Possible Directions in Electronic Publishing in Archaeology

Sebastian P. Q. Rahtz
Department of Computer Studies
University of Southampton, Highfield, Southampton S09 5NH

1. Introduction

Since their development in the late 1940s, the main work in the application of computers has been persuading the recalcitrant machines to do clever things, huge, enormous jobs that humans cannot manage, such as calculating mathematical formulae to a trillion points of accuracy, directing spacecraft to Mars or looking after the sale of sandwiches in Marks and Spencer. The very word, *computer*, is associated with calculation, with science and mathematics, things that men in white coats do—we would do better to remember Marshall McLuhan's phrase 'information processing' (which he came up with when IBM asked him what they did) or the French word, *ordinateur*, or 'arranger', altogether a more appropriate description of the machine. Archaeologists have largely followed the trend in their use of computers—for most people, there is still a feeling that one has to do something clever to justify the use of a computer. The last decade, however, has seen the growth of a feeling that computers can be used for really pretty dull things, things that humans can do, but cannot be bothered with; the theme of my paper is this use of computers, to do *useful* things, not *clever* things. UNIX hackers will recall that Thomson and Ritchie wrote the system a) for playing games and b) as a text-processing environment [1]. At a gathering like **Computer Applications in Archaeology**, I am probably preaching to the converted, but then again a glance at the contents of this conference reveals that many archaeological computer hackers are still trying to do clever things, rather than using a computer to give them extra time to do the clever things themselves.

That is the moral of this paper (all the best papers give their conclusion before they have even started) and its contents are not designed to be either new or important for the world of historical interpretation. What I shall be dealing with is the ways in which a computer should be used in the publication process; this does not mean I am going to talk about word-processing again—anyone who isn't bored with that subject can read my thoughts on the matter in the Office for Humanities Communication *Newsletter* [2] and the forthcoming issue of the *Archaeological Computing Newsletter* [3] . Instead I want to discuss the following topics:

---

[1] D. M. Ritchie, 'The Evolution of the *UNIX* Time-sharing System', AT&T Bell Labs Technical Journal vol. 63 no. 8 October 1984, 1577–1593.

[2] December 1985, issue no. 5, from Office for Humanities Communication, University of Leicester, Leicester.

[3] no. 6, Spring 1986, from Department of Computing, North Staffs Polytechnic, Blackheath Lane, Stafford ST18 0AD.

1. The more sophisticated uses of word-processing and text formatting programs, and computerised typesetting.
2. The use of electronic bulletin boards and networks.
3. The publication of databases in electronic form.
4. Whether we have to make fundamental changes in the way we "publish" archaeological data in an electronic medium.

The first two of these of these are, of course, equally relevant for anyone in the business of academic publishing.[3] I would argue, however, that archaeology is a special case in the extreme suitability of its information for the electronic media. Archaeologists, perhaps rightly, tend to be allied to the 'Arts' side of things, but the publication of their basic data is much more akin to that of their scientific colleagues; this has a number of characteristics:

a. The data is needed by other workers in the field as soon as possible.
b. The point of much publication is to disseminate information, not opinions. Traditional archaeological publication stresses the interpretation backed by data—should we not be offering the unadulterated data separately from the thoughts about it?
c. Above all, different parts of the audience want different things. However we arrange a printed publication, it remains strictly linear and fossilized, and some people will find what they want difficult to trace. Here, the random nature of electronic media is of immense advantage, as different views of the same data can be presented to different readers.

[3] Those who write for the 'real world' enter a different world, that of Eddie Shah and colour printing in Taiwan.

2. Improving the Printed Page

The use of a word-processor and a daisy-wheel printer for interim reports, local publicity and printers copy is well-known, although many people do not use the full capabilities of modern software, ranging from spelling checkers, through pre-defined formats up to the sophisticated facilities of the text-processing systems under UNIX which maintain a standard bibliography of commonly used references for insertion (in a number of styles such as footnotes or endnotes) in a text when called up by keywords. But for all the hype, word-processing is still glorified typing, and cannot mimic the facilities of the printer; to get close to the quality to which we are accustomed, the archaeological publisher (perhaps we should call him/her the AP) has to learn about book design, page layout, ligatures, tables, a *vastly* increased character set, hyphenation routines, points and picas. There are two schools of thought about all this: the first says that book-design is unimportant, and the information can survive any presentation. The misguided souls who run British Archaeological Reports doubtless subscribe to this philosophy. The second school says that, yes, design and layout matter, "but look how wonderful my Apple Macintosh is", and "have you seen the latest laser printer". Both, I would argue, are wrong. The first, because they cannot see how the design of books has evolved over the last 5 centuries to *enhance* the content, not just prettify it; these pioneering printers didn't invent concepts like page numbers and running heads for a joke, but to help their readers. Trials demonstrate that there is a *wrong way* to lay out, say, a dictionary, or an exhibition catalogue—bad design impedes and impairs understanding. And also because they are Philistines. The second school does not appreciate just how many variables are being juggled by a book designer and his typesetting machine; the fact that the publisher may use the same IBM micro or Apricot to input text and manipulate it, does not mean that he is using the hyphenation routines of Wordstar.

There are good reasons for the substantial differences between stand-alone word-processing systems and their printers, and machines linked to typesetters. It is essentially to do with a quarrel between two intelligences; on our word-processor, the computer controls the whole operation, and the printer does as it is told; if we send text to a typesetter, it has its own computer, which will proceed to overthrow the decisions of the first. If we give the first computer all the knowledge of the second, it has great difficulty (not insurmountable, however) in mimicking the output on its screen. Everybody who uses a word-processor knows the term 'wysiwyg'—'what you see is what you get'. Not quite so well known [4] is the term which describes the systems that one actually has to use when dealing with a typesetter, WYGINAWYW, or *'What You Get Is Not Always What You Wanted'*. Except on the most sophisticated machines, preparing typeset output is a 2 stage process—first, you prepare you text with embedded 'markup', and then you process it. This process is actually much more suitable for the designer; if things are expressed as markup, s/he need only change a few global definitions to alter the look of the whole thing.

So far we have concentrated on the *difficult* points about producing 'real' books with

[4] In fact, I only invented it on 12th March © S. Rahtz 1986; subsequent users should seek my permission.

an ordinary computer; what routes can we, in fact, adopt? The archaeologist who has a report to get out has three options:

- Prepare the text with best available word-processor, and send a the final draft to a conventional printer for complete rekeying and redesign.
- Prepare the text on a word-processor, but insert a greater or lesser degree of generalised markup, putting *visible* codes for things like paragraphs, headings and text emphasis—some publishers can advise on ASPIC or SGML, others have private codes (eg OUP [5]). The publisher then takes your raw text onto his typesetting computer and inserts (manually or automatically) appropriate commands for reformatting. [6]
- You can attempt to typeset the work yourself; this is appropriate under two circumstances: *either* the work is very complex, and you will spend more time explaining the layout to the printer than it would take you to do it yourself (perhaps it contains obscure scripts) *or* you work is highly formalised, and you have easy access to the hardware. The latter case should be more common in archaeology than it is; a large amount of what is in archaeological publication is held in databases, and people spend a lot of time programming their databases to produce nice results. Why not write a program to produce output for the typesetter? A database is an extreme form of generalised markup language—the structure of the tables and variables. A suitable algorithm imposed on this structure can produce highly complex and typographically satisfying output. Fig. 1 was produced in this way, from an IDMS database:[7]

Obviously, the first route is not very satisfactory—none of use wants to spend a lot of time proof-reading (although it should be borne in mind that, just because you write something on a word-processor and run a spelling checker on it, your work doesn't need skilled proof-reading. The writing profession seem agreed that the author is the worst person to proof-read). If you do decide to try your own typesetting, there are software choices; one way is to adopt the conventional printers technique of having a design in mind and inserting the correct codes in the right places to produce it. Thus each time you start a paragraph, you will put in a code for (say) 1 em of indentation—rather tedious if you later decide paragraph indents should be 2 ems; this is the method you will encounter if you use the facilities of the Monotype Lasercomp at Oxford. The alternative is to use software which thinks in terms of macros eg you define paragraph indentation once at the beginning and just use a code in the text thereafter—the definition of what a paragraph is can be

---

[5] *Oxford Rules for the Preparation of Text on Microcomputers* OUP 1984.

[6] see *Word Processing and Publishing Some Guidelines for Authors* by Peter Denley, British Academy 1986, for more detailed advice on this route.

[7] for further discussion of this example, see Elaine Mathews and Sebastian Rahtz, *Typesetting from a Database,* in Proceedings of the XIIth Conference of the Association for Literary and Linguistic Computing, Nice 1985), forthcoming 1986.

Figure 1 Lexicon of Greek Personal Names

changed at any time. Users of UNIX systems will recognise this as characteristic of the well-established *nroff* and *troff* packages; it is also the basis of the system used to set the printed version of this article, TEX. TEXwas designed by a computer scientist and mathematician, Donald Knuth, but that does not blind him to art; TEX was 'intended for the creation of beautiful books ...GO FORTH now and create *masterpieces of the publishing art!*'. [8] The importance of TEX is two-fold: a) it is available for microcomputers (the printed version of this was prepared on an IBM PC AT, and printed on a £200 Kaga dot-matrix printer), and b) it produces device-independent output—the files can be printed on any machine that understands Knuth's .dvi files. This standardisation of files seems to be an important trend, and it is to be hoped that the machine-independent *Postscript* typesetting code used on the Apple laser printer and other machine will become standard (TEX files can be converted to it).

Doing ones own design and typesetting is a very rewarding occupation (*pace* WYGINAWYW), but three points should be borne in mind: a) do you have nothing better to do with your time that worry about whether there should be a kern of .125 of an em between the E and the T? Isn't that why you bought a dishwasher? b) The cost of producing camera-ready pages is not great a part of the publishing expense. I recently typeset a book for a commercial publisher who told me that we had saved maybe £200 out of a total of thousands. There was some slight improvement in the speed of production. c) There are some very ugly books being produced by home typesetters[5]; do you want to

[8] D. Knuth, *The TEX Book* Addison-Wesley 1984, pp. v and 303.

[5] at a conference on typesetting in Oxford in 1985, two separate speakers from Reading University Department of Typography used portions of a book I typeset—*Middleton Stoney A North Oxfordshire Parish* by S. Rahtz and T. Rowley, Oxford 1984—as an example of

join their numbers? One of noticeable results of the IT revolution is a swarm of horrible publications. Perhaps the best comment I can make is to note that I probably spent longer on typesetting this paper than in composing it, because I wanted very much to produce both a readable and a beautiful product.

3. Networks and Bulletin Boards

If I have been negative about the assistance of computers in typesetting, I intend to be much more positive about electronic networks. As I noted in the introduction above, archaeological data is (or should be) often ephemeral, and of importance chiefly to fellow-workers. Why not get it to them quickly and bypass the traditional publication channels? Admittedly, typesetting ones own small publications and distributing them can be very fast and effective, but it is not cheap, and it requires manpower, which archaeology emphatically does not have. The two overwhelming virtues of the electronic distribution of material are as follows: firstly, the distribution is instantaneous. As soon as information is ready, others can access it; secondly, the publication is only virtual—the 'reader' does not get information he does not want thrust upon him/her, but can collect only what is of interest. I am only talking here, of course, about transitory data and electronic mail queries and answers—a more permanent electronic resource needs more careful thought (see below).

In an ideal world, the field archaeologists of Britain (for it is they who would benefit from this instant access to data) would all be linked together on a network. Until the millenium comes, there are three solutions: a) send floppy discs around the country (not much fun), b) use the publicly available bulletin boards, or set up a special one (as the home computer user groups do—eg the UK IBM User Group), and postboxes (Telecom Gold, Prestel—at a cost), or c) use the academic networks. All British Universities are linked together on a network (Joint Academic NETetwork), which has mailing and file transfer facilities and links to British Telecom's commercial network PSS (and thence across the water with IPSS); some, but not (as yet) many Polytechnics are on it. The use of JANET is free to academic archaeologists; other disciplines make extensive use of network mailing facilities [6] for disseminating information, so let us use it too.

For those who have never heard of it, let me add here another 'plug' for the Humanities Bulletin Board (HUMBUL) which was set up at Leicester University in December 1985 by the Office for Humanities Communication; it is designed to act as a place where information, views and queries can be posted about anything to do with computing and humanities subjects. Access is via JANET or dial-in lines, so anyone who has a modem can use it. Archaeology has not reared its head there yet amidst the details of literary computing conferences, the new OED and the Norwegian Computing Centre for the Humanities, but there is no reason why it shouldn't. HUMBUL is set up very simply using the Vax Help system, and once you get through can be used immediately. [7]

---

poor design and lack of proper knowledge.

[6] I receive an electronic digest of news and views about the Kermit file transfer protocol every week; it costs neither me nor its originator anything.

[7] For more information, contact Patrick Holligan, Primary Communications Research Centre, University of Leicester, Leicester LE1 7RH, HOL@UK.AC.LE.VAX on JANET.

4. A Roman Pottery Graphical Database

My third topic is the more deliberate dissemination of data in a controlled form, as a database. Many existing reports are based on computer databases, so why not let the readers have access to them? There are technical problems about linking everybody into a national database, but there is a simple solution in the short-term, which is distribution of material on floppy discs. As an example, let us consider a project about to start at Southampton University to publish a body of archaeological data as an integrated text and image database. It is based on cooperation between the Departments of Archaeology and Computer Studies, University, and we hope the result will be published by Oxford University Press. The initial project deals with a group of Roman pottery (for reasons given below) but the principles would apply to any body of archaeological data which currently relies on 2-dimensional line drawings as part of classification.

It is an integral part of this project that the result should be a commercial product which could be marketed by OUP in the form of floppy discs for a common machine. In the long-term we have two aims: a) to present all the current knowledge (including form shapes) about a given group of Roman pottery in an integrated computerised form, with an appropriate 'expert' front-end to interact with the archaeological user, and b) to establish a general methodology for the capture, storage and analysis of two dimensional archaeological shapes, using *generally available* computer hardware and software. Much archaeological data consists of descriptions and drawings of 3-dimensional objects, and pending the availability of commercial products to record 3-dimensional shapes, an interim system can store and recall the previously created 2-dimensional representations and integrate them with text. The effect, then, is that of a more efficient book—instead of leafing through either text or drawings and then cross-referencing by hand, a computerised system should allow us to browse through a combination of drawings and text in any desired order. Tools can also be provided for more objective analysis and searching of forms than is sometimes possible (ie limiting search to shapes where height is more than twice the diameter). Roman pottery is a good starting point as it on the one hand well studied and understood, and on the other criticised for lack of formal definitions of forms; for publication, its abundance guarentees a reasonable sale. It is important that archaeology should start to take seriously the possibilities of 'publication by database' rather than paper, and this exercise may solve some of the technical problems, and test the attitude of the market.

It must be stressed that this is *not* an attempt to set up an automated pottery processing system, nor an image-processing research project, but an exercise in 'electronic publication' of established data. Aim a) has several sub-stages which will be firm commitments:

i. A conventional database of the selected pottery group, which could be usefully be published by itself, consisting of purely numeric and textual data and a 'friendly' interface.
ii. The simple capture, storage and display of conventional pottery drawings, linked to the main database.
iii. The development of an 'expert system' for more sophisticated use of the underlying database.

iv. A system for input of the shapes of new forms for comparison against existing library. The ideal here would be the user sketching a shape on a graphics tablet, and that shape being matched (with an appropriate degree of fuzziness) against the database. The problems of automatically capturing appropriate data from a comparative pot held in the user's hand seem to me to mask the more important principle that the interpretation by the pottery worker of the shape when he or she draws it is a valuable exercise not lightly to be discarded.

When these tasks are complete, it may seem appropriate to expand either the graphical work to other parts of the archaeological publication, or to extend the expert system side into Roman culture generally. However, the *publication project* is unlikely to encompass more than i. and ii. of the above list.

We would foresee that immediate discussion should centre on the following subjects (in order):

a. Which group of pottery should be presented? This depends on the "readership", on available expertise and on the time available—we propose that the initial pilot scheme should cover a group on which the basic research has been completed, rather than attempting to integrate ongoing research, and the most obvious subject is samian, so the first database will be a computerised version of the standard samian forms.
b. The aims of the publication database must be established, ie what is the ideal user interface and content? Ideal working environments have to be balanced against achievable goals. It will be an important consideration whether the query interface will be solely textual or include some graphical input (drawing the shape one is interested in).
c. On the technical side, we must decide early on the method of storage of drawings, ie whether we are storing a snapshot or a manipulative shape. At this stage, we anticipate that we will have to deal with existing line-drawings and clearly the initial data capture will be best done with a frame-grabbing CCD camera and appropriate software. Relatively straightforward compression and storage techniques can be applied to the raster image to make it immediately useful, but it is also desirable for longer-term use to convert the data to a vector format, in which outline shapes can be manipulated or compared against other shapes in the library, or against user input.
Initial discussion, and comparison of the problem with a similar database being contructed in the Department of Biology at Southampton, suggests that a 512×512 raster image, appropriately compressed, should be used for the first publication, and the conversion to a manipulatable image should be subjected to a separate research project into pottery shapes.
d. A decision must be made about what hardware and software limits must be placed on the project in order to make effective sales in, say, early 1987.

The basic hardware will be a hard-disk single-user micro, and printer support will be a dot-matrix, though the suggested long-term configuration would include a laser printer.

The *development* hardware has to include the frame-grabbing camera, but of course the user only needs a normal machine. The same might apply to software—we are investigating the possibility of selling a 'runtime'-only version of the database software (under appropriate license), leaving the users no option of editing or addition—data updates would be on floppy discs by subscription from the publisher.

During the next year, the following timetable will be followed:

- Firstly, the conventional database will be established and filled with data. While time-consuming and important, this involves normal archaeological skills and normal resources.
- Secondly, an academic/publishing decision will be reached about the user interface to the data, and this will be broadly implemented after experimentation on a subset of the data (which could probably be made available very soon if all went well). Decisions will include minimum standards (speed, definition, access paths) for the ultimate product.
- Thirdly, the problems of the capture and storage of the drawings will be solved. The difficulty here will be finding hardware and software to do the job that is also within the price range of the public. It *might* be necessary to offer a subsidiary package of software and hardware add-on in conjunction with a manufacturer, but the aim must be to produce something that runs a predetermined 'lowest common denominator' configuration. To descend to concrete details, the initial development will be on a Research Machines Nimbus, using floppy discs (to enforce small size and speed), and the database system will be dBaseII—for good or bad, an MSDOS machine and a slightly old-fashioned database is what most people have. More importantly, the whole system will sit on top of Microsoft Windows; all the graphical work will be done with calls to the Windows software, which should give us the best chance of portability. Getting the digitised 'snapshot' of the relevant drawings (leaving aside problems of efficient storage for the moment), and re-displaying the image on request, is only a hardware problem.

After a few months or so, we would expect to see a useable database of normal data, and a separate imaging system. Then we can start combining them together ... The longer-term project will be developed as part of a proposed M.Sc. in *Archaeological Computing* in the Department of Archaeology.

5. Conclusions

This paper was not designed to be revolutionary; its purpose was to review three important areas where the 'trivial' use of computers can make a difference to the publication problems in archaeology. Clearly, typesetting by computer is here to stay, and whatever one says, more and more people will produce ugly books and papers on their Macintoshes and IBMs, and Gütenberg will turn in his grave again; the very best we can hope for is standardisation. once again, the UNIX exmaple is instructive—if one uses all the available tools, one can turn out a very fair imitation of a Bell Labs techinical paper, with consistent formatting, references and even writing style (the right proportion of complex and simple sentences, a Kincaid reading age of 12 years and no sexist terms). If we reached even that stage, the world would be a slightly better, if dull, place.[8]

Electronic mail and bulletin boards are also so self-evidently useful that they will be used by anyone with access to them. The way electronic publishing goes after that in archaeology is not yet clear; the concept of a conventional publisher selling floppy discs (as outlined above for the Roman pottery project) seems at first rather old-fashioned, but has many advantages. The venture by the BBC to sell video-discs of the Domesday will be similar, and obviously video-discs will be of great benefit in archaeology; in a few years, OUP will market the New Oxford English Dictionary in electronic form, and are making experiments now, eg with an electronic medical journal. Whatever the hardware and software, the basic method is crucial—does one distribute individual copies of a database, or allow access to a central version? The trend seems to be towards personal copies, and although this has enormous problems, of machine compatibility and of updating, it is probably still preferable to central databases, unless they are very large indeed.

Whether we buy a copy of the NMR on a floppy disc or access a Vax, there are some important 'philosophical' problems that will arise. Firstly, there is the question of copyright, already a thorny area after the disastrous Fortress House-prompted dive into microfiche, and certain to become a bear garden. If researcher X puts his painfully-gathered field-walking results onto a national computer, and synthesizer Y gathers them in, and presents them at a conference a month later, what recourse does X have? The result will be that only a subset of 'safe' data will be available, which is sad. Alternatively, we must simply stop being so possessive of our information, and accept that our true contribution to archaeology will be measured by our interpretation, which will be richer for shared information.

The second major problem that arises is the 'user interface' to archaeological databases. When all the data from an excavataion is stored electronically, it become necessary to design appropriate search strategies; to some extent, this brings up the old chestnut of standardizing on words and term, but that is a red herring—the real question is how to view an excavation on a computer screen. As a real-time graphical reconstruction? As a conventional database of "list me the contexts with more than 2 Roman coins in the west half of the site" type? As a collection of raw data, or an interpretation? As an

[8] anyone who wants to see the depths to which home typesetting can sink should call in to Blackwell's bookshop in Oxford and search out *Dons Dinner* by Marcelle Quinton, quite the ugliest piece of typesetting in sight, and with the grossest self-indulgent contents.

adventure game ("once upon a time there was just a little village—look, heres a picture of it—but what do you think would happen if you looked underneath the chuarchyard?")? There is also the question of relative importance of data; in a Cunliffe-type report, material that the author considers vital is put in running prose, not so useful material is put in an Appendix in small type, worthy but dull tables are put on microfiche, and the rest is consigned to The Archive. But if the database becomes available, how does the excavator bring out what he considers important?

I do not offer solutions to these problems. But I believe that they are real, and must be faced, if archaeology is to take advantage of electronic publishing, for which it seems so suitable.