

Making the LEAP: Linking Electronic Archives and Publications

Julian D. RICHARDS^{1,2} – Judith WINTERS² – Michael CHARNO^{1,3}

¹Archaeology Data Service

²Internet Archaeology

¹jdr1@york.ac.uk

²jcw7@york.ac.uk

³mdc502@york.ac.uk

Abstract

The aim of the LEAP project <<http://ads.ahds.ac.uk/project/leap/>> was to investigate novel ways in which electronic publication over the internet can provide broad access to research findings in the Arts and Humanities, and can also make underlying data available in such a way so that readers are enabled to ‘drill down’ seamlessly into online archives to test interpretations and develop their own conclusions. The LEAP project used the existing publishing infrastructure of the e-journal *Internet Archaeology* and preservation systems of the Archaeology Data Service (ADS) to provide four sustainable exemplars of multi-layered e-publications and e-archives. The four exemplars selected for the project were Medieval Whittlewood, The Troodos Archaeological and Environmental Survey Project, Silchester: Roman Town Insula IX, and Ancient Merv. Users can experience these project outputs from either the publication or archive level, and navigate seamlessly between the two. This paper reviews the four exemplars and addresses the editorial and technical issues raised by their implementation.

Keywords

ICT, digital publication, digital archiving

1. Background

The growth of the internet has been described as the most important event since the invention of the printing press. The Arts and Humanities, however, have been relatively slow to embrace electronic publication. Traditional publication of research projects often comprises a series of journal papers, a popular summary, an academic monograph, and sometimes an offline research archive, which is rarely accessible. There are growing numbers of online databases, but these are rarely linked directly to interpretative analyses. In addition, the Arts and Humanities community has no conceptual framework for evaluating online databases, which results in uneven and unsystematic reward for those who create them. Given that supporting data are invariably in electronic format, the advent of e-publication allows the combination of different forms of dissemination and the adoption of an integrated approach.

These issues are not unique to Archaeology. Other disciplines are now making greater use of a broader range of multimedia applications. In the historical disciplines, supporting data often exist in such quantities that traditional publication is prohibitively expensive and only a subset can be printed. This makes

it imperative to preserve a research archive which can be used to test and refine published interpretations. As the Arts and Humanities embrace electronic publication and develop electronic resources, there is a need to ensure wide dissemination of best practice, through a series of exemplars.

The potential of linking publications and archives has been discussed (Richards 2002) but few examples have been published (Clarke *et al.* 2003; Millett *et al.* 2000; Richards 2001; Theatron 2002). The LEAP (Linking Electronic Archives and Publications) project, funded by the Arts and Humanities Research Council (AHRC) under the ICT Strategy Programme, aimed to explore some model solutions, and archaeology, and the related cultural heritage disciplines, provided an ideal test-bed for exploration. The use of ICT is relatively advanced and there is a rich variety of data formats, including text and colour images, and also databases, GIS, VR visualisations, video, and geophysics.

2. Aims and objectives

The LEAP project used the existing infrastructure of the e-journal *Internet Archaeology* <<http://intarch.ac.uk>> and of the Archaeology Data Service

<<http://ads.ahds.ac.uk>> to provide four sustainable exemplars of multi-layered e-publications and e-archives, capable of wide implementation across the arts and humanities. In order to demonstrate the extensibility of the project across the Arts and Humanities domains, exemplars were chosen that were of thematic or cultural interest to other communities.

The exemplars attempted to provide a novel and imaginative additional form of dissemination for four research projects of high academic quality, and investigate the ways in which e-publications can be interactive, multi-layered and underpinned by supporting data. This was achieved for example, by deep linking to individual elements of the online archive (held by ADS) from the interpretative narrative (*Internet Archaeology*), embedding an online GIS of archive data within the narrative, or making data queries (linking) from within the text to the searchable data in the archive.

3. Exemplars

Two calls for papers were widely publicised, the first in June 2005, and the second in January 2006, with £4000 grants offered to facilitate staff buy-outs or the hiring of research assistants to assist with the preparation of the exemplars. Selection was undertaken by the Internet Archaeology Editorial Board and ADS Advisory Committee in conjunction with other Arts and Humanities Data Services heads.

Most applications were from archaeology and closely related fields, partially reflecting the higher level of ICT development and awareness within the discipline. Exemplars were selected on the basis of (i) academic significance, (ii) potential of value-added by the ability to drill down from synthesis to primary data, (iii) ability to deliver within the project timescale, and (iv) inter-disciplinarity. Priority was also given to projects that had received AHRB/AHRC funding but which might benefit from an additional means of dissemination.

4. Changing settlements and landscapes: medieval Whittlewood, its predecessors and successors <http://intarch.ac.uk/journal/issue19/jones_index.html>

This article presents an interpretative synthesis of the development of a medieval landscape in the English Midlands. The reconstruction of medieval village territories was only achieved by adopting an interdisciplinary approach employing a wide range of techniques. By presenting all the data on which reconstructions of these territories have been based, readers are able to test the veracity of the conclusions outlined here and in the monograph, and to identify the intrinsic strengths and weaknesses of each class of evidence. Readers are encouraged to explore their own research agenda and to develop different readings of the evidence on which alternative models of medieval settlement and landscape change can be built.

The link between *Internet Archaeology* and the ADS archive was through a web GIS interface. This allowed users to access and interrogate the spatial data from the archive directly alongside multiple interpretations within the article. In addition to the standard spatial and attribute data within the web GIS interface, spatial features were also linked to related images and finds tables when queried. This additional level enabled users to “drill down” from the interpretation within the article through to the spatial data of the web GIS and then finally into otherwise un-associated digital objects within the archive.

5. Joining the dots: continuous survey, routine practice and the interpretation of a Cypriot landscape <http://intarch.ac.uk/journal/issue20/taesp_index.html>

One of the major challenges facing intensive surface survey is how to interpret surface artefact scatters in terms of past human activities and relationships. This study uses web based GIS and database technologies to provide a complete landscape data set and a fully integrated interpretative text carefully grounded in current landscape theory. The material comes from the Troodos Archaeological and Environmental Survey Project (TAESP), which carried out intensive survey in the northern foothills of the Troodos

Mountains in central Cyprus between 2000–04. This survey covered all periods from the Neolithic to the present day, and a broad spectrum of disciplinary and interdisciplinary expertise.

Like the Whittlewood article, TAESP used a web GIS interface with spatial data served from an ArcIMS server hosted by the ADS. The TAESP project already had a design and layout in mind for the article, which was adhered to as closely as possible. The web GIS interface was set on the left side of the web page while the text was positioned directly adjacent to the right (Fig. 1). This was done to allow users the ability to read the interpretive text and look at the spatial data without requiring any scrolling or adjustment of their viewing window.

The data which made up the web GIS was very complex and consisted of 20 shapefiles and 55 rasters to create 350+ unique layers. To simplify this data and lighten the load on the ArcIMS server, views of the entire GIS project were established by the TAESP authors which were related to specific sections

of the interpretive text. That enabled the user to view maps by selecting highlighted text within the article and therefore see what data lead to a specific interpretation. The only shortcoming of this design was that users could not see layers not associated with that view. Therefore users were unable to add layers of data that, according to the authors, were not related to the interpretation. The article did, however, include a map which consisted of the entire spatial data set to allow for independent exploration of the spatial data set.

Users could also query the project's database from within the article. One way was through a standard database interface which allowed users to construct their own queries and explore the results, but links were also provided throughout the text which went directly to database entries.

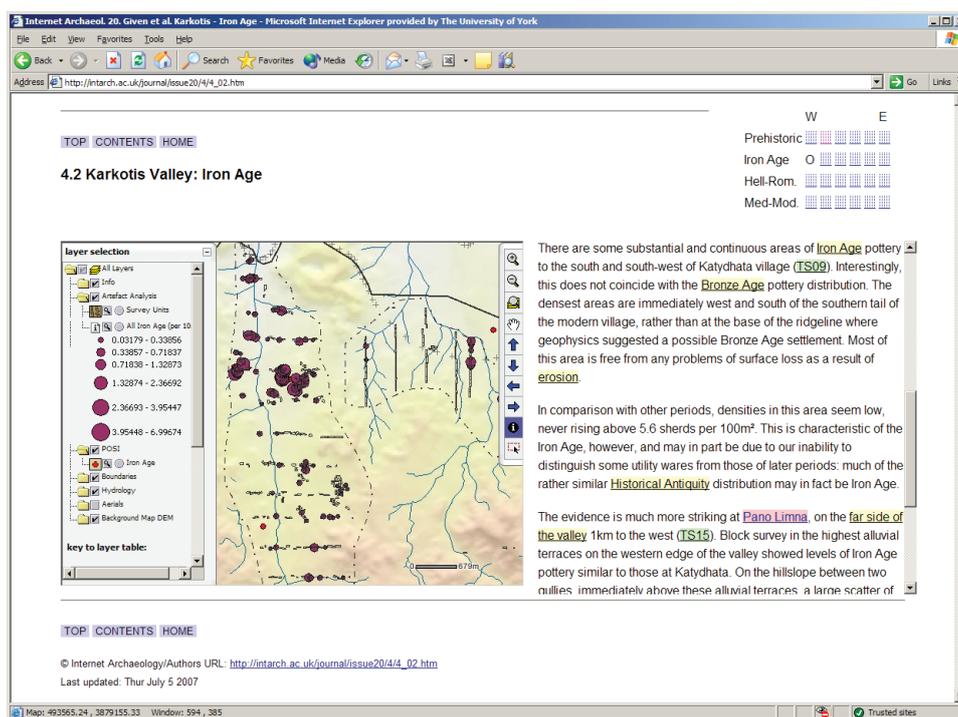


Fig. 1. A screenshot of the TAESP interface in internet Explorer 6.0. The top and bottom left of the screen (top, contents, & home buttons) show the general article navigation. The section 4 navigation grid (top right) allows users to select content based on temporal (Prehistoric to Medieval-Modern) or spatial (West to East) conditions. In the body of the page, the web GIS interface and interpretive text are located alongside each other. Within the interpretive text is yellow, green and pink highlighted text. The yellow text opens a different view for the web GIS, the green text links to the project database, and the pink text links to images from the archive.

**6. Silchester Roman town insula IX:
the development of an urban property
c. AD 40–50 – c. AD 250** <http://intarch.ac.uk/journal/issue21/silchester_index.html>

This article discussed the development of an urban property in the Roman town of Calleva Atrebatum (Silchester) from the late 1st to the mid-3rd century AD. Three successive periods of building with their associated finds of artefacts and biological remains were described and interpreted with provisional reconstructions of the buildings. Links were provided to a copy of the Integrated Archaeological Database (IADB), archived by the ADS, holding the primary excavation and finds records. A ‘snapshot’ of the IADB was necessary since the Silchester project is not yet complete and therefore the dataset is still changing. The authors will be re-submitting the full dataset but this could potentially cause referencing problems in the future if some of that re-submitted data has changed in any way (e.g. if archived material no longer supports a conclusion made in the narrative that links to it). The archive should grow and be added to, so it will be necessary to archive both the snapshot and final archive.

The copy of the IADB was hosted on an ADS server and was linked directly from the article. The links gave the reader the chance to dive directly into the archive and then navigate laterally through the archive. For example, a link from the text mentioning context 4774 would bring the user to the ADS archive and a page with data about that context. Context 4774 happens to be the fill of context 4790 and contains 4 small finds, all of which are linked to from the results page. This allows users to “drill down” into the archive from the interpretive text and then effectively “drill sideways” through related contexts and finds.

**7. The landscapes of Islamic Merv,
Turkmenistan: where to draw the
line?** <http://intarch.ac.uk/journal/issue25/merv_index.html>

Merv straddles one of the main branches of the ancient Silk Roads that connected Europe and Africa to the Far East. The succession of cities date from the 5th century BC to the present day. The article develops and tests the methodology of documenting interpretation (and uncertainty) based on a variety

of data sets. The publication will have a web GIS interface similar to the Whittlewood and TAESP projects as well as video and audio content, and sections of the text will link directly, for example, to images held by the ADS.

The LEAP project raised a number of questions due to the novel nature of dissemination, both from the publication and the archive standpoint. Many of these questions inspired useful and productive discussion on digital dissemination in general which went beyond *Internet Archaeology* and ADS standard procedures and which will feed back into future updates to those procedures.

8. Editorial questions

Resulting resources should be assessed and peer-reviewed. Since we were deliberately blurring the boundaries between the electronic archive and the publication, should the peer-reviewer also comment on the archive data to the same level of detail as they would the narrative ‘text’ and supplementary material? It is common *Internet Archaeology* practice to make all supplementary materials available for the referee. But when this also constituted the data deposited with the archive then this makes a truly full assessment a much larger undertaking by the selected referee.

The ADS has well-established procedures for accessioning data, which were applied to the data received as part of the LEAP project. Obvious errors are pointed out during this process to the depositor but depending on the size of the archive, this can become a very time consuming process, so realistic expectations must be applied to the accessioning process. While not every piece of data can always be investigated, a comprehensive selection can confirm the quality of the data while not dedicating an excessive amount of time to the process. More important is checking that information about the data’s provenance and associated metadata will allow future users to make sense of the data. However, further down the line, once linking work got underway on the narrative text (IA) with deep links being made to archive holdings, other errors in the data content could be observed and resulted in further updates to the deposited data (made by the ADS curator and not the IA editor) which would have otherwise gone unnoticed. This level of unpicking and scrutinising of data essentially resulted in a further stage of refereeing. This is something that is

not done with traditional publications as the entirety of the underlying data is usually not included, with publishers opting instead for a subset of the data in the form of maps, tables, and images.

One unforeseen outcome of the LEAP project was the question of how archives should be structured so that they could be linked with minimal intervention. This becomes a very serious issue if the e-publication and archive do not share a common infrastructure. Furthermore, it was important to manage author expectations. Those who were used to Desktop GIS, for example, found it difficult to accept the reduced capabilities of web GIS. Some early design visions clashed with accessibility and usability requirements so robust management was necessary.

Some more general challenges arose. The LEAP project worked with the intention that the narrative published in *Internet Archaeology* should tell a specific story that is dependent upon links to a data archive. But we realise that in doing so, we run the risk of forcing one particular interpretation upon the reader with the possibility that further research on the data may actually be harder to carry out. There is a fine line between supplementing the data and informing the reader and controlling the data they can look at.

9. Citation questions

An issue of citation highlighted by the LEAP project was the extent to which intellectual credit could or should be given to all authors of the integrated resource. Now that the archive is essentially part of the publication, does everyone involved in the creation of it, the databases or GIS, deserve to be listed as an author? If this were the case there could be dozens of authors due to the way archaeological projects are frequently conducted.

An equally difficult question was how queries or views into the archive data should be cited, or even if they should be cited at all. This situation is the product of the deliberate ambiguity of where the publication ends and the archive begins. If the archive should be cited in a similar manner to a journal article then this could be taken to its logical conclusion that any query run against a database should be cited. In reality the publication is merely linking to an archive, so while the interface into the data archive may be embedded in the publication, it is still simply an interface. As such, it was concluded that citation of the archive as a whole should be made when referencing data

from a linked electronic publication and where a researcher identifies something in the data that has not been mentioned in the interpretation, then both the archive and the publication should be cited.

10. Copyright questions

Copyright will always be relevant when an individual's work is being disseminated by another party. When a further party is added into the equation, such as with the LEAP project, the copyright issue becomes more confusing because the electronic publication or archive may have different or non-compatible licenses. *Internet Archaeology* and the ADS each have licenses which permit the sharing of data and resources, so this was not an issue in this instance, but it could be a key question for anyone trying to follow this path.

11. Sustainability questions

A significant concern of the LEAP project which cannot be answered immediately is the question of interface sustainability. Information and communication technologies develop at breakneck speed. Left in that wake are the web developers and designers trying to keep their interfaces and applications in a functional state. This is both a short term and long term difficulty, although the long term questions are of the most concern. There is a desire to keep interfaces consistent throughout time to avoid confusion on the part of the user. This becomes a very daunting and unpredictable task due to the nature of the web and its associated technologies.

Exacerbating the potential problems are the uncommon and bespoke interfaces necessary for web GIS and databases. While web browser developers endeavour to maintain backwards compatibility, this is not guaranteed; nor should it be assumed. As browsers evolve, maintaining these interfaces could become a massive undertaking. The ADS takes the approach of preserving the raw, underlying data rather than the interfaces through emulation. With the data preserved and migrated as formats evolve, the ADS can build new interfaces to present the material as necessary. It will be important that these new interfaces have a look and feel similar to the original interface to ensure continuity.

The ADS and *Internet Archaeology* share a common technical infrastructure and both currently

create their web GIS interfaces with ArcIMS and their database interfaces in Coldfusion. Both of these are proprietary solutions which have potential migration issues as they become obsolete. This will not affect the raw data but will potentially create more work for future editors/data curators as they are required to develop new interfaces for the data. The future migration of LEAP interfaces remains the biggest unknown for this project, and one which we will be unable to answer until that time comes. It might be that the migration is a simple 30 minute job 10 years from now or a 10 day job 30 years from now, but we will not know until we get there.

12. Culture change

Finally, before future LEAP-style publications can be produced, a shift in the way archaeologists prepare, create and think about data must occur. Traditionally the archive has been something of an afterthought, pieced together after the fieldwork is finished. If the archive becomes part of the publication, however, more care from the outset of an archaeological project must be taken with regards to data creation. This means that full metadata and documentation of the data must be created. Putting the raw data alongside the publication may force data producers to take more care in creating their data. Archaeologists can sometimes create data knowing that few outside the project will ever see it. This new exposure will hopefully force archaeologists to break the uneven data management habits that have historically afflicted the discipline.

Making the raw data more accessible also makes the author more open to scrutiny. This may not always be welcome. While one of the goals of the LEAP project was to allow users to draw alternative conclusions, archaeologists have been historically protective and defensive of their data. We hope that as more archives are linked in with publications, scholars will themselves become more open to allowing others to interrogate and re-interpret their data. To accomplish this will take more than a few exemplars, but rather a change in the way archaeologists think about their data in the larger context of the discipline.

13. Conclusion

The LEAP exemplars have already had significant impact in the ongoing debate about the development of

electronic publication. The number of people directly involved in the project through the four exemplars has had an immediate impact on the dissemination of good practice. Individual publications and archives have each seen high levels of usage, although it is too early to analyse general trends. Nonetheless, the web access statistics show high levels of usage for each exemplar from the outset, including individual archive files being downloaded an average of over 100 times per quarter. The access statistics also demonstrate that users are navigating from article to archive, but also that as many are navigating from the archive back to the article. Authors were keen to link from the narrative to archive file level but fewer links went in the other direction. Links from archive to article are possible and in future developments we intend to look at ways in which the archive could link back to the narrative.

References

- Clarke, Amanda, Michael Fulford and Michael Rains (2003). Nothing to Hide – Online database publication and the Silchester Town Life Project. In: Martin Doerr and Apostolos Sarris (eds.) *CAA2002: The Digital Heritage of Archaeology*, 401–4.
- Millet, Martin, Francisco Queiroga, Kris Strutt, Jeremy Taylor and Steven Willis (2000). The Ave Valley, northern Portugal: an archaeological survey of Iron Age and Roman settlement. *Internet Archaeology* 9. (http://intarch.ac.uk/journal/issue9/millett_index.html) Accessed 27 May 2008.
- Richards, Julian D. (2001). Anglian and Anglo-Scandinavian Cottam: linking digital publication and archive. *Internet Archaeology* 10. (http://intarch.ac.uk/journal/issue10/richards_index.html) Accessed 27 May 2008.
- Richards, Julian D. (2002) Digital Preservation and Access. *European Journal of Archaeology* 5, 343–67.
- Theatron (2002). <http://www.theatron.org/> Accessed 27 May 2008.