries, locate and examine the evidence relevant to such studies. After a considerable amount of travel, note taking and thought, a corpus would be published describing the vessels and the interpretations which resulted. This publication, possibly in more than one volume, would be substantial and finely illustrated, thus providing an invaluable resource for other scholars researching in the same and related fields. It would not only be found on the shelves of libraries, but also on those of other scholars.

By the end of the twentieth century the world is a different place. The isolated humanities scholar spending many years, perhaps decades, carrying out his or her research, producing occasional publications prior to and sometimes after the main body of work has seen the light of day, is no longer encouraged. The need to justify research, particularly by publication, and to complete it within a short time-frame, has become a factor of growing importance. The day of continuing, long-term funding is fading. Philanthropy is no longer justification-free, some return is expected and often quickly.

However, it is not just the sources of finance that have changed. There are now many more people conducting humanities research, so that the funding needs of all these researchers have increased considerably. With more researchers researching, there has been a consequent explosion in the amount of information available. More information is made available in more books, many of which are only found in libraries. A further consequence is increased specialization with

growing pressure to produce some unique insight into the world, past or present.

Although the ease with which a scholar can move around the world has improved, with faster travel and greater accessibility to more geographically disparate locations, it only takes a crisis like the Gulf War to curtail the possibility or even desirability of undertaking such ventures. Finally, the days of the quarto, colour illustrated leather bound book are long gone. The expense of publishing often precludes even the inclusion of a few colour plates in a small monograph.

Information Technology (IT) offers a number of options to overcome these problems and crises. Although IT is definitely not a universal panacea, it is certainly more than a palliative, and this paper describes one such option in the discipline of classical archaeology, and discusses the reasons why this particular solution is most appropriate and the lessons and implications which can be applied to the humanities in general.

THE BEAZLEY ARCHIVE DATABASE

The Beazley Archive had its beginnings in the work of Sir John Beazley, who spent a lifetime researching figure-decorated Greek vases made in the Athens region between the sixth and fourth centuries B.C. His many photographs, notes, drawings and books were bought by Oxford University in 1966 and were transferred to the Ashmolean Museum in 1970, where they were gradually transformed and augmented into what is now a vast collection of research material which is made available to many scholars. It is not just the classical archaeologist who uses this archive – historians (of the art, political and social varieties), dramatists, linguists and those who study ancient economies also use these objects as sources of information. The decorations not only illustrate mythological stories but also everyday life – ancient Athenians at work, rest and play. Their importance lies in the fact that much of western culture has its roots in this civilization, and they are a major source of visual representations of the ancient Greek world.

The Archive is by nature a pictorial data set in that it comprises some 250,000 photographs and many thousands of drawings of Greek vases, in profile, in detail, and from many different angles. Only a small number of these vases are in the Ashmolean Museum, so the Archive obtains photographs from other museums and collections that hold these objects. The computer was introduced in 1979 to handle the vase records using FAMULUS on an ICL mainframe operated by Oxford University Computing Service (OUCS). This was subsequently replaced by the network database management system (DBMS) IDMS, which was then superseded by the relational DBMS Ingres, this time running on the OUCS VAX cluster. Due to the limitations of

available computer systems, the various databases have been text based, although the information is stored in such a way as to overcome the lack of visual images by using Beazley's well developed descriptive scheme. The data structure of the current relational database is shown in Figure 1, and the contents of the computer-based archive are best described by explaining this structure.

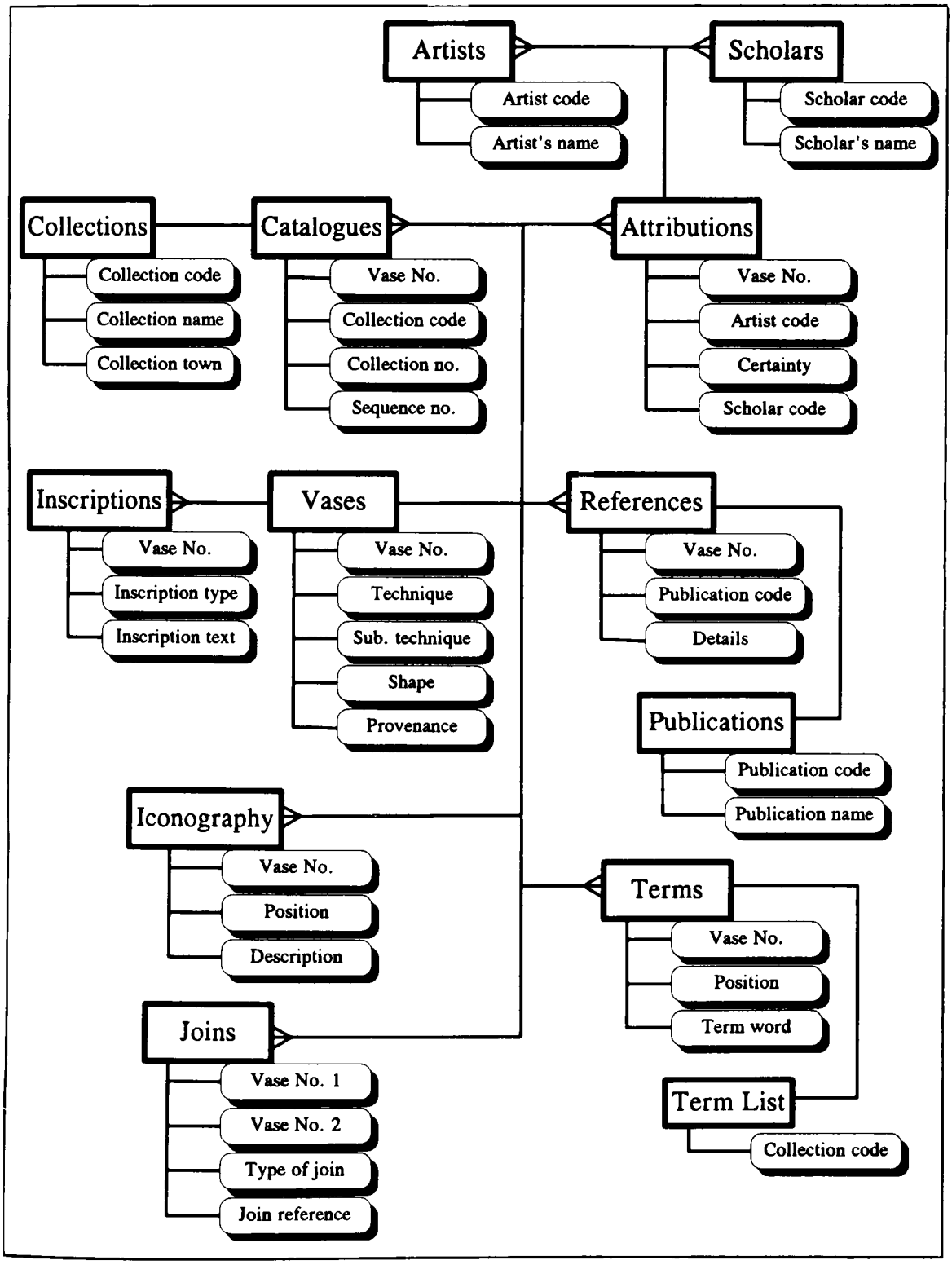

FIGURE 1: THE BEAZLEY ARCHIVE DATABASE STRUCTURE

The *Vases* table holds information relating directly to a vase, and all vases have a single entry in this table, each vase having a unique identification number which is assigned prior to entry into the computer system. Other information relates to the overall technique, that is either black or red figure, and details about the sub-technique, such as semi-outline, silhouette or applied colour, and shape which is stored as a keyword, sometimes with subdivisions, e.g. CUP, LEK (*lekythos*) or LEK S (squat *lekythos*). Finally, the provenance is recorded.

The *Attributions* table gives details about who the painter of the decoration may be, or the date range within which the vase is thought to have been made, along with the degree of certainty of the attribution and the scholar who made the attribution. As will be obvious, since several scholars can make different attributions about the same vase, each attribution is stored as a separate row in the table, although each row has the same vase number. The artist and scholar fields are coded and the two tables, *Artists* and *Scholars* respectively, contain the definitive lists of artists and date ranges that can be ascribed to a vase and of the scholars who have made such ascriptions.

The *Catalogues* table is basically a cataloguing history of where the vase has been since it was discovered. The collection code and collection inventory number are stored, with the full collection name and its location being stored in the *Collections* table. A serial number is used to give a chronological order to the entries for a particular vase.

The *Iconography* table contains the description of a vase's figure decoration, and the position of that decoration, such as obverse, reverse and inside. The *Inscriptions* table stores any signatures that may be on the vase, as well as any *kalos* names – both of which are transliterated from the Greek, as none of the database systems used have been able to cope with Greek character sets.

The *References* table stores all the published illustrations of a particular vase with the publication code and the page and/or plate number of the illustration. The definitive list of publications is stored in the *Publications* table. Again, it is obvious that there can be more than one published reference to a vase and so as many entries as are required are stored in the table, each with the same vase number.

The *Terms* table is based on the iconography of the vase and is basically a selected number of descriptive words which are used when searching for some form of decoration. Again, the position of the description is also stored such that a vase with the figure of Herakles on side A and a Hydra on side B would be distinguishable from vases depicting both figures in the same scene. The allowed list of keywords is found in the table *TermList*, although not all the words that are given in the Terms table will be found in the Iconography table.

As of September 1992, the Beazley Archive database covered approximately 80 Megabytes of VAX filestore. The number of records held in each of the tables is given in Table 1.

Table	Records
Artists	1,442
Attributions	70,924
Catalogues	50,494
Collections	1,382
Iconography	60,905
Inscriptions	1,335
Joins	19
Publications	1,140
References	58,097
Scholars	595
Terms	239,463
Term-List	1,514
Vases	42,840

TABLE 1: BEAZLEY ARCHIVE DATABASE STATISTICS

TECHNIQUE	SHAPE	FIND PLACE
R	CUP B	
TOWN ① (TEXH) Texas ② (NEWS) New York	MUSEUM ① Hunt ② Sotheby's	INV. No. ① 5 ② XXXX 7043

SUBJECTS
A: HYPNOS AND THANATOS WITH SARPEDON, AKAMAS
B: PYRRHIC, YOUTH PLAYING PIPES, YOUTH AND WOMAN
I: FLORAL PATTERN, PALMETTES

KALOS	SIGNATURE
	EUPHRONIOS EGRAPH [SEN]

ATTRIBUTION
525-475 (EUPHRO) [SIGNAT]

PUBLICATION (GetMus), 9 (1981) 24-26, FIGS. 1-6 (I, A, B)
(Hunt Coll), 54-57, NO. 5, FRONTISPIECE (I, A, B, PARTS)
(Soth Hunt), NO. 6 (I, A, B, PARTS)

7043

FIGURE 2: BEAZLEY ARCHIVE VASE CARD

THE DATABASE MANAGEMENT SYSTEM

Data compilation is carried out by trained classical archaeologists who initially fill in a record card for every published illustration of the vase they come across. The details written on the cards are then fed into the database. Figure 2 is an example of a card. The entire computer-based archive revolves around the vase number which, if stored incorrectly in any of the tables, will make a nonsense of the data. However, the data-entry system, which is written in the Ingres Applications-By-Form (ABF) fourth generation language (4GL), is so designed that the vase number is initially entered by the user who is then given the option to add, delete or update the record. There are five basic entry screens, relating to the attributions, catalogue details, iconography, inscriptions and published references. Once a vase has been dealt with, the user has the option to move on to another vase. For global editing, the Ingres interactive Standard Query Language program (ISQL) is used and the SQL command *update* issued. Understandably, only Archive staff are given the necessary knowledge and access to carry out such operations.

For Archive staff, the basic retrieval system works with a combination of the ISQL program and a specially written Pascal program which produces a vase report. In answer to a query, the Archive staff enter the relevant SQL commands so that a list of vase numbers is obtained. These are stored in a temporary table in the database for use by the report program. For example, if someone wants a list of vases that illustrate the Greek mythological characters Thanatos and Hypnos (Death and Sleep) the command would be:

```
INSERT INTO Temptable(Vaseno)
  SELECT Vaseno FROM terms
    WHERE Termword = 'THANATOS'
      and Vaseno in (
        SELECT Vaseno FROM terms
          WHERE Termword = 'HYPNOS')
```

It should be pointed out that this actually locates those vases where the two figures are found, it does not necessarily mean that they are on the same part of the vase. To distinguish between the parts of the vase the SQL command would be:

```
INSERT INTO Temptable(Vaseno)
  SELECT Vaseno FROM terms
    WHERE Termword = 'THANATOS'
      and Vaseno in (
        SELECT Vaseno FROM terms
          WHERE Termword = 'HYPNOS')
        GROUP BY Vaseno, Position
        HAVING COUNT(*) > 1
```

The resultant table *Temptable* is used by the report program to produce Figure 3.

```
BEAZLEY ARCHIVE DATABASE
REPORT ON VASE   7043
------------------------------------------------------------
Technique : R                        Shape  : CUP B


  Signature : EUPHRONIOS EGRAPH[SEN] (EG)
  Provenance :

  Attributed to EUPHRONIOS Painter  by SIGNATURE
  Attributed to 525-475 by x
Decorated Area : A  HYPNOS AND THANATOS WITH SARPEDON, AKAMAS
Decorated Area : B  PYRRHIC, YOUTH PLAYING PIPES, YOUTH AND
WOMAN
Decorated Area : I  FLORAL COMPLEX, PALMETTES

Cataloguing history
-------------------
 1    TEXH5           Texas, Hunt Collection
 1    MALGLOAN7043    Malibu (Ca.), The J. Paul Getty Museum
 2    NEWSXXXX7043    New York (N.Y.), Market - Sotheby's

Publication record
------------------
 J. Paul Getty Museum Journal, 9 (1981) 24-26, FIGS.1-6
(I, A, B)
 Wealth of the Ancient World, the Hunt Collections (Fort
Worth, 1983), 54-57, NO.5, FRONTISPIECE (I, A, B, PARTS)
 Sotheby's, The Nelson Bunker Hunt Collection, New York,
19.6.1990 (New York, 1990), NO.6 (A, B, I, PARTS)
```

FIGURE 3: BEAZLEY ARCHIVE VASE REPORT

MAKING GREEK VASES ACCESSIBLE VIA COMMUNICATIONS NETWORKS

Although the Archive has an international reputation and is heavily consulted by scholars from all parts of the world, until recently, unless scholars actually visited the Archive, queries had to be conveyed by mail or possibly telephone. Since the mid 1980s the U.K.'s Joint Academic Network (JANET) has allowed queries to be sent by electronic mail although, naturally, it took some time before scholars began to use this facility. In 1990, the growing potential of national and international telecommunications linking computers across the world made it a realistic possibility for scholars to use the Beazley Archive interactively and remotely at their desks. To this end, the Archive, in cooperation with OUCS, set up a number of accounts on gateways between JANET and the International Packet Switching System (IPSS) as well as on the National Science Foundation's network (NSF-NETRELAY) to allow users outside the U.K. to call

up the Archive. The current list of countries where users have access to the system includes France, Germany, the Netherlands and Switzerland in Europe, and Australia and the United States.

Once users have gained access to the Beazley Archive they are confined to using an interrogation program written in the programming language C. The aim has been to restrict the users to retrieving information, only Beazley staff are allowed to update the database in any way – although if errors are found by users they are asked to notify the Archive, by E-mail, and corrections will be made.

The interrogation program has three basic commands: LIST, FIND, and PRINT. The LIST command allows the user to find out what keywords are stored, thus the command

```
LIST TERM HE
```

will list, in alphabetical order, all the term words that begin with the letters 'HE', and so will find HERAKLES and HERMES, as well as HELMET. It is also possible to list the artist, scholar, publication and collection codes, as this is information which is controlled and standardized in the database.

The FIND command is the most useful as it is this that allows users to locate vases. Returning to the two figures used in the examples above, given the desire to locate all vases that are decorated with the figure of THANATOS a user would enter:

```
FIND TERM THANATOS
```

When the search is complete the user is shown the results as:

```
1. FIND TERM THANATOS [24]
```

where the first number indicates the search sequence number and the second is the total number of records found. The speed of the retrieval is relatively fast, although this depends on how heavily the VAX system is being used: in the above case the search took just less than a minute around about 2 p.m. In computing terms this may be a long time, but not when compared to a manual search.

The user can then enter a second search command, e.g.

```
FIND TERM HYPNOS
```

which may result in the program replying:

```
1. FIND TERM THANATOS [24]
2. FIND TERM HYPNOS [25]
```

It is possible to carry out the two set operations of intersection and

union on the resulting lists of vase numbers, or in logical terms, AND and OR. To find those vases which have both THANATOS and HYPNOS figures the user enters the command:

```
1 AND 2
```

which may yield the result:

```
1. FIND TERM THANATOS [24]
2. FIND TERM HYPNOS [25]
3. 1 AND 2 [19]
```

This latter operation takes very little time (less than a second), as the database is not actually used. The program compares two files of vase numbers, which were created when the original searches were carried out. Thus, even when the number of vases are in the thousands, the user will have to wait no more than a few seconds.

The PRINT command basically displays information about the vases that have been found in the latest search, which would be number 3 in the above sequence. There are several options – the main one being a complete Beazley Archive record of the vase (Fig. 3), although it is possible to list only the artists, collections, or references, as desired.

IMAGES ACROSS THE NETWORK

The concept of remote site access to data sets is certainly not new; on-line bibliographic databases have been fulfilling this role for many years, e.g. Dialog and BLAISE. However, the Beazley Archive is probably the first internationally accessible on-line database in classical archaeology, and also probably the first project to attempt to incorporate images into an internationally accessible database. It is in some respects blazing a trail in classical archaeological computing and possibly to a lesser extent in the humanities as a whole. However, when setting up such a system several issues come to the fore.

Although written information such as the history, composition and provenance of an object is important and necessary, when conducting research in a discipline such as classical archaeology, it is just as important to see the object itself. While one can use a description made by someone else, as the song says, a picture is worth a thousand words, or possibly quite a few more when the original is there. However, failing the actual object, the thousand word alternative is a suitable substitute. Consequently, the next stage in the development of the computerized Beazley Archive is incorporating an 'image-base' alongside the text database, which like the current database, will be accessible over international networks. However, with a potential target of 100,000 images, selected from more than a

quarter of a million, the creation of the complete image-base has a number of hurdles to overcome.

Although the current database is only some 80 Megabytes in size, once image files are created and stored the amount of disc capacity required will increase enormously, even allowing for very efficient compression algorithms. Assuming a size of 40 Kilobytes per image file for a thumb-nail and low-resolution preview, the amount of space required will be some 4000 Megabytes, or 4 Gigabytes. It is very unlikely that the Oxford University Computing Service would be willing to allocate space for such a large data set, while their willingness to support the software to handle the combined text and image base is also an unknown factor. The implication is that the database system should be transferred to an in-Archive system. However, this will have the knock-on effect of making the database inaccessible to international users. The interim solution will be to run both in tandem, with the textual database stored on two machines, one allowing international access and the other building up the image base. The in-Archive system will also be the data entry system, so that the OUCS version will be a second, periodically updated copy. This temporary solution also has obvious security advantages.

It is anticipated that within the next five years Information Technology, specifically in the form of computer telecommunications, will develop sufficiently for the Archive to have a direct link to the various networks, as well as allowing the transmission of images. Indeed, with the use of broadband technology and satellites the potential actually exists today, although not all the necessary elements are yet in place, neither does the Archive have the resources to take advantage of the current situation.

Another hurdle to be overcome is that of software. The current interrogation program is based on typing line commands, with the results also printed line-by-line. This means any communications package can be used to talk to the database. Once images are available, it will be necessary to have suitable software for receiving and displaying them. To some extent this is not a problem for the Archive to solve; the commercial world requires systems that will do this, for instance newspaper organizations, and such software already exists, e.g. NewsPhotoAccess. However, a certain amount of system development is required at both the transmitting and receiving ends, with the Archive supplying or specifying the appropriate communications package for users.

In order to alleviate this problem, the Beazley Archive, in its guise as a part of the Ashmolean Museum, has become involved in a European Community funded project called RAMA (Remote Access to Museum Archives). The intention is to develop a system whereby museums of different types, across Europe, can be linked by telecommunications in order to access each others' collections databases. An initial experiment between the Beazley Archive and the Musée

d'Orsay successfully took place in the second half of 1992, to attempt the transmission of data and images from France to the UK via an ISDN (Integrated Services Digital Network) line. The software is being developed by Télésystèmes (a subsidiary of France Telecom), and the project involves the participation of museums and telecommunication companies and computer consultancies in France, Germany, Greece, the Netherlands, Spain and the U.K. The project is due to finish at the end of 1994, with commercially available software being the end result.

GENERAL HUMANITIES CONSIDERATIONS

Other more general issues arise when considering the development of this system. The first is persuading people that it can be done, which tends to be linked to a question of financial resources. It is surprising how many people foresee hurdles; sometimes when these are investigated they turn out to be molehills, rather than the anticipated mountains. Inevitably, one stumbling block is money, but as already mentioned above, much of the hardware to do the job is being put in place by commercial, academic, computer and telecommunication organizations. An institution like the Beazley Archive can, therefore, concentrate on paying staff, buying the necessary in-house equipment, and paying for general running costs. If access to JANET and other academic networks continues to be free of charge, this will mitigate to some extent the cost of using the database, and users will only need to buy their own hardware and software, much of which they may already be using for other purposes.

Although the Archive liveware (i.e. staff) are currently handling queries and updating the existing database, the actual capture of images is another full-time job. To some extent this is a clerical role, although, since the quality of photographs varies and therefore some need *enhancement*, some technical knowledge is required. Also, to select areas of a vase which can legitimately be considered as suitable for *zooming* demands knowledge of the material itself: thus the function becomes more than just clerical.

The Archive intends to create an image base of some 100,000 pictures. If an image takes five to six minutes to scan, *enhance*, compress and store, then it is feasible that in an eight-hour day some fifty pictures can be transferred, or 250 in a week, or 13,000 in a 52-week year. The whole process will require about eight years, and this is almost certainly an optimistic estimate. Using more than one scanner would reduce the amount of time, but there is still the cost of paying for people to do this.

Another issue is the freedom of access to information, but this is also linked to money. The information held in the Beazley Archive is made available free of charge, except when the cost of paper, postage, and staff time for carrying out complex searches requires some

remuneration, but over the network the only costs are those involved in using telecommunication lines. The Beazley Archive is seen as providing a service for researchers; it is enabling research. The costs of staff and equipment are met from elsewhere, and the intention is not to make a profit from people using the Archive. For systems similar to the Beazley Archive, it would be understandable to levy a charge if staff and equipment costs were not met from other sources, although since the humanities are relatively poor, at least compared with the sciences, to make a charge would very likely discourage people from using it. Also there would be the administrative costs of billing, collecting payments, and currency conversions to be considered. Whether such benevolence is sustainable is open to debate, but in the case of the Beazley Archive there are no plans to change this arrangement at present.

A further issue that works against freedom of access is that of security. A single username gives international access to the Archive database, but it is very restricted in what it can do, basically only running the interrogation program. Additions and corrections to the database can only be carried out in Oxford, and this will ensure that the integrity of the database is kept. Also this username is not allowed to use the VAX operating system or any break keys. There is also a second security level of usernames and passwords specific to the database to protect it as much as possible. No doubt a determined hacker would be able to get through these protective measures, but to what end is debatable. Hackers, like mountain climbers, may do what they do 'because it's there', but frequently made copies of data and programs should provide some insurance against senseless acts, while constantly updated virus scanning software can locate other, possibly malicious, software. The system only has a single username and password, mainly because it is administered by OUCS. An in-Archive system would be more flexible: users could be given individual usernames, and a form of *big brother* monitoring would become possible.

Another issue is that of copyright, intellectual property, and what the information can be used for. Although the Beazley Archive has a large collection of photographs, it seldom has the negatives and rarely holds the copyright on these images. Currently there is no problem as the images are only available on-site, but once an image base has been established and is remotely accessible, this issue will need to be addressed. To some extent the problem is akin to photocopying, although the ease of copying files from one computer to another is far less time-consuming than copying an entire book, while computer files do not degenerate in quality.

It is only if the images are to be published that problems really occur. The proposed solution of the moment is for images to be stored in a low-resolution format for remote use. The quality will be high enough to see the detail on the screen, but insufficient for a satisfactory

screen dump: after all, what academic publisher would want to publish a poor-quality screen dump of a photograph?

Undoubtedly the copyright issue is a major world-wide problem which will be explored by the commercial and academic communities in any case, and so it should be possible, until a resolution of the issue is accomplished, for the Archive to continue with its planned solution. Although sound and moving images are not likely to form part of the Archive's data set, in the wider humanities context they too will raise the same problems of protecting copyright. (It is not possible to downgrade the quality of sound, as this would negate the reason for making it available. It is to be hoped that the recording industry will provide a solution, although since they failed to do so with copying ordinary cassettes and tapes, perhaps this is wishful thinking.)

The fact that the Archive rarely holds the copyrights of its photographic collection is the reason why it cannot use the Compact Disc publication route. For similar archives and collections where the ownership is multiple or uncertain, the use of an on-line system will overcome the legal difficulties, at least until governments and other bodies have resolved the issue satisfactorily. For institutions which hold copyright, publication may be the most appropriate solution.

A final long-term issue is upgrade and replacement of equipment over time. The Archive began its computerization program in 1979 and has gone through several changes of hardware and software. As long as the Archive exists and is used, it does not require a prophet to foresee that there will be changes as well as improvements in the technology of the future. Similarly, just as photographs fade, so computer-readable media go through an equivalent decay process. Transferring data and text is a relatively simple, if time-consuming, task. Eventually the same will be true of image files, but someone will have to carry out the process. This naturally brings us back to the financial question. Running a computer-based archive is like running a library or museum: it requires a long-term commitment on the part of those who provide the funds. This is not the place to discuss the issue of whether the humanities should be self-financing, but it is doubtful whether the Beazley Archive ever could be.

The Beazley Archive is a large, concentrated set of data from a small, specialized subject area, but a considerable number of scholars from different humanities disciplines actually use the information it contains. More significantly, these scholars are based all around the world, which makes it an important resource for the humanities. The purpose of Information Technology is to make such sets of data available as quickly as possible, in as much variety as possible. The Beazley Archive is attempting to marry the potential of IT with the needs and desires of classical scholarship, and so far has been very successful.

ABOUT THE AUTHOR

JONATHAN MOFFETT is Data Manager both for the Ashmolean Museum and for a number of British Academy funded projects based in Oxford, principally the Beazley Archive, the Lexicon of Greek Personal Names and the Supplement to LSJ. He is co-editor of *Computing for archaeologists* and of the Proceedings of the 1991 Computer Applications in Archaeology Conference, *Computer applications and quantitative methods in archaeology 1991*, ed. G. Lock and J. Moffett, International Series 577 (Oxford: British Archaeological Reports, 1992).

While accepting full responsibility for remaining inaccuracies, Dr Moffett wishes to thank Dr Donna Kurtz, Dr Thomas Mannach and Dr Marilyn Deegan for reading versions of this paper and making corrections and suggestions.

CONTACT ADDRESSES

Post: The Data Manager, Ashmolean Museum, Oxford, OX1 2PH
Phone: (0)865 278030
E-mail: MOFFETT @ UK.AC.OX.VAX
Fax: (0)865 278018