ANALECTA PRAEHISTORICA LEIDENSIA 28

PUBLICATIONS OF THE INSTITUTE OF PREHISTORY
UNIVERSITY OF LEIDEN

# INTERFACING THE PAST

## COMPUTER APPLICATIONS AND QUANTITATIVE METHODS IN ARCHAEOLOGY CAA95 VOL. II

EDITED BY
HANS KAMERMANS AND KELLY FENNEMA

UNIVERSITY OF LEIDEN 1996

Graphic design: Henk de Lorm

Computer graphics: Peter Heavens

Copy editor: Marianne Wanders



ISSN 0169-7447

ISBN 90-73368-10-3

Subscriptions to the series Analecta Praehistorica Leidensia and single volumes can be ordered from:

Institute of Prehistory
P.O. Box 9515
2300 RA Leiden
The Netherlands

# contents

## VOLUME I

**Archaeometry**



**Statistics and Classification**

Mike Heyworth
Seamus Ross
Julian Richards

# Internet archaeology: an international electronic journal for archaeology

## 1 Background

In 1993 the Higher Education Funding Councils (HEFCs) published a review of libraries and related provisions in higher education in the UK, chaired by Sir Brian Follett (Follett Report 1993). The review group devoted much attention to how information technology can help to meet the needs of library users and library management over the next decade. It proposed that the funding councils should jointly invest some £20 million over three years in support of a range of activities to further the development of the electronic library. It was subsequently announced that the Higher Education Funding Councils for England, Scotland and Wales, and the Department of Education for Northern Ireland had agreed to allocate £4.75m during the 1994/95 academic year to proceed with the Information Technology (IT) recommendations set out in the report.

The Follett Implementation Group on Information Technology (FIGIT) was established to manage the implementation of many of these recommendations (see JISC Circular 4/94, *Follett Implementation Group on Information Technology: Framework for Progressing the Initiative*). FIGIT works closely with the Joint Information Systems Committee (JISC) of the HEFCs. Instead of inviting competitive institutional bids for projects in each category FIGIT invited higher education institutions and other interested parties to submit expressions of interest in programme areas. Some of the institutions which expressed interest were invited to make a presentation to the appropriate FIGIT working party, and subsequently an even smaller number was requested to submit formal proposals for funding. One of the programme areas was electronic journals, where FIGIT agreed to fund a number of initiatives to improve the status and acceptability of electronic journals, and the promotion of new forms of electronic journals and opportunities for parallel publishing.

We initially expressed an interest in establishing an electronic journal for archaeology and following an apparently successful presentation on 23 February 1995 to the Electronic Journals Working Group we were invited to submit a formal proposal for funding. Our proposal to establish an international electronic journal for archaeology received funding. The intention is that the production and dissemination of the journal will be network-based, ultimately available to all via the Internet.

The journal will publish the results of archaeological research, including excavation reports (text, photographs, data, drawings, reconstructions, diagrams, interpretations), analyses of large data sets along with the data itself, visualisations, programs used to analyse data, and applications of information technology in archaeology: for example, geographical information systems and computer modelling.

Conventional publication via the printed page cannot do justice to the rich diversity of archaeological information. Electronic publication, by contrast, offers opportunities to overcome these difficulties. The journal will be fully refereed. It will set a high academic standard. Contributions will be provided by archaeologists throughout the world, and the journal will be aimed at an international audience.

## 2 Consortium membership

The bid was led by the Council for British Archaeology (CBA) in close collaboration with the British Academy and the University of York as well as several other Universities. The University of York will act as the host for the project. The project's Managing Editor will be based in the University's Department of Archaeology and the journal will be made available via a network server linked to the University campus computer network.

The CBA will act as the publisher of the journal. The Council was founded in 1944 and is both an educational charity and a company limited by guarantee. It is primarily a liaison body made up of over 400 archaeological organisations and 2800 individual members. It works to advance the study and care of Britain's historic environment, and to improve public awareness of Britain's past. Three of its focal areas of activity are education, information and publication. It has extensive experience of journal management and will act as the subscriptions manager. Guidance will be provided through its Publications Committee and the staff of its Publications department. The CBA will also assist in the marketing of the project and the generation of project publicity.

The British Academy, founded in 1901, is a learned society, which has responsibility by Royal Charter for promoting advanced work in all the humane disciplines. Among these is archaeology. It has supported much archaeological fieldwork, publications of archaeological research including study of archaeological collections and excavation reports, and supports Schools and Institutes abroad. The Academy will provide editorial guidance from among its fellowship, technical assistance through its IT department, and will make available the expertise of its Publications Committee and department.

The other consortium members are all UK universities with major archaeology departments (Durham, Glasgow, Oxford and Southampton). Together with the University of York, they will contribute the specific expertise of individuals from within their archaeology departments to both the editorial board and technical panel.

## 3 Why archaeology?

Archaeology is a particularly appropriate subject to promote the use of electronic media as it is multidisciplinary, with a wide variety of data types. Much archaeological work is by its very nature destructive — it is only possible to excavate a site once — and archaeologists therefore need access to primary data in order to repeat and test conclusions, and reanalyse data to apply alternative hypotheses. New computer tools are being developed to allow archaeologists to make statements about the data they collect which were not previously possible. Traditional methods of publication cannot provide the functionality that these new methods and data types require.

The electronic journal will allow archaeologists to distribute full excavation data, in addition to their interpretations, allowing other researchers to reanalyse the material to confirm conclusions or to draw new conclusions. It will also be possible to distribute photographs, drawings, and dynamic reconstruction images, together with the computer programs that were used to analyse the data; and this is particularly important where the analytic programs hide hypotheses that might have influenced the analysis of the data (Ross 1995).

## 4 Why an electronic journal?

Electronic publication offers new possibilities for the display and interpretation of archaeological data that is not possible through conventional publication. However, there is some need for culture change if electronic journals are to become common (RS/BL/ALPSP 1993; Vickers/Martyn 1994). Difficulties include: problems with refereeing and guaranteeing quality; lack of standards governing the citation of electronic publications; the fact that electronic documents appear to have an ephemeral character; difficulties of access; fundamental questions in the minds of academics on what constitutes a publication and what it means to be published; and that electronic publications cannot be accessed without other hardware and software (BLRDD/BA 1993: 30).

This electronic journal will gradually help to break down barriers to these new approaches. Three factors will cause this change. The first will come about naturally and results from the general availability of computing equipment, software, and wider access to electronic resources. The concomitant collapse of psychological barriers to digitised information will take place as more people realise that, with electronic resources, it is possible to do things with the sources that could never be done with their printed counterparts. The second factor has to do with guaranteeing quality. Journals with poor quality content have no financial viability, whether stored in print or available via the networks. An academic refereeing process is required to make certain that digitally distributed scholarly articles achieve the same standards of excellence found in their printed counterparts. Third, many objections to electronic publication are disappearing as publishing houses add electronic imprints and series to their lists. There is a fear, and one that probably has a great deal of justification, that publishing electronic material will lead to the scholarship being marginalised by colleagues. This journal will aim to make certain that this does not happen because it will be both refereed and sponsored by the UK's premier learned society for the humanities and social sciences, the British Academy. We aim to ensure that electronic imprints are accorded the same scholarly recognition as any print journal.

## 5 Journal content

The journal will present its material in four sections:

1. general articles on archaeological issues whether theoretical, methodological or analytical studies;
2. excavation reports and finds studies;
3. the application of new techniques, such as software tools or the application of visual methods to archaeological analysis;
4. reviews of technological applications such as databases and other services available on the network.

### 5.1 Software

By defining the needs of archaeological publications as multimedia based, and the goal of the journal to deliver multimedia information, it is clear that the delivery tools must be both capable of displaying this kind of material and of running in a client-server environment. Several other

characteristics are also essential: the interface model must be reusable and tools widely applicable; where possible based on currently available network access and retrieval software (such as WWW browsers); be capable of accessing and running other software packages, of displaying still and moving images, data, and text and its use must put minimal financial and training overheads onto the journal reader; no substantial development and deployment costs must be borne by the journal.

The interface required shapes our approach to the rest of the software. Four categories of software are involved:

1. the software the user will run to access the journal (described above);
2. the software to create the HTML marked-up journal (e.g., query forms, links between data, charts, etc.);
3. the application programs for interfacing between the Web forms or reports and the databases or other information services which lie behind them;
4. the Web server program.

Users should use WWW client software: either public domain versions such as Mosaic or a commercial version such as Netscape. The users will also need to use viewers for images, charts and graphical representations. Once again these will be public domain tools wherever possible. This constrains the server technology to a Web application — CERN Web server software will be adequate for this. However, for the development of the journal articles it will be helpful if the Managing Editor has access to layout and design software that will ease the production of multimedia networked applications. The project will use Silicon Graphics WebForce software for the development of the journal front-end. Although we are generally avoiding using bespoke or commercially developed software, WebForce has a rich set of tools to aid the implementation of Web-based articles such as those required for HTML mark-up.

Between the Web server technology and the application databases, spreadsheets, information, or software, we will need programs which take input from the reader and translate it into a query or request which can be used by the appropriate software as the basis of queries. When the result is returned from the program it will need to be formatted before it is passed to the Web client who made the request. These programs will be written on an ‘as-needed’ basis (that is in designing an article the software will be produced if data in the article requires access to the programs). The objective will be to keep bespoke programming to a minimum.

It will be necessary to produce access control software. This may need to be a special development, although we do not believe that the software itself will be unduly complicated. What will be needed is a way to verify that readers have a legitimate right to access the journal. We are currently investigating how this might best be achieved.

All this has the caveat that we do not feel that locking the journal into a single delivery technology would be wise. For this reason we are proposing that the delivery strategy should be reviewed on a regular basis. For instance it may turn out that the use of Portable Document Formats (PDF) will supplant the use of HTML marked-up files in the future and we would want to be in a position to migrate the journal forward.

### 5.2 Hardware

The journal will require three kinds of hardware: development, delivery and reader. The development hardware will be used by the Managing Editor to produce and test the journal. The complexity of integrating text, images, data, programmes and other digital sources requires access to a Unix workstation class machine which has excellent visual representation tools. The work will be time consuming and heavily screen based. To improve productivity the development machine must be capable of running authoring tools for Web production.

The project also requires a delivery machine which can host the journal and provide processing power to manipulate the data which forms the foundation of each of the articles. This machine must be capable of running a Web Server and a full range of database and visual representation tools. It must be connected to JANET/Super JANET by a high bandwidth (minimum 4Mb) line. Articles and software will be worked up on the development machine and periodically archived to the delivery machine. We believe that a separate project server for journal delivery is essential to provide independence from the future technical developments of the host site, and will also make access control significantly easier to manage.

The reader machines will be a definition of standard hardware that a subscriber must have access to if he/she is to read the journal. Current thinking is that the machine must be capable of running Mosaic or Netscape and other public domain visual display tools. This suggests that a minimum of a 486 DX33-based machine with 8Mb of RAM and 20Mb of free disk space, a 24-bit graphics card with 1Mb of video memory onboard would be required to read a single article in any issue. In the short term the journal will be designed to be delivered at one of two levels, users with faster connections will be able to take the high-level version whereas those with low speed modems will only be able to take the low-level version. (A critical element in the design of the journal will be ‘delivery time to the user’s desk’ of the articles.)

### 5.3 Standards

The key to the long term viability of the journal will be the choice of portable standards which will make the data and information accessible from a wide variety of hardware platforms and software packages over a long period of time. In practice this means that while each issue of the journal must appear to the reader as a coherent piece which could be read or browsed from beginning to end, it must be capable of being broken up into its constituent parts for archiving. By doing this it will be possible to store the components in standard and widely used file types whether these are images, text, data, or CAD files. No files will be stored in proprietary data formats except as a last resort or where the standard has been commonly adopted. This will make it possible for AutoCAD file formats and Kodak's Photo-CD image definitions to be used for reconstructions or site records.

Each journal will therefore consist of its Web definition, which will be accessible only as long as the current generation of HTML-based Web software is being used; the files which the user front-end accesses; and documentation which details the interrelationship between the components so that the article could be reconstructed if necessary from its parts. Where for instance an article depends upon its visual sources to make its argument, details of these will need to be preserved along with a description of the hardware that is required to view the material.

We do not wish at this time to list all the standards that we will use, since to do so would pre-empt work which will be done by the Arts and Humanities Data Service when it is established (Burnard/Short 1994). One of the main objectives of the Data Service will be to outline data standards that can be applied to the creation of data if it is to be stored for long periods. But for the record it may be worth mentioning just a few standards which will be applied. Text will be stored as HTML marked-up files, images as either Graphical Interchange Format (GIF) — assuming the current copyright problems are resolved — or Photo-CD files, where possible dbf for PC-based databases, or ASCII comma-delimited files, and dxf formats for CAD renderings. Surprisingly, the most commonly used kind of data has the least widely established data standard. This is tabular data and the problem is particularly acute where multi-valued fields are used or the data is stored in Unix-based database packages. This issue will need to be addressed by requiring all data to be converted into a standard SQL package which will be running on our server and special hooks will need to be written to access this material. In all cases documentation on the data formats and coding used will be included within all articles.

The question of data compression for both storage and delivery is important. For still image compression the journal will rely on JPEG and for moving images it will use MPEG because these are the industry standard. Whether we will be able to use this for material that is being delivered to the end user is doubtful because to uncompress files at the reader's machine would be both time-consuming and require substantial amounts of processing power.

The focus will be on data standards that are widely accessible and do not pose significant long term storage problems.

### 5.4 Project management

A major problem which faces consortium-based journals or projects is that they are basically a loose confederation of members without any firm foundation beyond the agreement of the participants to take part. This problem may be less important for some projects, but it is a pressing issue for an archaeological journal which must plan for the long-term accessibility of the information which it publishes. In an attempt to provide a bond between the consortium members and to ensure that the journal has a foundation which will guarantee the preservation of the material it publishes, we plan to establish a charitable trust to own the journal. This is common practice for ongoing non-commercial projects, such as journal publications. For example, *Antiquity* the main international print journal for archaeology is owned by the Antiquity Trust, a registered charity founded in 1927.

The trustees will be the members of the Steering Committee, who would be responsible for overseeing the entire project on behalf of the Trust. This has the advantage that no single organisation within the consortium has significant control of the journal, and that the journal is seen to be independent. No single organisation can unduly influence the project and the project is not dependent on the continuing existence of any one organisation. The constitution of the trust will ensure that it can only use the project income to develop the journal.

### 5.5 Refereeing

One of the main tasks of both the editorial board and the technical panel will be to oversee the refereeing process which is essential for quality control and academic credibility. We envisage a two-stage refereeing process, all undertaken electronically via the network, based on concept refereeing and product refereeing. Initial concept refereeing will be undertaken by the editorial board, together with external referees where appropriate, who will firstly assess the academic quality of the proposal, followed by the potential of a contribution for its suitability to be published in the journal and the validation of the underlying resources

that will be required. Contributions will be selected on their merits for electronic dissemination and their potential to make full use of the new media, as well as on traditional scholarly factors. The editorial board will also need to ensure that each issue of the journal is balanced, particularly in view of the varying quantity of production work to be undertaken by the Managing Editor that will be required for the different articles. Once a contribution has been provisionally accepted for publication then contact will be established between the Project Managing Editor and the contributor to work together to produce the final electronic article. It is assumed that in the early stages of the journal there will need to be significant input from the Managing Editor to prepare material for publication; however, as archaeologists become more familiar with electronic publication and the advantages it offers we assume that less work will be required to rework submitted material. The product refereeing will be managed by the Managing Editor, and must be undertaken over the network to allow referees to evaluate the contribution as it will be viewed by users, to ensure quality in content and delivery effectiveness. Refereeing will take place via the development server which is only accessible by the editorial board and technical panel members and a select group of referees. The whole refereeing system for the journal will therefore be an iterative process, undertaken entirely electronically, managed by the Managing Editor.

### 5.6 Dissemination

Protecting the copyright of electronic journals poses major obstacles to the widespread dissemination of information through computer networks. The journal will require contributors to grant it a non-exclusive licence to publish their material (including photographs and data) in electronic form. It will protect its own copyright and the intellectual property rights of its contributors by requiring users to accept licensing terms to access and use the journal. The licence will govern accepted usage as well as stipulating that readers follow defined rules of citation (including author and article titles, etc.).

A method of accepting the licensing terms will be provided online. The intention is to restrict the redistribution of the articles or the data included in them, but we will encourage secondary analyses of the data itself. We will encourage users to publish reinterpretations of the data disseminated with articles in the journal, and stimulate debate on interpretations both in the journal and through the associated listserver mailing list.

Copyright and intellectual property rights in the articles and data will be vigilantly protected to assure the journal a continued revenue stream from its collection of core materials. We will keep under review the possibility of using a variety of information fingerprinting techniques to make it possible for the journal to protect its material.

As mentioned previously, we are looking into ways of taking advantage of the Internet functionality to verify the rights of access of subscribers. This will be supplemented by a password protection system which is made simpler by the use of a separate network server for the journal.

### 5.7 Subscriptions

The main source of income for the journal will come from 'subscriptions' and other access charges. The focus will be on revenue stream diversity to ensure the maximum flexibility in charging for users and to provide a sustainable long-term revenue stream for the journal.

The majority of users of the journal will pay a straightforward subscription charge to access the journal issues for a particular calendar year. There will be individual and institutional charges, with the institutional subscription based on access availability for all members of the institution and the subscription being proportional to the size of the potential user-base. It is hoped to keep subscription charges low to maximise the subscriber base and to reduce library subscription costs. From studies of subscriber numbers to international print journals in archaeology we believe that it will be possible, within three years, to gain sufficient income to make the journal self-supporting.

Subscribers to the electronic journal would only have access to the issues for the year in which they subscribe, though in due course they could also purchase the rights to access issues from previous years. We plan to implement a scheme to allow free content searching of issues of the published journal. This will allow users to check whether there is any material of direct relevance to their research interests in the journal before they subscribe. We will then charge for retrieval of individual articles for a variable fee depending upon the type of article, whether access to a particular set of material requires processing on our host server, the quantity of data, and the currency of the material.

### 5.8 User and community feedback

The network server which hosts the journal will also run a listserver program which will allow the journal to have a dynamic discussion list based on the journal's contents. The listserver will encourage discussion on the academic contents of the articles, as well as providing feedback to the Steering Committee on the quality and applicability of the journal issues. We will also investigate the possibility of including a WWW-based letters section linked to the journal which would be dynamically updated and available on open access.

**6 Conclusion**

In conclusion, the objectives of the project are as follows:

a. a regular electronic journal
b. a detailed description of the process of establishing and managing an electronic journal
c. definition of a suite of access and navigation tools that will allow the readers to use the journal
d. a contribution to cultural change through the increased use of electronic media.

It is the intention that the first issue of the journal will be available within a year of the start of the project.

## references


BLRDD/BA. British library research and development department and the British Academy 1993 *Information Technology in Humanities Scholarship: British Achievements, Prospects, and Barriers*. London: British Library R&D Report 6097.

Burnard, L. H. Short 1994 *An Arts and Humanities Data Service: Report of a Feasibility Study Commissioned by the Information Systems Sub-committee of the Joint Information Systems Committee of the Higher Education Funding Councils*. Oxford: JISC.

Dempsey, L. 1994 Network Resource Discovery: A European Library Perspective. In: N. Smith (ed.), *Libraries, networks and Europe: A European Networking Study*, London: BLR&DD, LIR Series 101.

Denning, P.J. B. Rous 1995 The ACM Electronic Publishing Plan, *Communications of the ACM*, 38(4), 97-103.

Follett Report 1993 *Joint Funding Councils' Libraries Review Group: Report*. Bristol: HEFCE.

Lesk, M. 1994 Electronic Libraries and Electronic Journals, *The Tenth British Library Research Lecture*. London: The British Library.

Ross, S. 1995 Introduction: Networking and Humanities Scholarship. In: S. Kenna/S. Ross (eds), *Networking in the Humanities*, xi-xxiv, Sevenoaks: Bowker-Saur.

RS/BL/ALPSP 1993 *The Scientific, Technical and Medical Information System in the UK: A study on behalf of The Royal Society, the British Library and the Association of Learned and Professional Society Publishers*, Chairman: Professor B.R. Coles. London: The Royal Society, BLR&DD Report 6123.

Vickers, P. J. Martyn 1994 *The impact of electronic publishing on library services and resources in the UK: Report of the British Library Working Party on Electronic Publishing*. London: The British Library, LIRR Report 102.



Mike Heyworth
Council for British Archaeology
Bowes Morrell House
111 Walmgate
York Y01 2UA
United Kingdom
e-mail: m.heyworth@dial.pipex.com

Seamus Ross
The British Academy
20–21 Cornwall Terrace
London NW1 4QP
United Kingdom
e-mail: seamusr@britac.ac.uk

Julian Richards
Department of Archaeology
University of York
The King's Manor
York Y01 2EP
United Kingdom
e-mail: jdr1@york.ac.uk