The Editorial

Scholarly Online Publishing in Archaeology: the price of progress

by Dirk Brandherm

Contents

Taking Coordinates
Setting the Course
References

Abstract: While the idea of disseminating specialist knowledge over the Internet really caught on with some sectors of the scientific community, others have reacted rather less enthusiastic. The predicted world (wide web) revolution in scholarly communications and the associated cataclysm for scholarly print journals has yet to occur. Instead, issues of data longevity and long term data integrity have become a growing concern for many scholars. For archaeologists in particular, choosing the best medium to publish their data and disseminate the results of their research has not become any easier. The present contribution wants to raise awareness for long-term economical concerns of scholarly e-publication which previous authors have not always adequately taken into consideration.

Keywords: Archaeology, Data Longevity, Hypermedia, Scholarly E-journals.

Taking Coordinates

MPO has been online for almost two years now, which perhaps is sufficient reason for some reflections on its particular development and on the topic of scholarly e-journals in general. As the past few years have made increasingly clear, there can be no doubt any longer that e-publishing over the Internet offers a whole range of advantages for most, if not all, archaeological endeavours. Publishing archaeology on the World Wide Web cannot be labelled a trendy privilege of a small vanguard any longer. The first introductory textbook about the use of the Internet in archaeology has just seen print (ALTEKAMP / TIEDEMANN 1999), while earlier texts on the same subject had only taken the shape of shorter articles (GILL 1995; CHAMPION 1997; REHMET 1997; TIEDEMANN 1998). Meanwhile, some university departments even have started to offer specialised courses on the use of the Internet for archaeological purposes.
The advantages of publishing archaeology on the World Wide Web are obvious, and have been dwelt upon by so many authors on so many occasions that I do not intend here to expand in any great length on the subject, but rather shall restrict myself to briefly rehearsing the most frequently addressed issues.

Most (virtual) ink probably has been spilled on the possibilities offered by the inclusion of hypermedia features in scholarly e-publication (DENNING 1996; 1998; HOLTDORF 1999). In the editorial to the first issue of this journal, too, the potential of online hyperreferencing has extensively been dealt with by COLOMER (1998/99). So there is perhaps no need to ruminate about the subject again at this point, although further below some words will have to be said about the cost at which these blessings come.

Also for the debate about "freedom of speech" vs. "censorship" (aka. peer-review process), little can be added to what was said by COLOMER (1998/99), regarding the policy of MPO. Obviously, the World Wide Web as a democratic medium must provide freedom of expression. As you are able to see from the icon at the centre-bottom on each of its pages, our journal does actively support the Blue Ribbon Campaign for Online Free Speech. However, as VINCE (1996) has rightly stressed, furthering freedom of expression is not synonymous to inviting everyone to put on anybody else's site what they deem fit. Especially on the World Wide Web, peer review cannot mean discrimination of unpopular or deviate views, as today everyone is in a position to communicate his or her message to exactly the same potential audience just by putting up their own site. Also, as HARNAD (1997) has argued, peer commentary cannot act as an adequate substitute for pre-publication peer review, even if the first may grow in importance with the technical possibilities offered by the Internet.

One obvious advantage of e-publishing over the Internet is the potential speed with which texts can be made available to the public. While most printed journals in scholarly archaeology will have to draw a deadline for the submission of contributions anywhere between six and twelve months before the respective issue is to be published, e-journals can incorporate new articles continuously, and even where a peer-review process is involved, the time between the submission of the manuscript and its uploading as a published article will be measured in weeks rather than in months.

Some authors have taken this to herald a general revolution in scholarly communication, possibly leading to the eventual disappearance of papers and journals as we know them today. Such a view has been advocated mainly by HARNAD (1995) and GINSPARG (1997), based on experiences gathered with preprint archives and peer discussion list mainly in physics. However, as VALAUSKAS (1997) correctly pointed out: "Harnad forgets that comparing scholarly communication in the fast-paced world of high-energy physics to
the mere academic deliberations of humanists, social scientists, and non-physics scientists is dangerous”. For reasons explained further below, this would seem especially true for an academic discipline like archaeology, which does not derive the majority of its data from clinical laboratory experiments or abstract computations.

Also, it is important to keep in mind that, as a rule of thumb, the larger a document grows in size, less time is to be gained through e-publication when compared to conventional publishing in print, because the input of work which is needed to convert an electronic manuscript into a document ready to publish does not really differ that much between print and online media, its amount growing more or less proportionally with the size of the document, whereas the time needed to actually publish, i.e. print or upload a document is affected by its size to a much lesser degree.

With large documents, e.g. books, but sometimes also with extensive articles, the advantage in speed of publication in most cases will shrink down to the few weeks that are needed in conventional publishing to ship the copies from the publisher to the distributors. As large documents usually represent the result of many months, often of several years of work by their authors, this kind of difference will usually not be decisive in the choice of the medium for their publication.

Furthermore, when compared to more or less fixed pre-publication costs, the effective costs of publication, i.e. actual printing and shipping of the copies, should not be that decisive in our choice, even if the kind of ratio between pre-publication and actual publication costs that VARIAN (1998) cites probably cannot be applied without adjustment to many scholarly archaeological journals in Europe.

On the other hand, publishing large documents in print at some point usually will meet with some very clear-cut economical limits, as the costs of disseminating, i.e. printing and distributing, lavishly illustrated multi volume site reports are bound to become prohibitive if they cannot be spread over a sufficiently large number of copies. In archaeology, copies sold of scholarly monographs or journals will normally be counted by the hundreds, rather than by the thousands. E-publishing, imposing almost no limitations on text size and allowing for the inclusion of a very large number of full colour illustrations at very little additional cost, could be viewed as a solution here, but there are also other factors to be taken into account when it comes to calculating total costs of publication (see below).

Another obvious advantage of e-publication on the World Wide Web, quite apart from the potential gain in publishing time for documents of limited size, and only partially linked to economical considerations, is their ubiquitous availability. In order to work with the references provided to printed media, one will generally need physical access to some
kind of library or similar "warehouse of knowledge", most of which are forced by the sheer number of printed publications to specialise in particular fields. The role and positive as well as negative side effects of this kind of warehousing have recently been discussed by ALDENDERFER (1999). In order to access any document published on the Web, however, all one needs is a computer terminal or other Internet-capable device connected to it and corresponding software — at least such is the theory.

This has justly been welcomed as a potential step towards the emancipation of colleagues in poorer countries, as gaining access to the World Wide Web is relatively cheap, when compared to the costs of maintaining any of the traditional "warehouses of knowledge", although several authors have also pointed at new problems and possible pitfalls (ALDENDERFER 1999; GINSPARG 1997; HODDER 1999).

In order to lower pre-e-publication costs, ALDENDERFER (1999, 6) also has called for a much more extensive use of digital recording techniques in archaeological field work. This, of course, is diametrically opposed to the desire of getting colleagues from poorer nations more actively involved in the general advance of research. The kind of digital recording equipment he calls for, in economical terms is way beyond the reach of archaeological institutions in third world countries as in many not so poor nations as well, especially if one considers that today's cutting edge devices are likely to become outdated and perhaps even hopelessly incompatible to more modern standards in just a few years time.

The most important issue ALDENDERFER addresses in his paper, data-longevity, also is intimately linked to economic factors. With the widespread use of the World Wide Web and other e-media for publication, it is becoming ever more obvious that costs of publication cannot be defined any longer only as the costs of making a given information available to the public, but also involve the costs of keeping this same information available over time.

As opposed to the sciences, whose basic information is derived from experiments which by definition have to yield reproducible results, the basic information for our work is extracted from the archaeological record, mostly by destructive means, and therefore is unique in each of its pieces. Once lost, it cannot be retrieved, as no two archaeological sites are exactly the same and any given site can be excavated only once. A truism though it seems, these two simple facts must have far reaching consequences for how we choose to record and disseminate our data.

HOLTDORF (1999, 8) has recently argued that "living" World Wide Web pages provide a better guarantee for continued access to information than e-publications disseminated through physical data carriers, which are far more likely soon to join the impressive ranks
of the Dead Media brigade. While this may be true, publishing data over the Internet in itself does little to solve the basic problem of longevity.

The technical aspects of this complex problem have been widely discussed. The physical life span of data kept on the hard drives running on web servers has to be estimated even lower than that of copies recorded on high quality CD-ROM or DVD, which are estimated to range anywhere between 30 and 100 years. Compared to this, a journal or monograph printed on acid-free, buffered paper will easily survive for 500 years, about the same as the estimated minimum lifespan for microfilm. All these estimates refer to storage under favourable climatic conditions (ZIMMER 1999a). As acid-free, buffered paper remains hardly more expensive than acid varieties, and as it is now used for publication by just about all quality journals in archaeology, ALDENDERFER's (1999, 6) statement that electronic media in this respect compare favourably would seem somewhat misleading.

As CHIPPINDALE (1997) has pointed out in this discussion, books like Caxton's "Cronicle of England", published in 1485, still are there to serve our information needs today. If after 500 years one will be able to say the same about most scholarly texts published on the World Wide Web today, is doubtful at least. Some documents, like the pages of the pioneer archaeological e-journal "Archaeology Online" have already ceased to be available online after less than 1% of this period of time. The real problem here does not even rest in the life span of the physical data carriers. Copying digital data from one data carrier to another is cheap and easy enough while the respective technology is still available. Rather it is the limited lifetime of digital formats which is bound to give us and future generations most of the headache.

To assume that any of the databases in which most of the data from current fieldwork is stored, will still be readily accessible in 25 or 50 years, would be more than naive. And it will be exactly then, when the next generation of archaeologist shall want access to these data to confront them with their own theoretical background. So called "industry standards" have come and gone with intimidating speed over the past two decades, and there is no reason why we should expect this to change in the future. Industry must develop new products in order to survive, and cannot be relied on for providing continued backward compatibility for proprietary data formats over decades or even centuries.

As anyone who ever tried to transfer a document of moderate complexity from the file format of his old word processor into the file format of a more recent program will be able to tell, migrating data between different electronic formats can be a rather tricky affair. With complexity of the document and difference in age of the formats increasing, data integrity in many cases becomes impossible to preserve. This problem will become much more pronounced in the future, as new formats allow the creation of increasingly complex
documents, incorporating richer and more diverse data. Also, in order to preserve the advantages of hypermedia documents, not only will the data themselves have to be frequently migrated, but also the hyperlinks between the ever growing number of documents will have to be kept intact during this repeated process of migration. Partly because of these difficulties, emulation rather than migration has been proposed as a strategy for guaranteeing long term data accessibility (ROTHENBERG 1999). This approach, however, for a number of reasons would seem even more problematic, especially where networked data are concerned (BEARMAN 1999).

Open standards, like HTML or XML, in conjunction with standardised sets of metadata will likely go some way to alleviate the problems associated with data longevity and integrity, but they do not by themselves provide a conclusive solution for all times to come. Open standards, too, are subject to continuous development, thus necessitating data to be periodically migrated from one electronic format to another. Otherwise, the data is likely to become inaccessible within the average life span of the individual scholar. So, while we have every reason to support and encourage the use of open standards in scholarly e-publishing, this does not free us from the obligation of taking further steps to guarantee the longevity of our data and the accessibility of the conclusions drawn from our research.

Providing such a guarantee in the long run clearly goes beyond the possibilities of any individual, and HOLTDORF’s (1999) proposal to count on the commitment of the respective author as a possible solution for this problem remains rather unconvincing. Powerful institutional backup is of vital importance here, as migrating data to guarantee their longevity not only poises technical but also very real economical problems. Migrating complex data is expensive, and further technological advances are not likely to make it much cheaper (ALDENDERFER 1999, 6).

Therefore, with e-publications much more than with their conventional printed counterparts, when calculating the costs of any given publication, not only have the expenses for making this particular piece of information public to be taken into account, but also the long term costs of keeping it publicly available in a usable format. The costs of storing a printed volume and keeping it publicly available in a library for the lifetime of that particular book will range from approximately € 25 to € 40 at present value (VARIAN 1998). Extending the lifespan of books that were not printed on acid-free paper by deacidification and equipping them with a alkaline reserve implies additional costs of about € 16 per kilogram paper (ZIMMER 1999b). This will cater for the next 500 years or so.

With information disseminated over the World Wide Web, costs of course could be cut drastically by maintaining all the information at just one site, instead of distributed over the 500 odd copies of a scholarly book or journal issue. This, of course, would mean exactly
the opposite of what the Internet really has been about from its very beginning. The ARPA network, from which the World Wide Web as we have come to know it did evolve, was intended for keeping information distributed in an explicitly decentralised way, in order to guarantee data survival in case of war or natural catastrophes. To provide the same kind of security for digital data as do the 500 printed copies spread throughout the scholarly world, in theory at least, an equally large number of mirror servers would be needed. At first glance this may seem somewhat exaggerated, but the problem is a very real one. Had e.g. ZERAVICA's (1993) work on Bronze Age metalwork from parts of former Yugoslavia been published on the Web, uploaded say, on a server in Sarajevo – perhaps even with a backup kept in a government building in Belgrade –, likelihood has it that not only a large part of the original items catalogued in this work would have been irretrievably lost during the war, but that also their only coherent documentation would forever have ceased to be available.

Anyway, when it comes to estimating the costs of keeping information available in electronic format for half a millennium, the simple truth is that none of us has the faintest idea what the final bill will amount to. Judging from present experience, it might be considerably more expensive than conserving printed hard copies and reprinting them at the end of their lifetime, but we just do not know. In any case, considering this unknown, it cannot be taken for granted that electronic publication over the Internet in the end per se will result cheaper than conventional publication in print, as GINSPIARG (1997) and HARNAD (1997) have so zealously argued.

If at present we have no idea as to how much the "total cost of ownership" will be for those electronic bits of specialist knowledge over the years, it is also quite uncertain to what a degree the traditional concept of "ownership" will be applicable here. Some commercial e-publishers already have adopted a "pay-per-view" policy that clearly aims at eliminating institutional intermediators like university libraries and other "warehouses of knowledge". What happens when such a publishing company ceases to exist, or when migrating old documents simply does not pay for them any longer?

These concerns, obviously, do nothing to diminish the evident benefits of online publishing, but in order for scholarly e-publishing to come of age, these problems have to be resolved first.
Setting the Course

Now, what does all of the above boil down to for readers and publishers of archaeological e-journals now, not to forget potential authors? In more concrete terms: what does this mean for the present issue of Mediterranean Prehistory Online? From a comparison of the strength and weaknesses discussed above, it should be obvious that e-publication on the World Wide Web currently does not serve all needs of archaeology alike. It lends itself better for some purposes than for others, and for some purposes, publication in print still would seem much more adequate even (or especially) in the long term.

Although the dilemma is not easily resolved, it is rather straightforward to explain. On one hand, e-publishing on the World Wide Web not only needs to achieve a critical mass to make the advantages of hypermedia formats fully exploitable, but also to justify the important investments needed to guarantee the kind of long term data survival that archaeologists are looking for. Also, continued experimental work on how to put hypermedia techniques at the service of the advancement and dissemination of archaeological knowledge is important. On the other hand, where data longevity is concerned, keeping documents as simple as possible, and making minimal use of experimental techniques would seem the safest bet. No responsible archaeologist can be expected to entrust the results of many years of meticulous fieldwork, subsequent analysis and final synthesis to data formats which future generations might find difficult or even impossible to access. With documents of distinct preliminary character, like annual excavation or survey reports which eventually are to be superseded by a final publication of the project, the latter problem would seem less pronounced. Here, online publication can free valuable resources for further study of the evidence and preparation of the final site report.

The original concept of MPO did not provide for dividing its growing contents diachronically into individual issues. While perfectly feasible in theory, such an approach does present a number of very practical drawbacks. A single table of contents covering all the contributions at some point will cease to be of help to the reader, and is likely to become more of a nuisance than provide the quick overview one is looking for. While automated searches will easily turn up the desired document, they can be performed in any number of virtual "issues" with about the same speed and yielding just the same result as if searching just a single file. Dividing the contents of MPO into individual issues not only should provide some kind of mental handrail for the reader, but also will ease gradual migration of their contents to newer formats. For these reasons we have chosen to abandon the original concept at this point and to start a second issue.
As the two-year period covered by what thus has become the first issue of MPO (1998/99) clearly is not the kind of issue output frequency that most readers would expect from an e-journal, from now on new issues of MPO will be published annually. Furthermore, for those of our readers for whom offline availability of the original HTML documents is more important than online features like external hyperlinks, as from now, with the close of each issue the contents of the journal will also be made available on CD-ROM. Still, new articles will be uploaded continuously, as they are submitted by the authors and approved by the scientific committee. Thus, authors can continue to expect to see their articles published within a few weeks rather than months, as tends to be the case with conventional journals.

For the sake of simplicity and ease of eventual data migration in the future, we have made a number of further modifications, some of which will be rather evident to the reader at first glance, while others may be less obvious. With page layout for all of the journal still being done by means of conventional HTML, text formatting for the second issue of MPO is resolved entirely by CSS style sheets. If in comparison with the first issue you did not notice any difference as far as fonts and other text-formatting related features are concerned, that is the way it should be. If on your screen all the text from the new issue only shows up in your default browser font and such features as tables of contents and lists of references do not look the way you might expect, then you are probably not yet using a fully style-sheet capable browser.

We also decided to abandon the frame-based approach which was used for navigation with the first issue and to adopt a less intrusive means of navigation. Please read our Notes for Readers for more detail or click on the logo in the upper left hand corner of this page to try the new menu. For the sake of better readability, we have dropped the bitmap background of our pages, and as from this issue, all articles will not only be available as HTML documents for online reading, but also as PDF downloads, as a concession to readers who are not willing to completely have their reading habits subdued by the wired world.

Another feature, which we hope will be warmly welcomed by our readers, is the addition of a new section to the journal, providing reviews of conventional books as well as of e-media and software products. As with all other changes, we would like to learn what you think about this new section. Please let us know your opinion on any aspect of the journal. Your feedback is important to us!

The origins of MPO in the Early Prehistoric Migrations project have led to a strong bias towards the Palaeolithic in the articles that were published in our first issue, although it was clear from the start that the journal would welcome contributions related to all aspects of Mediterranean prehistory. While the editors hope that this bias will diminish with the current
issue, this also depends on the input received from authors. So, not only are we hoping for a stronger feedback response from our readers, but also for more submissions from authors not involved in the Early Prehistoric Migrations project.

Ultimately, the survival of MPO will depend on keeping high academic standards while at the same time broadening the scope of articles, which ideally should range from the Lower Palaeolithic to the dawn of Classical Civilisation in the Mediterranean Basin, and also include contributions concerning the general background of hypermedia and e-publication relating to Mediterranean archaeology. Especially those of our readers who are themselves conducting excavation or survey projects in the countries surrounding the Mediterranean, and who hitherto relied on their departmental or other institutional home pages for the regular presentation of preliminary results might want to consider publishing forthcoming reports on their ongoing work on the pages of MPO and take a look at our Guidelines for Authors.

To conclude with a word to all of our readers: the most striking novelty you recently experienced when trying to access the articles published in MPO probably will have been the need to register as a reader. Before the closing of this issue EU funding of the journal will run out, and in order to guarantee an economically sustainable future, we need to know our readership much better than we do at present. This does not mean that currently we plan to charge individual subscribers, but we need your information to develop a feasible strategy to fund the journal in future years. Be assured that the information you provide with your registration will only be used for the purpose of furthering the journal, and let me thank you for the trust you put in us by supplying us with that information.

References


