

DigiCULT .Info

Issue 6 A Newsletter on Digital Culture

ISSN 1609-3941

December 2003



INTRODUCTION

This is a rich issue of DigiCULT.Info covering topics in such areas as digitisation, asset management and publication, virtual reality, documentation, digital preservation, and the development of the knowledge society.

In October the Italian Presidency of the European Union promoted, in cooperation with the European Commission and the ERPANET (<http://www.erpanet.org>) and MINERVA

Continued next page



Technology explained and applied ...

- 6 3D-ArchGIS: Archiving Cultural Heritage in a 3D Multimedia Space
- 10 Zope at Duke University: Open Source Content Management in a Higher Education Context
- 16 PLEADE – EAD for the Web
- 31 Metadata Debate: Two Perspectives on Dublin Core
- 34 Cistercians in Yorkshire: Creating a Virtual Heritage Learning Package
- 45 SMIL Explained
- 47 Finding Names with the Autonom Parser
- 49 The Object of Learning: Virtual and Physical Cultural Heritage Interactions

News, events ...

- 23 Delivering *The Times* Digital Archive
- 30 NZ Electronic Text Centre
- 37 Cultural Heritage Events
- 52 New Project Highlights the Growing Use of DOIs
- 52 Digital Resources for the Humanities 2003
- 54 New UK Law for Preservation of Electronic Publications
- 54 Turning the Pages
- 58 Late Breaking News

Challenges, strategic issues, new initiatives ...

- 16 New Open-source Software for Content and Repository Management
- 19 The IRCAM Digital Sound Archive in Context
- 55 Towards a Knowledge Society

News from DigiCULT's Regional Correspondents ...

- 37 Bulgaria
- 38 Czech Republic
- 40 France
- 41 Greece
- 42 Italy
- 44 Poland
- 44 Serbia and Montenegro

Action in the Preservation of Memory ...

- 24 PANDORA, Australia's Web Archive, and the Digital Archiving System that Supports it
- 28 The Firenze Agenda (17 October 2003)

For subscription to the DigiCULT.Info please go to:
<http://www.digicult.info>



(<http://www.minervaeurope.org/>) activities, an international conference on digital memory preservation, *Futuro delle Memorie Digitali e Patrimonio Culturale* (16-17 October 2003 in Firenze). A major outcome of the meeting was the 'Firenze Agenda'. Building on the Spanish Resolution on Digital Preservation (2002, Document 2002/C162/02, <http://www.ibeurope.com/Newslink/311.htm#5850>), the Firenze Agenda attempts to drive forward work on digital preservation by encouraging cooperation between such players as ERPANET, Prestospace, Delos and Minerva. It identifies three areas of activity that could deliver 'concrete and realistic actions' over the next year or two. We have reprinted the Agenda in this newsletter in a new section, 'Action in the Preservation of Memory', as called for in the Firenze Agenda.

The main article in the inaugural showing of this new section is a piece by Margaret Phillips (from the National Library of Australia) on PANDORA and the Pandora digital archiving systems (PANDAS), which supports the collection of Australia's Web materials by staff at the National Library (NLA). The Australians have been among the leaders in developing strategies for preserving their documentary heritage as represented on the Web. In the middle of this year the NLA published a thoughtful review by Margaret Phillips (2003), *Collecting Australian Online Publications*, <http://pandora.nla.gov.au/BSC49.doc>, which provides illuminating background and contains a valuable examination of selection and collection strategies. As well as drawing our attention to the work of the NLA and the richness of the PANDAS tool, she notes the inaccessibility to many harvesting strategies of the 'deep Web'. The 'deep Web' is that information landscape characterised by Websites and associated information resources drawn from dynamic or static databases in response to specific user requests, or hid-

den behind password protect sites and firewalls. Two years ago Michael K. Bergman reported that 'The Deep Web: Surfacing Hidden Value' (*The Journal of Electronic Publishing*, 7.1 (2001), <http://www.press.umich.edu/jep/07-01/bergman.html>) may well be 500 times larger than the surface Web. Among his findings, based on data accurate as of March 2000, were the facts that the deep Web contained 7,500 terabytes of data and the surface Web only 19 terabytes, that on average deep Websites 'receive fifty per cent greater monthly traffic than surface sites and are more highly linked to than surface sites', that the 60 largest deep Websites were 40 times larger than the surface Web that the quality content in the deep Web is far greater than that in the surface Web, and that 95% of the deep Web is publicly accessible information. As a result we must conclude that, as attractive as comprehensive Web harvesting may be, it is far from comprehensive because it does not reach the hidden Web. Margaret Phillip's article reports that national libraries recognise this problem and that an International Internet Preservation Consortium has been established to develop common solutions.

Of course, the Internet is more than just a massive digital library waiting to be harvested, processed, stored and retrieved. Increasingly we recognise the central importance of the social space, context and interactivity that lie at the heart of the Internet. The physical and the virtual worlds are often contrasted, with the virtual world and its cyberculture viewed as uniquely different from 'real-world culture'. While it is true that there are characteristics of cyberculture that set it apart from more traditional measures of culture, the boundary between the two worlds has never been precise and continues to blur. The evolution of virtual social, information and economic spaces has demonstrated this with remarkable clarity. We are all aware that the Internet enables individuals

to share experiences, create social bonds, and construct 'imaginary communities' that take on social and cultural fabric (Ross, 2002). It is a fluid environment as anyone who has read Sherry Turkle's early 1990s study *Life on the Screen* and attempted in the past two years to investigate some of the same phenomena she describes will know. New spaces and practices are emerging all the time, older ones are disappearing, and it is transforming the ways we participate and interact.

Over a decade ago we enjoyed seeing the *Rediscovering Pompeii* exhibition twice: once in 1990 at the IBM Gallery of Science and Art (NYC) and again in 1992 at the Accademia Italiana delle Arti e delle Arti Applicate (London). Displaying a couple of hundred objects excavated from Pompeii, it gave visitors an insight into daily life in the first century AD Roman resort, illustrated how computer technology had revolutionised the analysis of archaeological evidence, and provided engaging interactives that gave visitors access to the wealth of information resources archaeologists had collected about the Pompeii region. While our understanding about how technology and digital objects can be presented has advanced considerably during the past decade, at the time the Pompeii exhibition offered an exciting early indication of how interactives in an exhibition setting could transform the experience of visitors and how underlying databases could provide users with access to information about material culture and its distribution. The sumptuous catalogue of the exhibition (*Rediscovering Pompeii*, 'L'Erma' di Bretschneider, c. 1990, Roma) provided some valuable insights into the use of information technology on the project (pp. 105-128). Occasionally I wondered what had become of the underlying databases. At the Firenze conference (see above) Alessandro Ruggiero presented 'Preservation of digital memories: risks and



Canadian National Archives

emergencies – Six Case Studies’, which satisfied my curiosity. The system had initially been switched off even before the exhibition, although it was switched back on briefly to enable the show to be created. While several of the case studies discussed by Ruggiero have been noted in other published literature, the Database of the Consorzio Neapolis (IBM and FAIT), designed to enable the exploitation of environmental and artistic resources in the area around Mount Vesuvius, has not received similar attention. The two-year project, which began in 1987, to collect and encode the data, engaged 110 staff and cost some 36 billion Lira (or at the current conversion rate 18,000,000 euros).

Ruggiero’s report on the recovery of this resource provides a window to challenges involved in digital preservation and recovery. It took some two years to recover the database and cost roughly 200,000 euros. The work ‘was made possible only due to the availability of a mainframe system similar to the original one’ and the accessibility of oral history and guidance from participants in the original project (Ruggiero (2003), p.30). Accumulating good case

studies about data loss and recovery is a challenge, as Luciano Scala, Director of the Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche, noted in his Preface to the Ruggiero study. He reported that, despite using contacts, engaging in wide-ranging discussions with colleagues and conducting interviews, ‘it was often hard, if not impossible, to gather enough information to create a comprehensive record of the problem’ of digital loss.

These case studies are essential if we are to develop an understanding of the points of risk to digital objects, and the costs and possibilities of data rescue.

Rescue of digital materials has become a major industry and *DigiCULT.Info* will examine the topic in a future Newsletter. Some of the participants at the excellent Canadian Conservation Institute Symposium 2003 on Preservation of Electronic Records: New Knowledge and Decision-Making (in September in Ottawa) were offered an illuminating visit to Tunstall and Tunstall Data Recovery (<http://www.datarecoveryservices.com>).

Staff there shared their experiences, ‘war stories’, and a few artist’s tricks – buying up hard disks for spare parts, and techniques of observation. The Symposium itself is worthy of mention because its forthcoming publication will bring together some of the latest thinking on digital preservation. Even those of us who had the privilege of hearing the papers will be rewarded by reading them (e.g. Ulrich Lang’s ‘Volatile Memory’).

In mid-November ERPANET co-hosted a workshop on Trusted Repositories with the Accademia dei Lincei (Roma) in its 16th century Palazzo Corsini. Talks by Robert Kahn from the Corporation for National Research Initiatives (CNRI) (Virginia, USA) and Professor Tito Orlandi opened the workshop. Materials from these and other workshop presentations can be found on the ERPANET Website (<http://www.erpanet.org>).



Robert Kahn and Tito Orlandi, at Accademia dei Lincei (Roma)

A few weeks before the workshop, decades of work by Professor Tito Orlandi, well known as an avid supporter of the thoughtful use of ICT to facilitate humanities scholarship, on the study of Coptic Literature and Civilisation resulted in the release of *Corpus dei Manoscritti Copti Letterari* (CMCL). This online subscription database provides resources on Coptic Manuscripts and Literature, with full texts (and translations into Italian), bibliographic sources, and information on the Coptic civilisation at large.



Tunstall and Tunstall (Nepean, Ontario) Disc Store



Disc Crash, Tunstall and Tunstall (Nepean, Ontario)

DigiCULT Forum has considered digital asset management systems from several perspectives – as a thematic issue (number 2) in September 2002 and again in March 2003 in our first Technology Watch Report. Paul Conway, Director of Information Technology Services at Duke University Libraries, has offered us an in-depth look at Zope, an open source content management system (CMS). While he has examined the use of Zope from a higher education perspective, many of the issues that he raises also apply to Cultural Heritage institutions – in particular, his conclusion that content management technology can be a ‘useful lever of innovation for libraries in a federated institutional environment’. In particular he argues that ‘university libraries have new opportunities for leadership and influence on campus.’ Could we extend his argument further by claiming that through extending its access to heritage materials cultural institutions can extend their influence? Our French correspondent notes that the New Cultural Portal launched by the French Minister of Culture is a Zope-based solution.

In the last issue Andrew McHugh introduced the topic of open source software and in this issue we take up the topic again with an article on PLEADE, a flexible

multilingual presentation framework for Encoded Archival Description (EAD) documents. PLEADE provides a set of tools to build dynamic Websites.

Given the diversity of our domain, *DigiCULT.Info* has great difficulty selecting which conferences to cover and would be delighted to receive suggestions and even conference reviews. Once again this year we report on the Digital Resources in the Humanities Conference. DRH2003 was held in Gloucester, UK, in September. The continued success of the DRH conferences reflects the increasing dependence of academic researchers and heritage professionals on ICT.

The numbers of academic institutions building humanities and arts informatics programmes is also growing. For example, this December marks the second anniversary of the launch of the New Zealand eText Centre based at the Victoria University. In a recent interview its Director Elizabeth Styron described her strategic approach to creating a sustainable unit. Her efforts are establishing an academically grounded centre that delivers effective services for public sector bodies, cultural heritage institutions and commercial enterprises.

In October the DigiCULT Forum released its *Thematic Issue 4: Learning Objects from Cultural and Scientific Heritage Resources*. The Telematics Centre at the University of Exeter has contributed to this newsletter a further article on this topic which examines issues surrounding learning enabled through virtual exhibitions. The discussion touches on such topics as the way the interaction of ‘virtual objects promoted an appreciation’ of physical objects and encouraged in virtual visitors a desire to see the real (or physical) original. They report, much as the Education team from ARKive did in the last newsletter, that interactives need to add value to content, not merely point to it.

Similarly Virtual Reality plays a significant role as heritage institutions communicate about the heritage and research cultural remains. DigiCULT has covered some of these topics in earlier issues of this Newsletter (e.g. Theatron in Issue 4) and in its first *Technology Watch Report* (February 2003). In this issue we report on *Cistercians in Yorkshire*, a project which is creating learning materials centred around 3D reconstructions of Cistercian buildings. The theme of virtual reality is further investigated by colleagues from the Cultural and Educational Technology Institute (Greece) in their presentation of work developing



3D multimedia tools for archiving cultural heritage materials from landscapes to objects. One of the strengths of their work is that it enables access to and display of other data types alongside the models. This theme of visualisation is also taken up in the closing article by the Director of the Hungarian Academy of Sciences' Institute for Philosophical Research, Professor Kristóf Nyíri, as he examines the development of the knowledge society.

DigiCULT is funded under the Fifth Framework Programme of the European Commission and, as with all such activities, the Commission appoints a Project Officer to assist our work. DigiCULT's first project officer, Axel Szauer, retired early from the Commission this autumn. One might not automatically mark the passing of a project officer, but Axel Szauer did much to ensure the success of our activities, helped us to avoid pitfalls that can bedevil projects, and gave us support and encouragement. We were sorry to see him go and wish him all the best.

Names provide a rich source for scholars, but culling them can be a slow process. Our first encounter with this scholarly domain was with the Lexicon of Greek Personal Names ([http://www.lgpn.](http://www.lgpn.ox.ac.uk/)

[ox.ac.uk/](http://www.lgpn.ox.ac.uk/)). This project, run by Peter Fraser and Elaine Matthews at the University of Oxford, aims to collect, document and make accessible (e.g. publish) for further research all surviving ancient Greek personal names from the earliest written materials to about the sixth century AD. The project has been using computer technology since shortly after its foundation in 1972 to facilitate the development of the lexicon. While the LGPN has used computer technology to facilitate its manipulating, processing and presenting, in other cases it may be feasible to automate the process of culling names from sources. Colleagues at the Netherlands Institute for Scientific Information Services have developed a tool for extracting names from literary texts that will enable research into how authors use names (e.g. types, frequencies, characterisation).

Similarly, research into the authenticity of texts and the attribution of authorship has been substantially aided by computational tools, whether it be the studies led by Sir Anthony Kenny on Aristotile, Gerard Ledger on Plato, or Don Foster on variety material from poems purported to be by Shakespeare, to more contemporary fiction and non-fiction. You may remember that, in part, because of his efforts, although less computationally based than

on other occasions, the authorship of that seasonal poem 'The Night Before Christmas' was re-attributed to Henry Livingston after having been assigned to Clement Clarke Moore for a century and a half. After 150 years, 'The Night Before Christmas' remains, even if a little too materially so, a wonderful story of warmth and seasonal cheer. All of us at DigiCULT extend you our best wishes for the season and hope that each of you may have a happy and successful 2004.

Seamus Ross & John Pereira
Editors, DigiCULT.Info



Christine Ataíde (1994), *Árvore de Natal*, National Palace of Pena

CONTACT THE PEOPLE BEHIND DigiCULT FORUM

Salzburg Research
Forschungsgesellschaft m.b.H

salzburg|research

DigiCULT Project Manager

John Pereira, john.pereira@salzburgresearch.at

DigiCULT Thematic Issue Scientific Editor

Guntram Geser, guntram.geser@salzburgresearch.at

DigiCULT Network Co-ordinator

Birgit Retsch, birgit.retsch@salzburgresearch.at

HATII – Humanities Advanced Technology and Information Institute

Director, HATII & ERPANET

Seamus Ross, s.ross@hatii.arts.gla.ac.uk

DigiCULT Forum Technology Assessor

Martin Donnelly, m.donnelly@hatii.arts.gla.ac.uk

DigiCULT.Info Content & Managing Editor

Daisy Abbott, d.abbott@hatii.arts.gla.ac.uk

DigiCULT Web Developer

Brian Aitken, b.aitken@hatii.arts.gla.ac.uk



Copyright Notice:

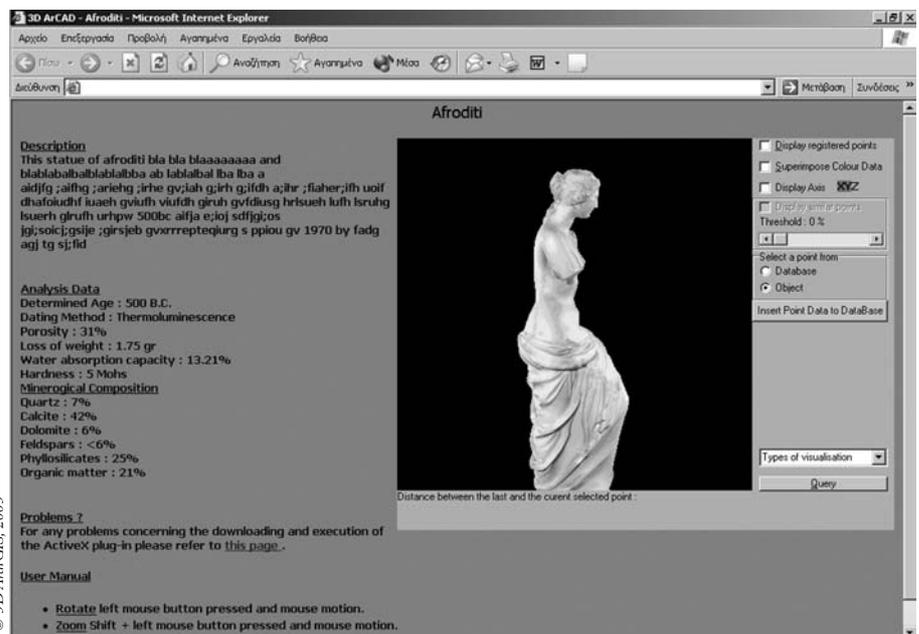
Introduction © HATII (UofGlasgow), Compilation © HATII (UofGlasgow), Images and other media © as noted at image or media access point, Layout and design © Salzburg Research, Individual Articles © named authors or their institutions. This issue may be freely re-distributed provided there is no charge for it and that it is redistributed in full and without alternation.

3D-ARCHGIS: ARCHIVING CULTURAL HERITAGE IN A 3D MULTIMEDIA SPACE

NESTOR C. TSIRLIGANIS,
FOTIS ARNAOUTOGLOU,
ANESTIS KOUTSOUDIS,
GEORGE PAVLIDIS,
CHRISTODOULOS CHAMZAS
CULTURAL AND EDUCATIONAL
TECHNOLOGY INSTITUTE, GREECE

3D ArchGIS is an experimental application that attempts to implement the capabilities and promises that new technologies bring to the field of documentation and preservation of cultural heritage. It enables scientists to map and browse physicochemical information on the surface of 3D scanned archaeological artefacts. In addition, it is a Web oriented application that provides functionality and features not commonly available from similar applications. The aim is to develop a flexible and user-friendly tool for combining and displaying various data types alongside 3D models – in some respects, an online GIS tool for objects of any size and shape, from a ceramic fragment to an archaeological site.

Preservation and dissemination of cultural heritage has always satisfied a multitude of psychological, aesthetic, social and political ambitions of humanity. The aspiration of the human race to dominate time was manifested early on with a genuine and ardent attempt to record, preserve and spread its present achievements and its cultural heritage. Vision, memory and narration were the original means used to serve this purpose, and remained basically unalterable until the revolution that accompanied the addition of writing to vision, stone and paper to memory, and



© 3D ArchGIS, 2003

Figure 1. 3D ArchGIS plugin

reading to narration. The introduction of these ‘supplements’ made the communicated information more objective, enduring and precise. Further technological advances like typography and photography have since been used to expedite the process.

More recently, the advent of new technologies and their applications have radically altered the way information is stored, archived, retrieved and presented. The enormous impact this could have on the registration, documentation, presentation and, ultimately, preservation of cultural heritage was appreciated and explored early on. Systematic recording of the physical and chemical characteristics, typological description, and historical information of cultural objects led to the first databases, mainly for research purposes.

Digitalisation of 2D images of the objects enriched the stored textual

information and made it more appealing to the public. Physical and chemical characteristics were, however, still interesting only to a limited number of researchers. Images combined with historical excerpts in the form of digital catalogues became the standard in promoting private collections and museums. Catalogues with deeper descriptions and scientific facts were also used for educational purposes and typological research. Multimedia brought a new era with virtual worlds. Relatively simple catalogues, enriched with video and graphics, were transformed into virtual museums, while multimedia databases offer a multitude of information. Yet this wealth of information remains, to a great extent, bound to 2D media.

Great advances in 3D technologies offer new opportunities to record every detail of cultural heritage in high precision, and to present it in a more

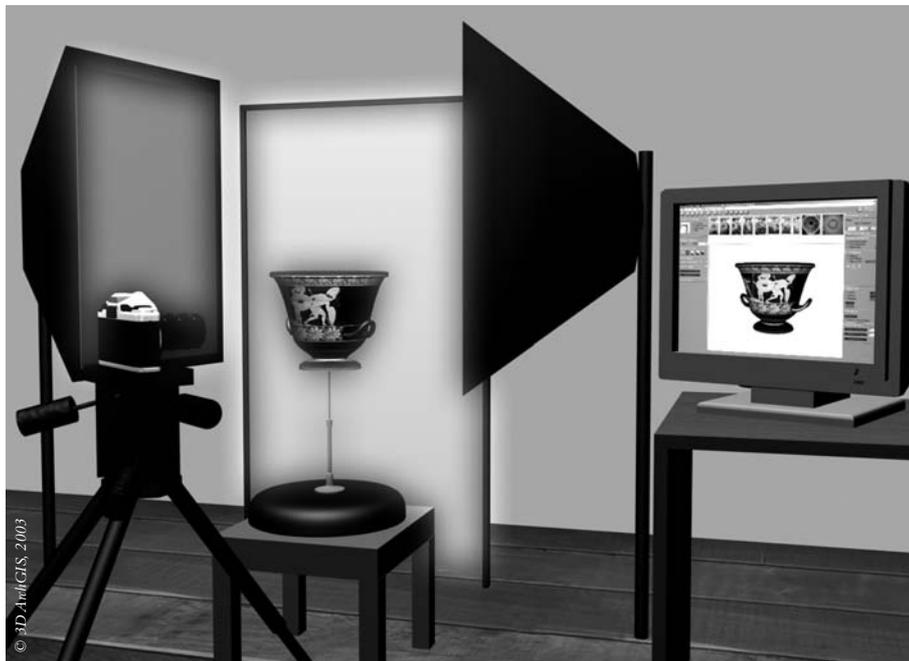


Figure 2. 3D Photographic scanner

attractive way. It is not only, however, the new imaging methods that help in the documentation and preservation of cultural heritage. In the early stages of archiving, information was derived mainly from the human senses, primarily vision and touch, since they described the perceived prominent and lasting characteristics of the articles. Advances in science and technology enriched this information and made it more complex, substantial, measurable, reliable and replicable. Today, innovations in instrumentation make possible the extraction of even more accurate, point-wise information of physicochemical characteristics and mechanical properties of objects.

The combination of modern measurements, 3D imaging, and mapping provides a field for the development of new ways to register and present information that can once again revolutionise the documentation of cultural heritage, leading to integrated and complete recording with the capability to visualise data not only macroscopically, but also in a point-size fashion, and enables the virtual reconstruc-

tion of an artefact in every conceivable detail. The effect that such a reconstruction will have upon scientific research, dissemination of knowledge and public interest is profound.

THE CONCEPT OF '3D-ARCHGIS'

The main requirements of a modern information retrieval system are: immediate access to distant stored data; an intelligent mechanism to turn raw data into useful information; flexibility to allow users to submit queries of varying complexity with a number of options; and finally a user-friendly environment. All these requirements point directly to Internet databases. There are many databases with specialised content providing services over the Internet. Specifically, a cultural database is a database that can provide information related to cultural objects, monuments, museums, heirlooms, and so on. At present, the majority of these databases' data are limited to 2D pictures and drawings, textual descriptions and tables of data, all used to describe the archived object.

The Cultural Technology Unit of the Cultural and Educational Technology Institute (<http://www.ceti.gr>), with the collaboration of the Institute's Multimedia Unit, initiated an attempt to incorporate the latest technologies and methodologies into an integrated documentation environment for cultural objects. This began with the development of a multimedia database initially focused on archaeological ceramic and glass artefacts. The database should include detailed 2D and 3D images of archaeological finds accompanied by morphological-typological descriptions, historical and scientific data such as dating measurements, mechanical properties and stoichiometric analysis that – where appropriate and possible – will be mapped on the 3D image.

A combination of different technologies were incorporated to implement this task: 3D geometry and texture acquisition technologies; 3D point-wise surface data acquisition technologies; relational database system technologies; Virtual Reality technologies (e.g. VRML); dynamic User Interface technologies realised through Java and PHP languages. Finally, borrowing ideas from GIS systems (where 3D information is structured in layers and the user can visualise any kind of information layered graphically over the base background layer), the effort was directed to the development of a system where a user can visu-



Figure 3. micro X-Ray Fluorescence (μ -XRF)



Elements	Value
SiO2 %	21
Al2O3 %	13.32
Fe2O3 %	23
Na2O %	2
K2O %	2
CaO %	1
FeO %	38
Li2O %	41
MgO %	4
MnO %	1.12231
Cu ppm	3111
Zn ppm	231
Co ppm	12313
O ppm	122
Ba ppm	2222
Si ppm	6783

Figure 4. Entering data

alise not only the physicality of an object, but also how it is described in the physico-chemical database (for example, the surface distribution of iron in the pigments of the decoration). This led to the name '3D ArchGIS' Cultural Database.

FEATURES OF 3D ARCHGIS

3D ArchGIS is an application based on the client-server architecture. The client component has been implemented as a plug-in for Microsoft Internet Explorer (see Figure 1). Web browsers are one of the most widely used and well-established platforms for presenting multimedia content while supporting numerous data formats, from simple text and images up to live streaming video and 3D graphics. Such a sophisticated client-side software component requires an equally powerful server that will be able to handle the complexity and size of the data. Modern multimedia databases can handle huge amounts of data in a very efficient manner. Thus, Web browsers are considered among the programming community as an ideal platform for the development of database oriented applications in situations where the main scope of the system is the global distribution of multimedia information. The current system of archaeological artefact archiving presents historical information in textual format accompanied by typological data alongside a realistic 3D representation of the artefact on the same Web page. A typical 3D scanning system (see Figure 2) is limited to export geome-

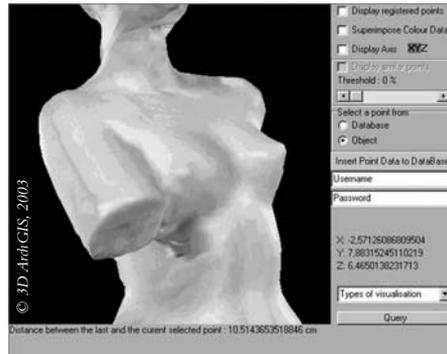


Figure 5. A new point selected

try and texture data (also known as 3 to 4 dimensions of data) in a Web compatible format like the Virtual Reality Modelling Language (VRML). However, the Institute's infrastructure makes possible the extraction of information from an artefact that goes beyond the typical 3 to 4 dimensions – 3D ArchGIS can be considered as an enhanced 3D model viewer with unique features. The system allows data mapping on the surface of the 3D model and at the same time composes a unique tool for browsing and retrieving these data. The information used for 3D mapping is grouped according to its physicochemical attributes, retrieved using the micro X-Ray Fluorescence (μ -XRF) technique against the object's surface to extract the chemical decomposition (see Figure 3).

The initial study involves the digital acquisition of geometry and texture from the actual artefact. Then it is parsed to the 3D ArchGIS to map the data retrieved from the μ -XRF system. The mapping of the data is performed manually at present (Figure 4). A region of interest is selected (Figure 5) on the surface of the 3D object that corresponds to the real object point where the measurements have been taken. The contribution of 3D ArchGIS to this task is that it enables study of the 3D model of the artefact and users can easily pick the 3D coordinates of the sampled point just by clicking on the surface of the 3D model (Figure 6).

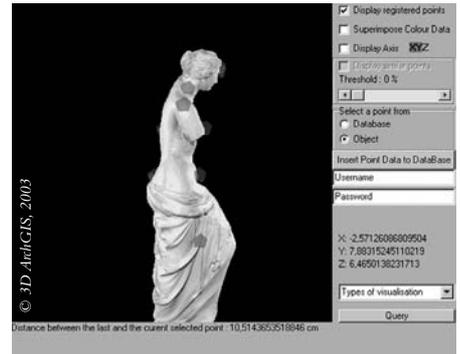


Figure 6. Registered points

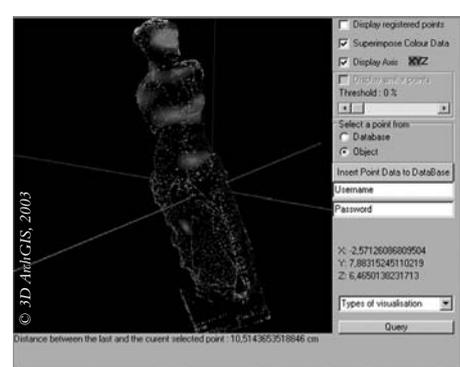
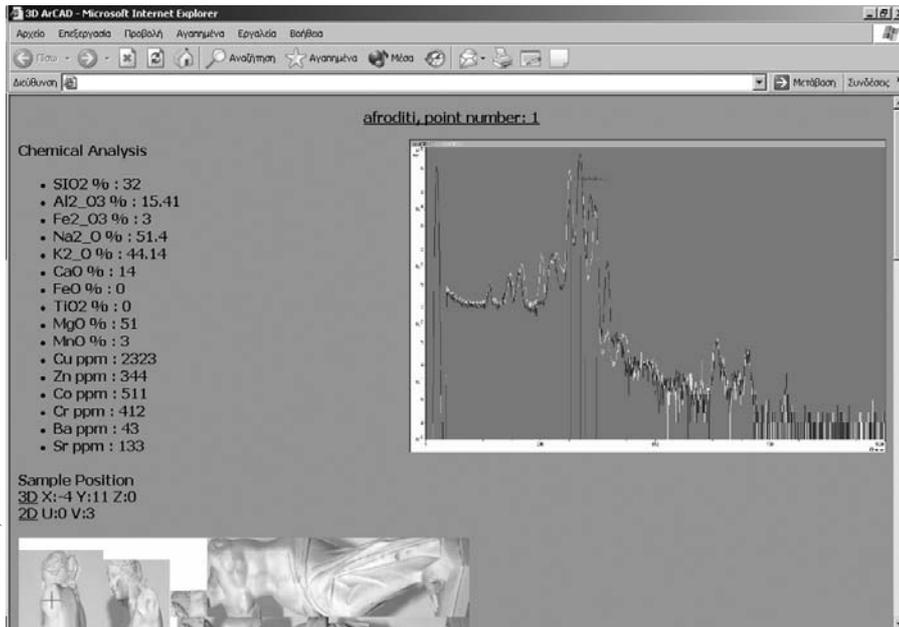


Figure 7. Colour mapping based on data

3D ArchGIS can be used for presenting all this mapped information or as a simple 3D viewer for displaying the 3D scanned artefacts on the Internet. The user can see the object from any angle, zoom in and out, and browse its surface for data. Furthermore, it is possible to visualise the object at spectrum bands invisible to the human eye, or even to observe the chemical elements distribution as mapped on the surface of the object (Figure 7). Furthermore, a point from a region of interest can be selected in order to grant access to the entire μ -XRF analysis for that point which includes chemical composition data followed by the respective spectrograms (Figure 8).

The system has the capability to block unauthorised users from inputting new data sets into the database; however, browsing and querying the database is a free access service with restrictions on the results. Querying is implemented through



© 3D ArchGIS, 2003

Figure 8. Data from a point

a friendly graphical user interface (see Figure 9). In this way, a laboratory which conducts an extensive study on an archaeological artefact can provide the entirety of the information regarding that artefact in a very valuable and efficient way to scientists all over the world.

TECHNOLOGY BEHIND 3D ARCHGIS

As mentioned earlier, the client component of 3D ArchGIS is implemented as a Web browser plug-in. The advantage of running the application as a plug-in within a single window is that it allows the user to work in a consistent and familiar environment without the need to switch between separate application windows. Furthermore, the versatile nature of a plug-in software component permits the application to be incorporated into bigger Web projects and allows real multimedia visualisation of the entire data set. This very useful feature is, however, also a restrictive factor that leads to some drawbacks. As a browser plug-in, 3D ArchGIS is restricted to Web oriented technology. The most widely used technologies for development in browser plug-ins are Java applets and ActiveX. Java applets are plat-

form independent but the fact that accelerated OpenGL graphics are not supported by default raises the issue of the manual download and installation of large files (Java2 & Java 3D) by inexperienced users.

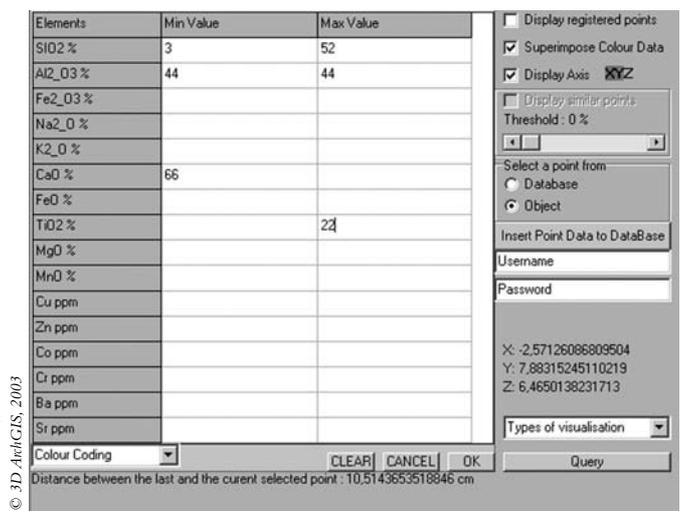
ActiveX plug-in is a well-established method for developing software components that can be executed within other applications (modularity). It is supported by a variety of software development platforms such as Microsoft Visual Basic, Microsoft Visual C++, and Delphi; however, the main disadvantage of this method is that it is operating system dependent (only available from Microsoft Windows version 95 and later) and browser dependent (Internet Explorer and Netscape Navigator).

During the initial requirements analysis and specifica-

tions, the target group of the application was identified as a group of 'average, semi-experienced computer user[s]'. Therefore most of these users would have Microsoft Windows OS driven computer systems with Internet Explorer (installed as default during the OS installation) and the 'in window' accelerated OpenGL graphics support. The majority of Microsoft Windows operating system users steered the development process to the ActiveX plug-in solution.

For the real time rendering of 3D graphics Open-GL libraries are used. Open-GL (<http://www.opengl.org/>) is a well-established industry standard and is supported by most operating systems and graphics hardware developers. Thus, a large portion of the code is platform independent and the rendering of 3D graphics is hardware accelerated via the drivers of a standard graphics card.

3D ArchGIS retrieves the 3D object to be explored from a remote server, along with information from a database where the latest measurements and scientific data for each artefact are regularly updated. The communication between the server and the client is bi-directional (see



© 3D ArchGIS, 2003

Figure 9. Easy querying GUI

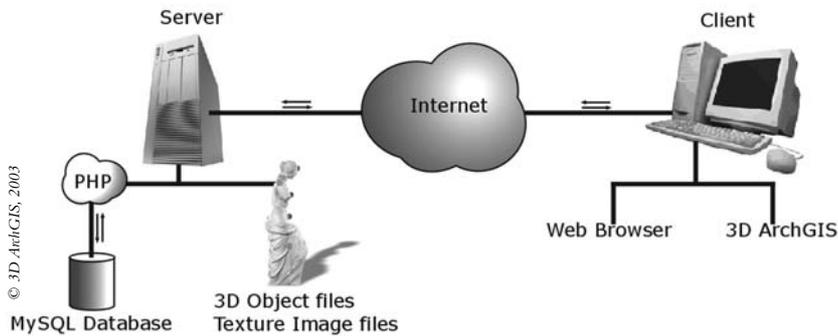


Figure 10. Communication Scheme

Figure 10). 3D ArchCAD communicates with the server by performing calls to PHP (a server based script language) pages and retrieving the output (simple text) from the server as a data set of input parameters. This technique is implemented by making use of operating system specific calls (which makes this part of the code platform dependent).

FUTURE DEVELOPMENT

3D ArchGIS is an experimental application and still under development. Currently, it is on the second *beta* implementation. Future implementations of 3D ArchGIS will address operating system dependency, which is one of the most serious limitations, and the needs of users with low bandwidth Internet connections.

Addressing bandwidth challenges will require innovative methods of data compression, including the 3D models, combined with progressive loading and display. At its current implementation the application is focused on mapping physicochemical data on the surface of the 3D models; future versions are designed to be more flexible on the type of mapped data and will allow mapping not only on the surface of the model but on its entire volume as well.

CONCLUSION

3D-digital archiving and presentation provides scientists and the general public with powerful means for the registration, documentation, preservation, display and dissemination of our cultural heritage.

Modern measurement instrumentation combined with multimedia databases and the appropriate software tools can offer a very rich and coherent description of a cultural object – a description that meets not only the needs of a virtual museum, but also the very demanding and detailed requirements of an in-depth, precise and accurate scientific study, even from remote locations.

ACKNOWLEDGEMENTS

The 3D ArchGIS project at CETI is a collaborative effort of specialists in various fields of science and technology, as is clearly shown through the professional expertise of the members of the developing team: Dr C. Chamzas (Electrical and Mechanical Engineer), Dr N. Tsirliganis (Nuclear Physicist), Dr D. Tsiafakis (Archaeologist), Dr Z. Loukou (Chemist), G. Pavlidis (Electrical and Computer Engineer), A. Tsompanopoulos (Electrical and Computer Engineer), K. Stavroglou (Electrical and Computer Engineer), D. Papadopoulou (Chemist), V. Evangelgelidis (Archaeologist), A. Koutsoudis (Multimedia Engineer), and F. Arnaoutoglou (Programmer). We believe this mixture of specialties reflects the demands of archiving of Cultural Objects in the twenty-first century. [BACK TO PAGE 1](#)

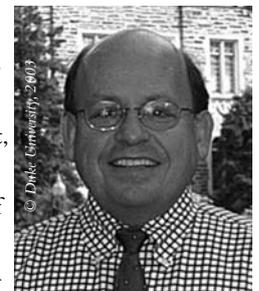
ZOPE AT DUKE UNIVERSITY: OPEN SOURCE CONTENT MANAGEMENT IN A HIGHER EDUCATION CONTEXT

PAUL CONWAY
DIRECTOR, INFORMATION TECHNOLOGY SERVICES, DUKE UNIVERSITY LIBRARIES

Duke University is about to embark on a major initiative to deploy an open source Web content management system (CMS) as an enterprise Web application. Two years in the planning, the university library has played a pivotal role

in defining the need, establishing the technological boundaries of the initiative, and marshalling the resources for the library to take advantage of Duke's overall investment in CMS software and services. This brief article explores the particular opportunities and challenges that are represented by the choice of Zope as the university's content management framework tool and of Zope Corporation as a provider of development services. The conclusion of

this article is that Web content management technology, if the fit is right, may be a particularly useful lever of innovation for libraries in a federated institutional environment.



Paul Conway

DUKE UNIVERSITY CONTEXT

Duke University (<http://www.duke.edu/>) is a private higher educational institution founded in 1924 in Durham, North Carolina. As with any modern research university, Duke is highly decentralised. Administrative departments are often proudly independent and have, over time, evolved subcultures and ways of accomplishment that may appear mysterious to the uninitiated. Decentralised administration extends to all aspects of the university, including technology operations and Web publication programs. The Web at Duke is a sprawling network that totals over 750,000 individual pages and hundreds of databases that generate Web pages dynamically. Over the past decade, virtually all HTML encoding has been hand-crafted. The university's news and communication department manages the Duke homepage, but the front-page system is a minuscule proportion of the whole. Responsibility for creating and managing virtually all pages beyond the front door is widely distributed across the campus. In the university library alone, over 70 individuals have authoring rights and responsibilities for portions of the 30,000 pages that reside on the library's Web server; some of these staff spend substantial portions of their week on content creation tasks. Duke is just now beginning to focus on the very real resource and content limitations of the university's Web space. A near consensus exists at Duke on the value of increased support for building and maintaining Websites and increased capability to share content internally, reduce inefficiencies across the campus, and improve the overall quality of our technology face to the outside world.

Duke conducted a review of the state of the university Web in October 2001. The internal report articulated a

vision of a unified Web presence supported by 'a sound Web strategy and an environment that automates some of the collaborative contribution'. The report identified four overarching benefits to a more unified approach to the Duke Web, including more effective branding, customised content, department autonomy with purpose, and improved quality of the overall effort. One of the critical mechanisms for achieving these benefits, according to the report, is using software to support Web development and content sharing – precisely the claims of the content management system industry.¹

WHAT IS A CONTENT MANAGEMENT SYSTEM?

In the United States, the term 'content management system' or 'Web content management system' increasingly has a distinctive meaning. For European readers, a CMS might best be thought of as a subset of the larger concept of 'Digital Asset Management System' or DAMS. The *DigiCULT Technology Watch Report 1* states that 'DAMS employ technologies such as commercially available database management tools to handle and manage resources, allowing users to discover them with ease and speed and owners/creators to monitor their usage and version histories.'² Content management systems are indeed database driven tools, but the focus is on publication processes rather than on search and discovery.

University of Washington professor Bob Boiko begins with a high-level view of content and its effective management. 'Content, stated as simply as possible, is information put to use. Information is put to use when it is packaged and presented (published) for a specific purpose. More often than not, content is not a single "piece" of information, but a conglom-



Perkins Library at Duke University

eration of pieces of information put together to form a cohesive whole. A book has content, which is comprised of multiple chapters, paragraphs, and sentences. Newspapers contain content: articles, advertisements, indexes, and pictures. The newest entry to the media world, the Web, is just the same; sites are made of articles, advertisements, indexes, and pictures all organised into a coherent presentation.³

A content management system is the technical environment (hardware, software, expertise) that supports the systematic processing of digital content from authorship to publication. For our purposes, 'publication' is delivery to users via Internet browser technology, but Boiko makes it clear that, once digital content resides in a CMS, it can be repurposed for any number of uses, including feeding to digital printing presses or reformatting for transmission as 'fixed' digital entities, such as PDF documents. 'At first blush CMS may seem like a way to create large Websites, but upon closer examination it is in fact an overall process for collecting, managing and publishing content to any outlet.'⁴

In 'Content Management Systems: Who Needs Them?'⁵ the authors acknowledge that the boundaries are fuzzy between document management systems, knowledge management systems, groupware and other enterprise information management systems. They place the core functions of a content management system into four categories:

1 Duke University, *State of the Web*, 17 October 2001. <http://www.oit.duke.edu/CMSsub/docs/StateofTheWeb.pdf>

2 Seamus Ross, Martin Donnelly and Milena Dobrova, *DigiCULT Technology Watch Report 1: New Technologies for the Cultural and Scientific Heritage Sector*, February 2003, p. 42. <http://www.digicult.info/pages/publications.php>

3 Bob Boiko, 'Understanding Content Management' in *ASIS Bulletin*, 28 (Oct/Nov 2001). <http://www.asis.org/Bulletin/Oct-01/boiko.html>

4 Bob Boiko, *Content Management Bible* (New York: Wiley & Sons, 2001).

5 Paul Browning and Mike Lowndes, 'Content Management Systems: Who Needs Them?' in *Ariadne*, 30, 20 December 2001. <http://www.ariadne.ac.uk/issue30/techwatch/intro.html>



1. **authoring** (creating Web content in a managed and authorised environment);
2. **workflow** (management of the steps between authoring and publishing);
3. **storage** (authored content components, in multiple versions, in a repository);
4. **publishing** (dynamic delivery of stored content to the Web).

DUKE GOALS FOR A CONTENT MANAGEMENT SYSTEM

At Duke, as may be the case in any widely decentralised organisation, the capabilities of an enterprise software tool with the wide impact of a content management system are demanding. The team charged with identifying a CMS tool for the university developed a list of key goals driving the software selection process.

- **Site Maintenance.** Enable non-technical staff to update and maintain their Web content more easily and efficiently using a variety of computing platforms and Web development tools.
- **Consistency.** Provide a consistent Web site management environment that will handle content creation, style, visitor usability, policy, workflow, versioning, and revision control for decentralised Web infrastructures and content authors.
- **Sharing.** Facilitate the syndication (sharing and reuse) of Web content, with the appropriate editorial accountability, by offering central facilities to index and cache content that originates anywhere on the Duke subnet.
- **Automation.** Reduce inaccurate, outdated, redundant or unauthorised content through automated content management processes, versioning and workflow.
- **Accessibility.** Make it easier for all Duke Web efforts to achieve and maintain compliance with evolving standards (ADA and W3 compliance, University and departmental privacy policies, etc.) and provide the tools (repositories of approved templates, images and logos,

etc.) needed to support efforts to balance creativity and flexibility with brand consistency at the departmental, school and University levels.

- **Flexibility.** Make it easier for all Duke Web efforts to support the full range of Web content sources (development tools, databases, etc.) and Web display devices (graphical browsers, text-only browsers, kiosks and emerging small-screen platforms such as cell phones and PDAs).
- **Implementation.** Give first priority to the immediate needs of early adopters and second priority to anticipated future needs of an emerging CMS community.

FUNCTIONAL REQUIREMENTS

The overall goals for the campus content management initiative defined a discrete set of functional capabilities that must be included in any system, whether bought or built locally.

- **Content Authoring.** Provide for the smooth and efficient creation and editing of content; integration with common editing and word processing tools; ability to import existing content, metadata, multiple file types.
- **Presentation.** Flexible use of common development tools; creation or import of templates, standard forms.
- **Syndication.** Support the sharing and reuse of content across campus.
- **Workflow.** Support various forms of workflow and facilitate the management of content throughout its lifecycle (from creation through publication and syndication to deactivation and archiving).
- **Versioning.** Provide revision control mechanisms, both in support of collaborative content manipulation and in support of change management and content reversion.
- **Accessibility.** Content stored or indexed within content repositories must be easily accessible during Website development, and access needs to be provided

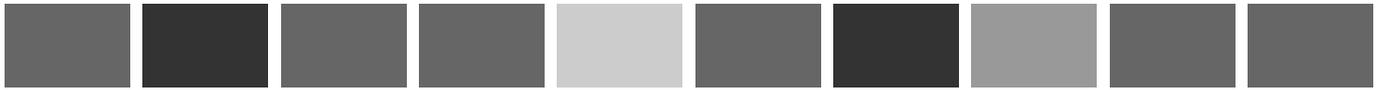
to a wide range of content objects both within and outside traditional CMS content repositories.

- **Delivery/Publication.** Provide content to participating delivery servers and to facilitate the publication and management of delivered content.
- **Site Management.** Provide mechanisms for monitoring and reporting on the use of managed content; provide mechanisms for automatically generating high-level 'site maps' depicting the relationships of pages and objects to a depth of at least four layers of referral.

TECHNICAL REQUIREMENTS

Together with the distributed institutional structures of the university, baseline functional requirements helped determine the technical parameters of a content management system.

- **Federated Architecture.** Multiple independent servers running a standard, accepted CMS software package put together in coordination with a central CMS support organisation.
- **Interoperability.** CMS systems must interoperate as seamlessly as possible with a wealth of existing content already available through a number of well-travelled and well-recognised Duke campus Websites. System-driven mandates for change should be at the discretion of the content creators, where possible.
- **Integration.** Because the level of technical expertise varies dramatically among campus users, flexible integration with third-party development and authoring tools is critical. Some users will have a high degree of technical expertise and demand fine-grained control over their content; some will have a low degree of technical expertise and need to create and deliver content in the most natural and 'matter of fact' fashion possible.
- **Scalability.** Because neither the size of the user population nor the extent of



the eventual growth of a CMS on campus is predictable, a CMS must be easily scalable and expandable.

- **Openness.** The University has a strong interest in ensuring that solutions deployed at the enterprise level support open standards, both to protect against vendor lock-in and to ensure the long-term utility of campus software implementations. Of particular interest to the University are open-source solutions that feature the active involvement of developer and user communities and vendors that demonstrate a willingness to train local users and participate in open standards development.
- **Modularity.** The major components of a campus implementation must be capable of functioning with relative independence and have modular components that can be adopted in varying combinations by various campus units.
- **Enterprise.** Enterprise-level strategies for selecting vendors, platforms and support mechanisms will be used to specify a CMS product.

Based upon the requirements described above, a campus-wide committee managed a vendor selection process that

began with a thorough review of the existing technical literature and a set of consultations with industry analysts. The committee assessed products from fourteen companies and then issued a formal Request for Proposals (RFP) to four vendors. Representatives of these companies attended a half-day bidders conference to clarify the requirements and then submitted full responses to the RFP. A comprehensive review of the four proposals led the committee to recommend the selection of the Zope Corporation (<http://www.zope.com>) to build a Duke-specific content management system based upon the open-source Zope content management framework (<http://www.zope.org>).

Duke University signed a contract with the Zope Corporation to develop a product called 'Zope4Edu' with Duke as the first user. 'None of the leading CMS packages seemed quite right for our situation,' said Tracy Futhey, Duke's vice president for information technology and chief information officer. 'Our challenge was political and institutional as much as technical. We needed a system that was very flexible, with open-source code, that

would encourage people to share resources and work together. Zope was responsive to our needs and we're now working with them to create a CMS solution that we hope will transform Duke's online activities.'⁶

ZOPE, OPEN SOURCE AND THE CMS COMPETITION

Web content management tools are big business in the US marketplace today. Industry analysts have content management trends on their radar screens and have produced assessment reports over the past three years that document increasing hype, marketplace mergers and consolidation, and the relatively rapid maturing of the product environment. Three of the largest technology analysts have articulated the case that an open source content management system **might** be a viable option, at least until the commercial sector stabilises. Among these analysts, however, only Tony Byrnes' independent *CMS Watch* has given substantial attention to any of the open-source CMS tools. In some ways, the failure of open-source CMS to register a share of marketplace or to compete directly with large-scale enterprise applications is understandable. How do you identify 'market share' when the only product being sold is a service? How do you measure return on investment when the core software product is freely available for download?

Freely available and extendable software has been around in many forms virtually since the inception of computing technology. In particular, the educational and research emphasis of the early evolution of Unix has fostered a robust community of developers and users that has freely shared and cooperated on open-source software projects for many years. The more formal concept of 'free' software as an organised and *licensed* initiative is credited to Richard Stallman, who founded the Free Software Foundation in 1984 and introduced the GNU project. Today,

INDUSTRY ANALYSTS

CMS Watch. <http://www.cmswatch.com/> 'CMSWatch is an independent source of information, news, opinion and analysis about Web content management.'

CMS Wire. <http://www.cmswire.com/> 'Content management news, commentary and product information.'

Butler Group. <http://www.butlergroup.com/> 'Europe's Leading IT Analyst Company providing Analysis without compromise.'

Faulkner Information Services. <http://www.faulkner.com> 'Faulkner provides in-depth technology information services to public and private sector organisations worldwide. Our products include subscription and custom-developed reports, studies and databases.'

Gartner, Inc. <http://www4.gartner.com/Init> 'Gartner, Inc. is a research and advisory firm that helps more than 10,000 clients leverage technology to achieve business success.'

⁶ Zope Corporation press release, 9 Oct 2003.
<http://www.zope.com/News/PressReleases/DukePR>



OPEN SOURCE CONTENT MANAGEMENT TOOLS

Enhydra Project. <http://enhydra.enhydra.org/index.html> 'Enhydra is the first and leading Open Source Java/XML application server. It was initially created by Lutris Technologies, Inc. Over four years in development it was open sourced on January 15th 1999. They christened the technology "Enhydra" after the California sea otter (Enhydra Lutris) - a popular inhabitant of Santa Cruz. After Lutris switched to closed source development and finally failed on the market, the community took over the further development.'

Red Hat Enterprise Content Management System.

<http://www.redhat.com/software/rha/cms/> 'The Red Hat Content Management Solution (CMS) combines powerful Web content management functionality with the flexibility to tailor deployments to the specific production environments and processes of each organisation. As the needs of the organisation grow, additional functionality can be integrated through other Red Hat Applications packages, including Red Hat Portal Server and collaborative plugins.'

Zope. <http://www.zope.org> 'Zope is a framework for building a special kind of Web application. Generally, a Web application is a computer program accessed with a Web browser; the special kind of Web applications built with the framework provided by Zope can also be thought of as dynamic Websites that both provide static information to users and allow those users to work with dynamic tools.'

Competition in the open-source arena is narrow but intense. Zope is one of three enterprise-scale open-source content management systems. In the United States, Red Hat markets a comprehensive Web content management product suite that is derived from the *Ars Digita* tool kit launched in the 1990s by Philip Greenspun and his associates. The Enhydra system started life as proprietary software and then emerged as open-source software when the parent company failed.

Zope is multidimensional in its conception and challenging in its implementation.⁹ On one dimension, Zope is an open-source Web application platform written in the Python programming language. People who know and use Python praise its elegance; most developers are unfamiliar with the language. The open-source version of Zope, now in release 2.7.0, is fully and enthusiastically supported by a worldwide development community that coordinates its activities through Zope.org.¹⁰ Zope is also a corporate entity that markets configuration services for the Zope platform and specialised products built on top of the Zope content management framework. Some of the major developers of Zope and the content management framework are employed by the Zope Corporation.

On yet another dimension, Zope is a content management framework that provides the tools and pre-programmed elements required to create a content management system. Plone

GNU-licensed 'free software' is routinely leveraged by thousands of developers on dozens of hardware and open-source platforms. At its heart, the open-source model is based on access to the source code of an application and the rights of users to freely modify the software for their own use or redistribution as they see fit.⁷

The success of any given open-source product is as much related to the existence of an active community of developers willing to share their work as to the fact that the source code is free. Open source pioneer Eric Raymond notes that the people who constitute a community systematically harness 'open development and decentralised peer review to lower costs and improve software quality.' More important, he clarifies that the development community manages development directly through a trusted-peer network. 'The owner(s) of a software project are those who have the exclusive right, recog-

nised by the community at large, to *re-distribute modified versions*' of the software.⁸

Complicating the open-source environment is the existence of private companies whose livelihood involves selling services to other companies that are attracted to a given open-source product but need help installing, configuring or customising the software. For example, in the United States, Red Hat is the premier commercial marketer of the Linux open-source operating system (<http://www.redhat.com/>).

ACCESS TO ZOPE

- Zope, the community.** <http://www.zope.org>
- Zope, the company.** <http://www.zope.com>
- Zope, the developers.** <http://cmf.zope.org/>
- Zope, the European connection.** <http://www.eurozope.org/>
- Zope, the lab.** <http://www.zopelabs.com/>
- Zope, the magazine.** <http://www.zopemag.com/Issue001/index.html>
- Zope, the orientation.** <http://www.zopenewbies.net/index.html>
- Zope, the Zen.** <http://www.zopezen.org>

7 For more information on open-source software, see 'Open Source and Free Software Solutions: Reliable, Standardised, Free' in *DigiCULT.Info*, Issue 5, October 2003. (http://www.digicult.info/downloads/digicultinfo_issue5_october_2003.pdf)

8 Eric S. Raymond, *The Cathedral & the Bazaar; Musings on Linux and Open Source by an Accidental Revolutionary* (Sebastopol, CA: O'Reilly & Associates, 1999), p. 89.

9 Paul Browning, 'Zope - A Swiss Army Knife for the Web?' in *Ariadne*, 25, 24 September 2000. <http://www.ariadne.ac.uk/issue25/zope/intro.html>

10 Chris McDonough, 'Gain Zope Enlightenment By Grokking Object Orientation', 12 October 2000. <http://www.zope.org/Members/mcdonc/HowTos/gainenlightenment>



(<http://www.plone.org>) is an example of a content system that is built using the Zope content management framework and distributed on an open-source licence. Plone could serve as an intranet and extranet server, as a document publishing system, a portal server and as a groupware tool for collaboration between separately located entities. The 'CMF Dogbowl' Website (<http://cmf.zope.org>) is the central information point for developers of, and for, the CMF. The site is additionally intended to be used as a 'fishbowl' site for the CMF, in the sense that Zope Corporation is committed to develop Zope 'in the fishbowl', with all significant product development visible and accessible to interested members of the Zope community.

THE LIBRARY VISION FOR CONTENT MANAGEMENT

Duke University Library has been an active participant and an enthusiastic supporter of the University's pursuit of an enterprise content management solution for the campus. A viable Web content management system, supported fully by the university, is a key technology tool for the library. In the emerging plan for digital library infrastructure, software that supports the systematic management of Web space is one of three critical components (see Figure 1). A CMS will be a more cost-effective way to manage the library's Web gateway and a more effective way to share digital content across campus.

- A CMS will empower library staff to produce more and better content with the same or fewer human resources devoted to the task.
- The version control features of a robust CMS may provide a viable alternative to 'archiving the Web' by managing the underlying content of the dynamic Web rather than managing fixed HTML representations.
- By separating the content of the Web from its presentation via a browser, a CMS will allow the libraries to focus

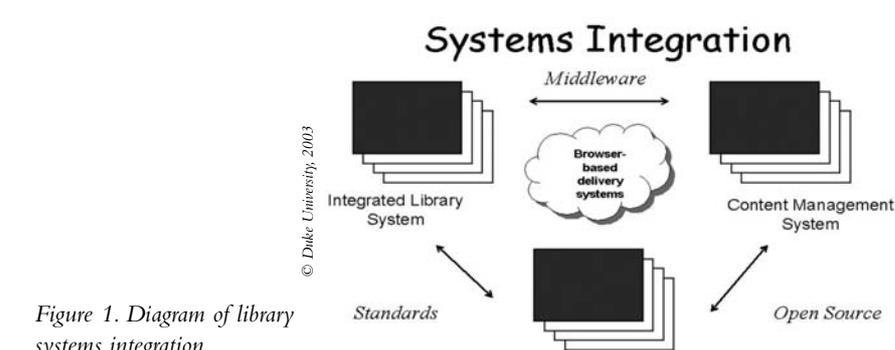


Figure 1. Diagram of library systems integration

responsibility for particular tasks of Web publication in areas of expertise that the library already has in place – in much the same way as a magazine publisher is organised into specialised departments.

- For the libraries at Duke, a campus-wide CMS may dramatically improve its ability to distribute the value-added products of its staff, including resource guides, image and text resources created locally, and research guides and other tools.
- Perhaps most important, a CMS, when fully operational, will give the library an opportunity to present its Web content in ways that respond directly to the information needs of its end-users and give users the ability to determine their perspective of the library. Previously static Web content can be presented dynamically to portray the library as a landscape of digital content, as a service organisation, or as a suite of tools to support resource discovery and use – to name just three of the possibilities.

CONCLUSION

The role of the library in an institution of higher education is changing. Fuelled in part by technologies such as content management systems, which permit more interesting and dynamic ways of disseminating library content, university libraries have new opportunities for leadership and influence on campus. Wendy Lougee, University Librarian at the University of Minnesota, argues that libraries must use their technological

resources to engage scholarship directly. 'We see the library becoming more deeply engaged in the fundamental mission of the academic institution – i.e. the creation and dissemination of knowledge – in ways that represent the library's contributions more broadly and that intertwine the library with the other stakeholders in these activities. The library becomes a collaborator within the academy, yet retains its distinct identity.'¹¹ The content management initiative at Duke has laid the groundwork for the library to lead from one of its principal strengths, namely its innate mastery of metadata and the fundamental role such data plays in the management of information content, including dynamic Web content.

A substantial investment in Web content management technologies remains a risk at this point in time. Not only is the CMS market changing almost too rapidly to gauge, but the cultural conditions in higher education that are required to embrace a dynamic and open Web space are undeveloped and untested. In partnering with a purveyor of services to the open-source community, Duke is gambling as to whether it can form a successful development community within the campus environment and change the culture of Web content creation from the inside out.

Information on the Duke University Libraries' implementation of the Zope content management system is available from <http://www.lib.duke.edu/its/index.htm>.

BACK TO PAGE 1

¹¹ Wendy Pratt Lougee, *Diffuse Libraries: Emergent Roles for the Research Library in the Digital Age* (Washington, DC: Council on Library and Information Resources, August 2002), p. 4. <http://www.clir.org/pubs/abstract/pub108abst.html>

NEW OPEN-SOURCE SOFTWARE FOR CONTENT AND REPOSITORY MANAGEMENT

FEDORA

The Fedora Project (<http://www.fedora.info/>), a joint venture between Cornell University and the University of Virginia, has released an open-source digital object repository management system. Based on the Flexible Extensible Digital Object Repository Architecture (FEDORA), the system can support a repository containing one million objects, managed via a graphical user interface. The management and access APIs are implemented as SOAP-enabled Web services and among other features include an OAI Metadata Harvesting Provider.

Fedora digital objects conform to an extension of the Metadata Encoding and Transmission Standard (METS, <http://www.loc.gov/standards/mets/>).

Objects can be submitted to the repository in XML and are persistently stored in the repository as XML files. The Fedora extension of the METS schema can be viewed at <http://www.fedora.info/definitions/1/0/mets-fedora-ext.xsd>.

The system is currently being tested by some North American institutions. For more information and to download the Fedora software, visit <http://www.fedora.info/>.

PLEADE AND NAVIMAGES

Two new open-source tools were unveiled on 7 October 2003 by the Centre historique des Archives nationales (CHAN, <http://www.archivesnationales.culture.gouv.fr/chan/>). The first, Pleade (<http://www.pleade.org/>), was developed

by French companies AJLSM and Anaphore as a generic platform for searching and browsing XML/EAD archival finding aids. A detailed report on the functionality and development of Pleade follows.

The second tool, Navimages (<http://sdx.archivesdefrance.culture.gouv.fr/gpl/navimages/index.html>), enables users to manage, view, search and browse a series of digital images. It is composed of three modules (the image collector, document base, and image viewer) and was designed to be closely compatible with Pleade.

The new software was presented to cultural heritage and information professionals and will become well recognised as an aid to collection management.

BACK TO PAGE 1

PLEADE – EAD FOR THE WEB

MARTIN SÉVIGNY, AJLSM, & FLORENCE CLAUAUD, CENTRE HISTORIQUE DES ARCHIVES NATIONALES

On 7 October in Paris, the PLEADE project was presented during a special meeting open to all cultural heritage professionals in France and elsewhere and to their private partners. PLEADE will help institutions put archival finding aids encoded in EAD on the Web, by providing a set of tools to build dynamic Websites.

EAD – Encoded Archival Description¹² – is technically an XML data format for archival finding aids. However, for

archivists and archival institutions, these three letters have a lot more meaning and EAD can easily be qualified as a real enabling technology. One of the first reasons is that EAD is an international and generic data model, maintained by a large working group where archival professionals and IT specialists meet, and where user needs drive the evolution process. Moreover, this data model is now completely compatible with the ISAD(G) – General International Standard Archival Description – conceptual model, published in 1993 by the International Council on Archives and used worldwide.¹³ The two standards are based on the principle of multilevel archival description, leading to hierarchical finding aids. These characteris-

tics have quickly put EAD in front, and this model is used today by a wide variety of institutions with archival holdings.

PLEADE is more than a software component; it is a project and, more importantly, an open-source project. It means that the software itself is distributed with an **open-source** licence (namely the GPL) and the sources (including documentation and examples) are available online.

PLEADE was initially developed by, and is supported by, two private companies:

- **Anaphore** (Barbantane, France, <http://www.arkheia.net>), the leading

¹² For information about the EAD, please visit the Official EAD Website at the Library of Congress: <http://www.loc.gov/ead/>.

¹³ The standard is available online: <http://www.ica.org/biblio.php?pdocid=1>.



French provider of software for archival institutions. Their flagship product, Arkhéia, provides features for managing archival repositories, but also includes a module for building finding aids which can then be exported in EAD format.

- **AJLSM** (Bordeaux, France, <http://www.ajlsm.com>), an important provider of services and tools for managing documentary information using XML and normalised solutions. AJLSM specialises in cultural heritage information systems, and is the main developer of the SDX platform, the underlying technology for PLEADE.

During the past few months, two other institutions have contributed greatly to the development of the software by testing and debugging it, and by providing valuable comments and feature requests:

- **The Centre historique des Archives nationales** (Paris, France, <http://www.archivesnationales.culture.gouv.fr/chan/>)
- **The Archives départementales de Maine-et-Loire** (Angers, France, <http://www.cg49.fr/themes/archives-departementales/>)

Currently, PLEADE is available as a second release candidate of version 1.0. This version has been thoroughly tested, and can be downloaded from the development Website. A final release of version 1.0 should be made available in December 2003.

MAIN CHARACTERISTICS

PLEADE is a highly configurable and multilingual Web publication framework, including a search engine, for EAD documents. There are a few important concepts in this short description that we should emphasise:

- **EAD documents:** As PLEADE is built for EAD documents, everything is optimised for the principal characteristics of

the EAD structure, such as hierarchical browsing and discovery of information.

- **Publishing framework:** EAD documents can be published on the Web, enabling users to browse and then read the finding aids – and their accompanying digital artefacts – online.
- **Search engine:** PLEADE is suitable for publishing large collections of (small or large) finding aids, thanks to the availability of a powerful search engine, letting users find the most relevant archival components directly.
- **Highly configurable:** PLEADE is designed as a generic application and, although it is usable straight ‘out of the box’, all kinds of configuration and local adaptation are possible.
- **Multilingual:** By using Unicode-based technologies (XML, Java), PLEADE is ready for multilingualism, both in the user interface (French and English are currently available) and for full-text searching.
- **Web:** PLEADE uses a Web architecture (such as the Web itself, an intranet, even a single computer or removable media such as CD-ROM) for publishing the finding aids. This basically means that PLEADE itself is installed on a Web server, and the users connect to the application using a standard Web browser.

FUNCTIONAL OVERVIEW

An institution may use PLEADE for its specific characteristics such as searching and displaying EAD documents only, or it can build its own Website with the platform. In order to do this, PLEADE provides four general features: publishing static information, along with organising, searching and displaying finding aids.

Once installed, PLEADE may be seen as a traditional Website, as it includes facilities for **publishing static information** and organising this information in a configurable menu. This static information may be a home page, opening hours for

the public, general conditions for reading the archives, general information about the finding aids, etc. The pages containing the static information are in XHTML format, and content is displayed according to the general layout or ‘skin’ chosen for the site.

PLEADE also has an interesting feature for **organising finding aids** in subsets and groups of subsets. Let us suppose that an institution wants to organise the available finding aid collections such as those related to private companies, persons and public institutions. These subsets can be defined and grouped in PLEADE, and documents can be added to one or more subsets. Once accomplished, administrators can easily configure in the interface a list of finding aids per subset and provide specific search forms for documents within a subset. Since no subset is hard coded in PLEADE, administrators can define their own subsets.

Searching EAD documents in PLEADE is possible in various ways. In the proposed interface, a simple search bar is available in all pages, letting users make simple full-text searches across all finding aids published by PLEADE. Although some users may find this feature sufficient for their needs, administrators will often want to provide more advanced search forms. For such needs, PLEADE includes a search form generator, dynamically building and displaying HTML search forms from easy to create XML configuration files. With the examples included in the PLEADE distributions, building specialised search forms should not be a problem for anyone. All relevant EAD documents are available for searching.

With any search engine, one must define what the searchable units of information are. In PLEADE, these units must be determined at the time of indexing (and not dynamically at query time) and, to do this, the administrator is provid-



ed with various options such as choosing one or more archival levels (e.g. fonds, file, item), one or more hierarchical levels (e.g. any third-level component) or even an EAD element (c0, c1, etc.). This means that the user will see meaningful search results, at the file level for instance, but these results will always be contextualised (the titles of its parents are presented also).

Displaying EAD documents in PLEADE usually involves a three-pane window. The top pane presents the bibliographic information about the finding aid; the right-hand pane contains the contents, split into searchable units in order to avoid large pages on slow network connections; and the left-hand pane contains interactive tables of contents for browsing the finding aid by the component titles and, if available, by indexation terms found in the documents. Browsing and reading the EAD documents is possible with a standard Web browser, like any other feature of PLEADE, as the documents are dynamically converted from XML to HTML.

The above description of the main features of PLEADE should not conceal the fact that everything is highly configurable, from the interface 'look and feel' to the way the documents are indexed. For instance, when an administrator adds a document (or a batch of documents sharing similar characteristics), he or she may choose the subsets it belongs to, how the document will be fragmented for searching and displaying, how to build the title of the searchable units from the various EAD elements, whether a table of contents should be used and, if so, what its label will be, whether the indexing terms should be used for browsing and/or global searching, what presentation formats to use from among a list of available formats (two are currently available), the set of phrases for labelling the EAD element contents for display, how to process hypertext links

between one EAD document and another, or between one EAD document and the digital surrogates of the archival material, and so on.

TECHNOLOGICAL OVERVIEW

Following the tradition of open-source software and modern software development methodologies, PLEADE is a thin layer built on top of some other technologies. The most important ones are, from the bottom up:

- a Web server, such as Apache, for handling HTTP requests from a Web browser, and sending back the information as HTML pages;
- a Java servlet engine, such as Apache Tomcat, and beneath it a Java virtual machine;
- and SDX (<http://sdx.culture.fr>), a search engine and publishing platform for XML documents, based on the Cocoon 2 infrastructure.

PLEADE can thus be seen as a layer on top of SDX, a layer that lets anyone publish EAD documents on the Web. This layer takes control of tasks such as indexing, organising and displaying EAD documents. This layer is quite generic and can be easily configured to suit different needs or layouts. And since SDX supports the use of thesauri for searching, along OAI-PMH harvesting or repositories, these features could be easily added to PLEADE.

As PLEADE is based on Java technology, it can be installed on any platform supporting a Java virtual machine, including Linux and most of UNIX, Windows and Mac OS X. It is thus easy to install it on a personal computer for testing or presentations, and also on high-end servers for public access.

THE PLEADE PROJECT

Developing successful open-source software involves a lot more than technologies. For PLEADE, the main

objective is to build a virtual community of sponsors, developers, contributors, integrators and users, each playing an important role in the evolution of the project. One striking aspect of this collaboration is the amount of contributions needed from non-technical people.

Currently, PLEADE uses its own Website and the Sourceforge platform for open-source projects as the main tool for building this virtual community.

- At <http://www.pleade.org>, one will find general information about the project, its actors, various events relating to the project, and up-to-date documentation. Most of the information is currently in French, but translation into English and other languages should occur early in 2004 (contributions welcome).
- At <http://sourceforge.net/projects/pleade/>, one will find links for downloading the latest distributions of the software, and a mailing list (and its archives) for the PLEADE community. The Sourceforge services are available in various languages including English, and the mailing list accepts posts in any language.
- At <http://ajlsm-sdx.hopto.org/sdx-21h/pl-demo/>, there is a standard installation of PLEADE, including the example documents distributed with the platform. No specific configuration or adaptation has been made for this installation.

If anyone requires to know more about PLEADE or wishes to contribute, the first step is to subscribe to the mailing list and then send a message. All contributions and questions are welcome.

[BACK TO PAGE 1](#)

THE IRCAM DIGITAL SOUND ARCHIVE IN CONTEXT

BERNARD STIEGLER,
MICHAEL FINGERHUT,
NICOLAS DONIN
INSTITUT DE RECHERCHE ET DE
COORDINATION ACOUSTIQUE/MUSIQUE¹⁴

Domestic and institutional archiving and dissemination of texts, images and sound (as well as 3D representations, haptic sequences and so on) have been deeply affected by the advent of digitisation, which has opened up vistas unforeseeable a few decades ago, mainly due to its potential of quasi-infinite perfect reproduction (sometimes called 'hyper-reproducibility'). Digitisation has indeed induced a profound breach in the technology of memory from the way in which it had emerged in the industrial age, in particular regarding music.

At the turn of the nineteenth century, the development of analogue reproduction – in particular the phonograph – allowed for the establishment and development of a powerful record industry which made possible widespread access to music of all kinds (including music directly derived from this technology, such as jazz, or reinvented by the record industry, such as baroque music). The separation between artistic production and cultural consumption has had very positive effects: a large public – at least in the industrialised world – has acquired a vast (if shallow in places) knowledge of music history and its cultural diversity.

However, in turn, reproducibility has also caused a huge regression in the musical audition capacity of this public: the

record buyer-consumer does not need to know anything about music in order to listen to it, nor share with the music producer-composer any instrumental or formal, practical or theoretical knowledge. He cannot read or play music any more; his ears have no eyes and hands, as it were. He is nescient.

Digitisation has brought about a major change in the fundamentals of the whole process, and therefore constitutes a vast opportunity for change, in particular with respect to music, and even more so learned music, and this concerns us directly at IRCAM¹⁶ (<http://www.ircam.fr>).

IRCAM IN THE DIGITAL AGE

This thesis, put forth by Glenn Gould as early as 1966, has led us to redefine at IRCAM the policy of production, use and broadcast of digitised documents, and in particular of sound documents. It doubtlessly affects the process of music creation; the issue of digitisation is therefore not just a question of patrimonial archival and collection production or of access modes, but a reinvention of the whole field of *lutherie* (or instrument-making).

IRCAM is thus in transition from the age of computer music to that of music creation in the era of generalised digitisation, which will affect, among other things, home HiFi systems, media for offline dis-

*'As limited as it is, the manipulation of dials and buttons is an act of interpretation. Forty years ago, a listener could only turn his turntable on or off, and maybe, if it was an advanced model, adjust the volume somewhat. The variety of controls available nowadays requires a capacity for analytical judgment. Yet these controls are still very primitive in comparison with the means of participation that will be offered to the listener when the current very sophisticated techniques still in the labs are integrated into domestic appliances'.*¹⁵

tribution as well as online access, and, in due course, the modes of listening, both professional and amateur. We are therefore working on tools for computer-assisted music composition as instruments of description, indexation and representation of musical sounds, including means of genetic traceability of the works, among other things. We are also developing instruments for musicological analysis, as well as an authoring environment for producing personalised formalised and transmissible 'signed listenings' of a musical piece (by a musicologist, teacher, conductor or musician), to be used as listening guides or a graphical track on a musical DVD to be used in the HiFi appliances of tomorrow. In addition, we are putting together a Web radio channel, which will allow access to these analyses as well as to music as it is being created or the *avant-première* production.

It is within this new context that the IRCAM multimedia library has evolved into a laboratory and become part of a wider project about new *lutherie*, at the heart of which is the digital document.

¹⁴ Bernard Stiegler is a philosopher and the director of IRCAM. Michael Fingerhut, mathematician and computer scientist, is the director of the IRCAM Multimedia Library and its Engineering Bureau. Nicolas Donin, a musicologist, is head of the Formulation du Musical research team.

¹⁵ Glenn Gould, 'Idées' (1966) in Bruno Monsaingeon *Le dernier puritain* (Fayard, 1983), p. 88. Translated from the French by the authors of this paper.

¹⁶ *Institut de recherche et de coordination acoustique/musique*, founded in 1976, is a not-for-profit organisation and an associate department of the Centre Pompidou (the French national centre for arts and culture).



Figure 1. IRCAM Audio Player

THE EMERGENCE OF THE MULTIMEDIA LIBRARY AT IRCAM AND ITS DIGITAL COLLECTIONS

From its inception, IRCAM has been engaged in scientific research into, and technological development for, all the phases of music creation and production, from the elaboration and exploration of ideas by the composer, through its actualisation in a paper and/or electronic score, its live performance on stage by instrumentalists and computers, and its off-stage life-cycle. The R&D department is working in close collaboration on research and artistic projects with composers trained in its Pedagogical department, with instrumentalists and with other artists. The public performance of their resulting compositions is then handled by the Production department.

This abundance of projects and activities generates a wealth of intellectual production, in the guise of music scores and recordings of performances on the one hand, and scientific publications and conferences on the other. In 1995, the Multimedia Library was established to organise and integrate into the existing research library as many as possible of the archives of this production, thereby ensuring both preservation and access in con-

text. In 2002, the Hypermedia Studio was put in place, with the goal of exploring novel means of, and tools for, 'active' listening to digitally recorded music, for musicological research, teaching and individual appropriation. In conjunction, the External Relations department has been actively seeking partners,

via European projects and other means, to share, develop and valorise know-how, infrastructure and content.

THE MULTIMEDIA LIBRARY SYSTEM

The basic building blocks of the system, uninterruptedly operational, scalable, and evolutionary since 1996, consist mostly of off-the-shelf hardware¹⁷ and software¹⁸ components. As the modular design allows for replacement without interruption of service, the system has regularly been subject to upgrades reflecting the evolution of standards, software and hardware, converging towards an increasingly generic integrated solution for archival access and valorisation of cultural content.

From the very first version of the system, user access has been provided by means of Web technology,¹⁹ which allows in-house and external access (search, navigation and retrieval), in French and English, to a variety of multimedia content. All databases can be queried through forms, and each item (data, metadata) has its own URL. Domain-specific search strategies are provided. For example, the documentary database allows for searches by genre and subgenre, period, instrumentation, and so on. The library catalogue can be browsed according to the classification

scheme, and 'walked through' via a 3D virtual representation. Digital documents can be searched by type, availability and other musically pertinent criteria. However, while all the metadata is available from anywhere on the Internet, each item of data is access controlled, depending on the particular rights associated with it.

From within the library, special interfaces have been designed to provide convenient and secure²⁰ access. The Web browser, based on Internet Explorer, has been adapted to remove all unnecessary and unsafe controls,²¹ while the multimedia player, providing a familiar metaphor of a (physical) CD even for archives which never existed in this form (see Figure 1), shows the structure of the musical works (title, movements), information about the performers, as well as the booklet (for CDs) or the programme notes (for IRCAM concert archives).²²



Figure 2. Score annotation tool. The scanned score is annotated with the help of a graphical palette (on the left) specifically designed for musical usage. The annotations can be placed and moved along the time axis (lower part of the figure).

17 Servers and local clients: PC. Underlying network and protocol: Ethernet (100 Mb/s and 1Gb/s), TCP/IP. Good quality audio boards (Digigram) and headphones (Senheiser).

18 Web server: Apache, with PHP-MySQL extensions; Loris (library software system); 4D (composers database).

19 Accessible at <http://mediatheque.ircam.fr/>, subject to restrictions due to copyrights.

20 Securing the content was a precondition to be allowed by the rights holders and their representatives to provide this access.

21 The only available controls are 'back', 'forward', 'refresh' and 'stop'. It also takes over all of the screen and does not allow for escaping to other programs or to the system. This allows for securing both the platform and the content in this public space.

22 This player also works as a VCR when viewing movies.

FROM PRESERVATION TO ACCESS

All IRCAM performances (concerts, conferences) have been recorded since the late 1970s by the Production department, first on analogue, then on digital tapes for archival purposes. After the Multimedia Library system became operational and an adequate agreement was signed with the appropriate music rights societies, the following steps were added to provide for both digital preservation and access to these hitherto sealed archives:

1. Transfer to a mixed-mode CD²³:
each tape is transferred (by means of digitisation in the case of old tapes) to one or several audio tracks of a writable CD; the description of the event in a structured form is written in a computer-readable file in the first (data) track of the CD.
2. For each tape, two identical CDs are produced. One is archived, the other is transferred to the Multimedia Library.
3. The audio contents of the latter CD are extracted, compressed and transferred to the online area. The description contents are imported into the appropriate database.
4. The online metadata is then verified, validated and cross-linked with appropriate online resources.

As a result of this process, every document that is archived is made available online. We are currently developing various automation strategies for the last two steps above so as to facilitate retrospective online transfer of archives.

FROM ACCESS TO INDIVIDUAL APPROPRIATION

While books in a library are – at least – contextualised by their neighbours on the shelf, the issue of ‘nearness’ acquires a different meaning in the digital space; objects can have many neighbours with whom they can be put in context, either manually, automatically or both.

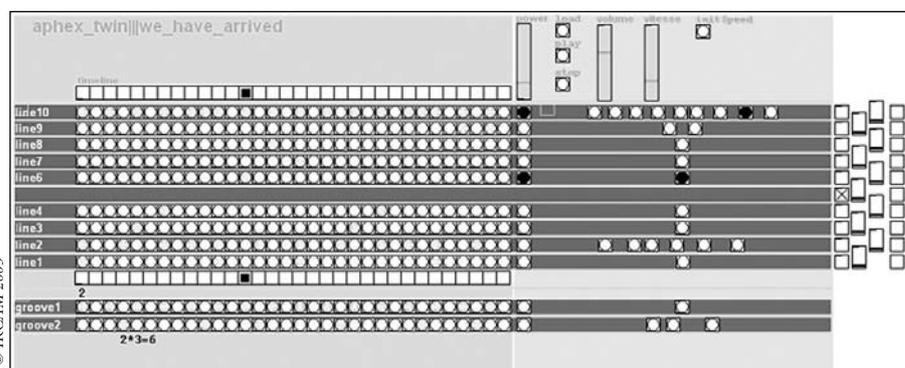


Figure 3. Signed Listening of an electronic music loop by Andrea Cera. The loop is manually decomposed into several audio layers, representing (as in sequencers) along the time (horizontal) axis the impacts. Some of the lines can be masked (here, the 5th one), so as to allow one to focus on other parts of the loop. Some operations such as slowing down can be used to refine the listening and synchronize it with the transcription. A condensed view of a phrase is inferred (in the right half) from the initial transcription. The simultaneous representation of several layers (bottom lines) allows one to search for and characterize implicit regularities in the transcription.

During the first stages of the project, this has been achieved by providing cross-links between related objects through their related metadata; the library record for a score provides links to the online biography of the composer and a documentary note about the piece, which in turn can link to all available online recorded versions, both from the IRCAM audio archives and commercial CDs.

The increased availability of off-the-shelf multimedia authoring tools has, therefore, allowed for the creation of higher-level integrated objects, such as musicological analyses (text, audio, diagrams), interactive score-and-performance presentations, Websites and so on – all based on the existence of the underlying sound archive and the ready availability of its material.

The question of the digital availability of these cultural contents within a technological context can be re-addressed nowadays in order to integrate usability in better and novel ways. Providing access to collections means not only making them available through structured and adaptive digitisation, but also allowing the user to

appropriate those elements s/he is interested in, through an individual, singular and unpredictable process.

The IRCAM Hypermedia Studio aims to provide tools for these purposes, based on technological experiments directed by users: musicologists, composers, music lovers. In particular, the project of Signed Listeners, launched in 2003, strives to propose instruments for personal listening purposes. Specialists of different types of musical listening are asked to imagine, using commercially available or IRCAM-designed technology, a representation of their individual way of perceiving a given piece or corpus. With the help of an engineer and a musicologist, each one authors a hypermedia mock-up²⁴ with the following characteristics:

- it exemplifies both a way to listen to, and a way to analyse, music;
- it allows the end listener to reproduce and/or alter the author's manipulations of the musical material (synchronisation, annotation, semi-automatic analysis of the signal, of the musical structures etc.), thereby becoming an author in his or her own right.

²³ A compact disc which comprises audio and data tracks.

²⁴ These are not necessarily functional objects, but models of what a finalised tool should look like after adequate development, allowing experts to test their ideas visually and aurally in a collaborative way.

These experiments (using various technical set-ups including Web plug-ins, DVD and installations) will allow us to determine both the generic and specific components of such a system, the former laying the ground for a possible access tool for all types of (musical) collections, the latter for specific corpora. For example, annotating a scanned score synchronised with a recording is generic to any corpus of music pieces that have both a score and a digitised recording (Figure 4 shows such a mock-up done at IRCAM for the MUSICWEB²⁵ European project). On the other hand, manual transcription of an electronic or repetitive music loop, with or without filtering the spectrum so as to reveal hierarchical rhythmic patterns, is specific.

TOWARD A NEW CONCEPT OF MULTIMEDIA TEMPORAL OBJECTS

Providing simultaneous access to content and to tools as a systematic process in this project is a necessity, in particular as IRCAM is a place for artistic creation and production of complex temporal objects. This is especially true of works belonging to the family of the new music of the second half of the twentieth century which mixes score and software, for which the modes of listening are rarely if ever codified and the analysis tools underdeveloped. As well as the general issues of access (e.g. indexation, navigation) to temporal objects there also arise questions of more or less local, more or less singular characteristics (modes of annotation, musical structures and so on) specific to learned Western music. This is why IRCAM is working on a new concept of a Web Radio (see Figure 4) closely related²⁶ to the Signed Listening project. The main stream (structured like a traditional radio programme, hence the name) provides pertinent access to digitised content. The 'auditor' will thus be able to explore the



Figure 4. Mock-up of Web Radio (example of a program about Edgard Varèse's *Déserts*. The main stream alternates interviews and musical excerpts (represented by the two colors). It allows the listener to access the full source documents used by the producer (here, an excerpt of *Le Sacre du Printemps* provides access to the full recording, represented by a longer horizontal bar below). The listener-spectator can thus navigate between a main stream and stored streams. In addition, he can appropriate himself any stream by annotating it; here, four lines of annotations mark excerpts from the program for different reasons (favorites, important citations, thematic indexation...).

content by accessing proposed ways of listening (by means of Signed Listeners provided by specialists) and by appropriating the content, and thus formalising his or her own listening on top of, or in parallel with, the one proposed by the editor.

NEW LISTENING TOOLS FOR AUDIO AND MUSICAL ARCHIVES

Music, and particularly contemporary music, raises acute and essential questions about the temporal object in the digital era – its exploration, annotation and restitution in all of its synthetic complexity, by individual users with singular modes of approach. The coordinated musicological, scientific and technological research programmes at IRCAM aim first at addressing them specifically, and then produce normalised and generic access tools for the listener, for whom the archive is a multiple-entry object, which can be interpreted in different, even contradictory, ways.

To allow for these entry modes by means of open and usable listening tools, IRCAM is about to coordinate the Semantic HiFi project.²⁷ The goal of this project is the development of a new gener-



A view of the Multimedia Library, as seen in its virtual 3D interface (<http://mediatheque.ircam.fr/infos/vrml/index-e.html>)

ation of HiFi systems offering novel ways of searching content, interacting with the musical material, and visualising and personalising the musical data.

The cultural and educational challenge specific to IRCAM-created music is vital: analysing it allows users to explore issues of listening and composition that are common to almost all kinds of music. It allows for the comprehension of most of the musical objects as complex temporal entities. The listening tools which will be developed in this context will thus open up new ways of accessing learned music, as well as learned access to 'simpler' music.

BACK TO PAGE 1

²⁴ These are not necessarily functional objects, but models of what a finalised tool should look like after adequate development, allowing experts to test their ideas visually and aurally in a collaborative way.

²⁵ New communication and information technologies in the music classroom, coordinated by the AEC (Association Européenne des Conservatoires, Académies de Musique et Musikhochschulen).

²⁶ Publication in the Web radio is done through a scenarisation file, which is an XML description (some existing standards are tested for this representation) of the sequencing of the various media and excerpts. The Signed Listeners can be seen then as either associated media or main stream.

²⁷ In collaboration with Sony, the Fraunhofer Institute, Native Instruments, the Universitat Pompeu Fabra (Barcelona, Spain) and Ben Gurion University (Israel).

DELIVERING *THE TIMES* DIGITAL ARCHIVE

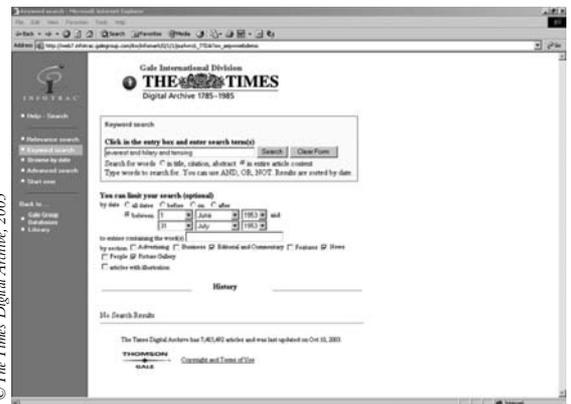
REG READINGS,
PRODUCTION DIRECTOR OF THOMSON
GALE AND MANAGING EDITOR FOR *THE
TIMES INDEX*, TALKED *DIGICULT.INFO*
THROUGH THE PROCESS OF BRINGING
'THE THUNDERER' TO THE WEB.

After two years of collective effort in Reading (UK), India and the USA, the digitisation of *The Times* from 1785 to 1985 is nearing completion. This article is a very brief account of Thomson Gale's (<http://www.galegroup.com/>) endeavours to create a searchable backfile of the oldest continuously published daily newspaper in the English language. Every headline, article and image, every front page, editorial, birth and death notice, advertisement and classified ad that appeared within the pages of *The Times* (London) is now easily accessible – a complete virtual chronicle of history for this period from the very first stirrings of the French Revolution to the release of Microsoft Windows, from the reckless driving of stagecoaches to motorway madness.

The digitisation process starts with our two Meikel M500 greyscale scanners and 1,250 reels of 35 mm microfilm. The individual pages are captured in greyscale and saved as 300 dpi bi-tonal TIFFs at a rate of six frames a minute. By carefully controlling the duplication of the microfilm masters and applying our own algorithms for gamma correction (light balance), noise removal and edge enhancement, we have been able to generate high-quality images for good OCR results. The images from the scanners are then de-skewed, cropped and de-speckled using

parameter-driven tools that we have developed in-house. Pages that were badly cropped prior to microfilming are manually cleaned and the title and page number electronically inserted. These processes not only make the final image look more attractive, but improve the accuracy of OCR on the text and reduce the overall file size for delivery over the Internet. Cleaned images generally range in size between 600 Kb and 1.6 Mb depending on content. Large display adverts or picture pages with a high black content approach the higher end of the range with some occasionally reaching 3 Mb in size.

The page images are delivered on DLT (Digital Linear Tape) to our contractors in India for article clipping, OCR, and the creation of fielded metadata made up of publication name, year, date, issue number, page number, title, subtitle, author, column position and illustration type. A category type is also added (e.g. 'Letters to the Editor' or 'Sport') to further segment the material and help end-users better limit their searches. Clipping the page into articles generates co-ordinates or positional information for each



© The Times Digital Archive, 2003

Figure 1: Keyword Search screen for *The Times* Digital Archive



© The Times Digital Archive, 2003

Figure 2: Sample Results screen. Showing the context in which an article was originally published (the article is highlighted in the thumbnail to the left) came high on the 'wish list' of project advisers.

article on the page and the OCR process provides the co-ordinates of every word within each clip or article. In this way articles can be extracted 'on the fly' from full-page images and the selected search terms highlighted on the delivered image. This approach obviates the need to store both full pages and clips on the image servers. The images, clipped images and XML containing the OCR'd text and

metadata are returned to the UK for checking before being forwarded to our US site for indexing and uploading to our own servers.

The process may appear both very technical and very mechanical, but software can only do so much. Many months were spent manually checking nearly every image for text that was too faint, too damaged or too dark to OCR well. Perhaps I should also mention the detailed examination of hundreds of pages to establish rules for the segmentation of articles and the management of a growing list of criteria for article categorisation. We have seen eighteenth-century issues without page numbers, issue numbers that appear more than once, mastheads printed upside down and articles with continuations in earlier columns or pages. Where the scanners have been unable to extract any detail from the film pages we have gone back to hard copy, but in many instances the originals are even worse.

In a project of this size and scope, planning is everything. Background work started over five years ago on copyright permissions, experiments with microfilm duplication, software enhancements, clipping and categorisation rules, XML and DTD creation as well as the design and build of the delivery platform. The endless design meetings and conference calls, though tedious at the time, really underpinned the project and ensured its success. And from a personal point of view it has been great fun: I am now a good deal more knowledgeable about the intricacies of image processing and file organisation and have a head stuffed full of pirates, pickpockets and ladies' maids.

To find out more about *The Times* Digital Archive, visit <http://www.galegroup.com/Times/>.



© The Times Digital Archive, 2003

Figure 3: The article delivered is the second of the three results shown in Figure 2. The search terms are highlighted in colour in the metadata and the body of an article - especially useful in longer articles.



© The Times Digital Archive, 2003

Figure 4: Page 6 of *The Times* of 8 June 1953. In the 'Browse by Date' search, mousing over a headline on the right-hand side outlines the related article in the large thumbnail image. Clicking on either the headline or within the outlined area will deliver the article selected.

PANDORA, AUSTRALIA'S WEB ARCHIVE, AND THE DIGITAL ARCHIVING SYSTEM THAT SUPPORTS IT



National Library of Australia

MARGARET E. PHILLIPS
DIRECTOR DIGITAL ARCHIVING,
NATIONAL LIBRARY OF AUSTRALIA

THE ROLE OF NATIONAL LIBRARIES IN WEB HARVESTING

The primary role of a national library is to develop and maintain comprehensive collections of documentary material relating to its country and its people, to preserve them and to make them available

for research now and in the future. During the last ten years, national libraries and other 'memory' institutions have had some difficult decisions to make regarding the extension of this role to electronic formats.

National libraries have accepted responsibility for online publications and a small number have embarked on programmes to develop collections, or 'archives', of them. Having accepted the responsibility, proceeding with it has been



no easy task. In the print world the principles and procedures are straightforward and well established and usually supported by legal deposit legislation, requiring publishers to deposit one or more copies of every publication with the national library. Preservation is also a relatively simple task with print material. As long as attention is paid to the security of the collection and it is kept under controlled conditions of temperature and humidity, there is an excellent chance that most of it will last hundreds of years.

The situation to be addressed with online publications is much more complex (and therefore much more costly) in every respect and national libraries have responded to it in a variety of ways. One of the first and fundamental decisions to be made is whether to archive online publications comprehensively (whole domain harvesting) or selectively. There are advantages and disadvantages with both approaches²⁸ and some national libraries such as Sweden, Norway, Finland and Iceland have taken the comprehensive approach, while Canada, Australia and Japan are archiving selectively.

As Web archiving technologies and practices mature, hybrid, more sophisticated approaches will be possible. An ideal situation would be a selective archive complemented by periodic snapshots of whole or part of a country's domain and/or subject-based snapshots such as the Election 2002 Web Archive²⁹ developed by the Library of Congress and the Internet Archive. The National Library of Australia (<http://www.nla.gov.au/>) keeps a close watch on developments in whole domain archiving with this possibility in view. It is also interested in the more focused approach being explored by the Bibliothèque Nationale de France, which would use automated methods to identify and archive significant sites.³⁰



National Library of Australia

Another basic decision is whether the publications will be acquired, or 'ingested', by deposit from the publisher or by Web harvesting using a harvesting robot or 'crawler'. Almost all archives established to date rely on Web harvesting, but the National Library of the Netherlands has built a major archive based on the deposit model.

Regardless of approach the archiving library must:

- identify what is being published;
- decide what is worth saving;
- define roles and responsibilities of different stakeholders;
- protect archived publications against loss;
- preserve the integrity and guarantee the authenticity of publications; and
- ensure long-term access and usability.

PANDORA: AUSTRALIA'S WEB ARCHIVE

The National Library of Australia selectively harvests Australian online publi-

cations for inclusion in PANDORA, Australia's Web Archive (<http://pandora.nla.gov.au/>). It does so in cooperation with eight other PANDORA partners, all of whom have responsibilities for collecting and preserving Australian documentary heritage. The mainland State libraries, the Northern Territory Library, ScreenSound Australia (the National Film and Sound Archive) and the Australian War Memorial are all participants.

Each of the partner agencies identifies, selects, catalogues and archives publications for which they accept responsibility as defined by their selection guidelines. After seven years of archiving Australian online publications, the National Library reviewed its guidelines³¹ in early 2003 and set priorities to focus effort on six main categories:

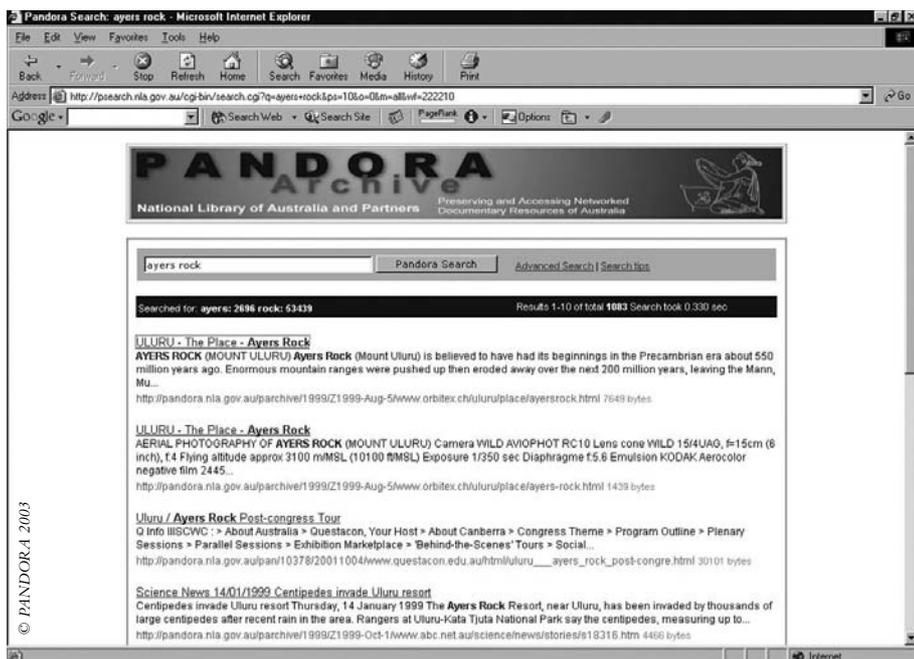
- Commonwealth and Australian Capital Territory government publications
- Conference proceedings

²⁸ A summary of the advantages and disadvantages of both the comprehensive and selective approaches is given in sections 3-5 of Margaret E. Phillips, *Collecting Australian Online Publications* (2003). <http://pandora.nla.gov.au/BSC49.doc>

²⁹ Library of Congress. *Election 2002 Web Archive*. <http://www.loc.gov/minerva/collect/elec2002/>

³⁰ Julien Masanès, 'Towards Continuous Web Archiving: First Results and an Agenda for the Future' in *D-Lib Magazine*, vol. 8, no. 12, December 2002. <http://www.dlib.org/dlib/december02/masanès/12masanès.html>

³¹ National Library of Australia, *Online Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia*. (2003). <http://pandora.nla.gov.au/selectionguidelines.html>



PANDORA interface showing the results of a search for Ayers Rock.

- E-journals
- Publications of tertiary education institutions
- Titles referred by indexing and abstracting agencies
- Topical sites:
 - (a) sites in nominated subject areas (defined in Appendix 2 of the selection guidelines) that will be collected on a rolling three-year basis; and
 - (b) sites documenting key issues of current social or political interest, such as election sites, Sydney Olympics, the Bali bombing.

The selective approach to archiving enables PANDORA partners to realise some important objectives:

- Each item in the Archive is quality assessed and functional to the fullest extent permitted by current technical and resource capabilities.
- Each item in the Archive can be fully catalogued and therefore can be included in the national bibliography.
- Each item in the Archive can be made accessible owing to the fact that permis-

sion to archive and make publications available via the Web has been negotiated with the publishers.

- The 'significant properties' of resources within the archive can be analysed and determined both for individual resources and for classes of resources. This enhances our knowledge of preservation requirements and enables strategies for preservation to be put into place.

Many of the titles in the Archive are gathered on a regular basis to capture new content, and each new gathering is referred to as an 'instance'. The PANDORA Archive is now over half a terabyte in size, contains almost 5,000 titles and over 9,000 'instances'.

With the permission of publishers, titles are harvested using HTTrack, a freely available Website offline browser and mirroring tool.³² A small number of titles are not available on the Web but are distributed by e-mail, and these are received directly from the publisher. Most titles in the Archive are freely available on the Web, but approximately 100 are com-

mercial publications. To protect the publishers' interests, access to these titles is restricted during the period of commercial viability nominated by the publisher. During this period the titles are available for consultation in the reading room of the archiving partner only, and the PANDORA digital archiving system manages this process automatically.

PANDORA DIGITAL ARCHIVING SYSTEM (PANDAS)

To achieve the greatest possible efficiency in creating a collaborative, selective archive, the National Library of Australia has devoted considerable effort to developing a digital archiving system to manage the process. The original intention was to buy a system off the shelf, as defined in the Information Paper, issued as part of the Digital Services Project.³³ The initiative did not find any system that met PANDORA's needs and the Library had no alternative but to develop one itself. In June 2001 the Library implemented the first version of the PANDORA Digital Archiving System (PANDAS) and version 2 followed in August 2002. Subsequent enhancements have brought us to version 2.1.3.

The PANDAS software supports the following archiving functions:

- Managing the metadata about titles that have been both selected and rejected for inclusion in the archive;
- initiating the gathering of titles selected for archiving;
- managing the quality checking and problem fixing process;
- preparing items for public display and generating a title entry page;
- managing access restrictions; and
- providing management reports.

The software is Web-based and enables all eight partners from their remote locations to carry out all of the tasks necessary to download and store titles in the central archive located on the National

32 Xavier Roche, HTTrack Website Copier: Open Source Offline Browser (<http://www.httrack.com>).

33 National Library of Australia, *Digital Services Project: Information Paper* (1998). <http://www.nla.gov.au/dsp/infopaper.html>



Library's server. The software runs on a standard PC and with Internet Explorer 5.5 or above.

The collection managers use PANDAS to store descriptive, administrative and preservation metadata about the Web site being archived, such as:

- descriptive information (title, URL, persistent identifier, publisher details);
- permission status (whether permission to archive was granted or denied by the publisher);
- type of resource;
- status (selected, rejected, still being monitored);
- subjects (selected from a list of broad categories);
- collection name (if the title is being included as part of a collection of sites on a particular event or topic, such as an election campaign);
- restriction information (e.g. access in-house only for a specified period);
- gathering schedule.

The gathering schedule can specify that a site should be archived once only, or it can specify a gathering frequency (monthly, quarterly, annually, etc). It is also possible to enter a date list so that an item is gathered on specific dates.

The PANDAS software allows a variety of Website copying software to be 'plugged-in' and allows the collection manager to control the scope and timing of the gathering process. As part of the gathering process, technical metadata about the file types captured is recorded by the system. This information is used to generate automatically public notes listing any plug-ins required for viewing the site.

PANDAS currently does not have any automatic quality assurance feature such as a link checking feature. Collection managers must check the whole site manually to identify problems, such as

Javascript not working correctly, missing files, or broken links due to coding errors or case sensitivity.

After a site has been gathered and checked for completeness, it is prepared for public access in the Archive. A script is used to disable external links, forms and mailto addresses automatically. When users click on one of these links in the archived copy, a screen will appear giving them the option to leave the archived version of the site and access the link to the live site (if the latter still exists).

A public 'title entry page' is generated for new titles, or is updated when a new instance of a title is gathered. The title entry page indicates whether the title is being gathered regularly, provides a link to the publisher's live site if it still exists, lists all the archived instances of the title, provides information about software required to view it, and indicates any access restrictions. It also provides a link to the publisher's copyright statement if there is one, and to the National Library's own generic copyright warning. Subject and alphabetic title listings are automatically updated when new titles are added to the Archive.

The PANDAS software manages access restrictions on the titles that require them. Restrictions may be applied for commercial reasons, for privacy or cultural reasons, or as part of a policy decision for certain categories of material. Items can be restricted for a specified time period to use within the Library building only, or can be password protected so that designated users can obtain access. PANDAS manages these restrictions and allows only those researchers in designated locations (based on IP address) or with the required password to access the archived versions. As time periods for restrictions expire, the system automatically updates the title entry pages which indicate the changed access conditions.

FUTURE DIRECTIONS

The selective approach to archiving is very labour-intensive. While PANDAS has enabled us to realise some efficiencies and to increase productivity, it still requires manual entry of metadata and processing of titles one by one.

In order to be able to deal with the increasing volume of online publishing, particularly in the government sector, the Library is undertaking a pilot project to explore the possibility of automating or semi-automating the processes of identifying, selecting, describing and archiving Commonwealth government publications. To automate identification, selection and description, it is working with seven government agencies to examine the metadata they produce and to refine workflows for making it available to the National Library for inclusion in the National Bibliographic Database (NBD).³⁴ The aim is then to distil from this a small number of best practice models that all government agencies could use to contribute metadata to the NBD. In a second phase of the pilot project, metadata would be extracted from the NBD and batch loaded into PANDAS for archiving. Before this can be done, PANDAS will need to be further developed and this will be undertaken as part of the development of version 3.

The main driving forces for the future development of PANDAS are:

- the need to be able to ingest a much larger volume of online publications and their associated metadata, automatically or semi-automatically;
- the need to improve discovery of and access to items in the burgeoning Archive in a variety of ways;
- the need for PANDORA to be able to interface with data acquired other than by the current selective harvesting method. This may include large amounts of publisher-supplied data, partial or whole domain harvests, and publications

³⁴ The National Bibliographic Database is a database to which approximately 1,100 Australian libraries contribute catalogue records and library holdings and it is made available through the Kinetica service (<http://www.nla.gov.au/kinetica/>).



and databases from the 'deep Web' obtained as a result of the deep Web archiving initiative described below;

- the need for a collection manager interface to the preservation metadata, and mechanisms to support preservation processes; and
- the need for better and more flexible reports from the system.

ARCHIVING THE 'DEEP WEB'

To date, titles archived in PANDORA have mostly been gathered using a robot that follows html links from the root URL submitted as the starting point for gathering.

A very large part of the Web consists of sites not accessible via html links but stored in structured databases, which it is necessary to query for presentation of information. These queries are typically made via selection of options from drop-down boxes or by entering terms into a search engine. This requires intelligence that gathering robots and search engine indexers do not have at this stage. The hidden nature of data stored and accessed in this way has led to its being known as the 'deep Web'.

PANDAS EVALUATION SYSTEM

Because libraries and other collecting institutions around the world are beginning to take on the responsibilities of collecting digital publications, and because as yet there are few digital archiving systems available to assist them, the National Library of Australia has received a number of enquiries about the use of the PANDAS software. The Library has decided to make the software available on a cost-recovery basis, although it will be limited in the number of agencies it can accommodate. To enable interested libraries to assess whether PANDAS does in fact meet their business needs, the PANDAS Evaluation System has been developed and has recently been made available to the first evaluating library. A maximum of three agencies at any one time can evaluate the software.

The 'deep Web' presents special challenges to national libraries that are endeavouring to identify and archive a nation's output in this form, whether they are taking the selective or comprehensive approach (both of which rely on harvesting robots). The National Library of Australia has embarked on a research project to find ways of taking in, managing and providing access to publications and Websites structured as databases. It is also participating in the Deep Web Archiving Group of the recently formed International Internet Preservation Consortium³⁵ and will collaborate with other members in developing solutions and sharing results.

CONCLUSION

The business of building a digital archive, managing it in the long term, and providing access to it is a complex and costly operation. A digital archiving system, which manages workflows, stores the digital objects themselves, as well as descriptive, administrative and preservation metadata, and provides management reports, is essential to support it. This system must develop hand in hand with the archive itself, growing to facilitate expanding business requirements and enhancing productivity at every stage. The National Library of Australia is about to embark on another development cycle for its digital archiving system.

[BACK TO PAGE 1](#)

THE FIRENZE AGENDA (17 OCTOBER 2003)

Reprinted without amendment by DigiCULT as the Agenda itself requested.

In response to the challenges of preserving digital memory, a group of experts have proposed an agenda with focussed objectives addressing creation, preservation and access issues for both digitised and born-digital objects. The Italian Presidency, the European Commission, ERPANET, and the MINERVA project are the promoters of this initiative in the philosophy of eEurope and linked to the National Representatives Group (<http://www.cordis.lu/ist/ka3/digicult/nrg.htm>).

The agenda covers a short period (12-18 months) identifying concrete and realistic actions which respect the interests of museum, libraries and archives, and the differences between media formats. It is an open process integrating ongoing actions and the voluntary efforts of participants. The experts have identified some initial responsibilities for each of the actions, and progress will be reviewed in one year. The Firenze agenda has been submitted to the National Representatives

Group (NRG) for endorsement and to encourage each Member State to support the initiative.

Action Area 1 of the agenda addresses problems and risks. Probably the most important task today is to create awareness about risks and problems among decision-makers at all levels. Area 2 considers ongoing initiatives and currently available technologies, whereas Action Area 3 examines the legal and regulatory

³⁵ The International Internet Preservation Consortium was established in July 2003 under the leadership of the Bibliothèque Nationale de France (<http://www.bnf.fr/>).



implications and responsibilities which must now be addressed, and who is responsible for finding solutions for digital memory preservation.

ACTION AREA 1: CREATE AWARENESS & COOPERATION MECHANISMS

The first step is to build community and consensus, focussing initially on decision-makers. The expert group will work with user communities to capture their requirements and suggestions and develop suitable approaches to address those needs. They will build consensus beyond the current audience, towards professional bodies and establish cooperation with other initiatives in the field. The expert group will report regularly on progress and planning.

- Workgroup activity and events, e.g. workshops, seminars, papers, helpdesk and electronic fora;
- Report of activity to NRG every six months.

Co-ordinators: ERPANET and MINERVA projects.

ERPANET: coordination and Web services for the expert group including forum, projects and literature assessment, the help-desk advisory service, erpaEprints; workshops on 'trusted digital depositories' (Rome, November 2003); seminar on 'scientific digital objects' (Lisbon, December 2003); publication of the workshops' and seminars' final reports.

MINERVA: network of the national representatives for promotion and for data collection; workshops or seminars or events organisation on the field like Bibliocom on 30 October; NRG report by the end of 2003 including a slot to present progress of the initiative.

DELOS: annual digital preservation summer school; annual state-of-the-art studies from the digital memory preservation cluster.

PRESTOSPACE: workshops or seminars or events organisation on the field; annual state-of-the-art on digital preservation for audio-video archives. Workshop on 'user requirements for audio-visual preservation', March 2004, Amsterdam.

ACTION AREA 2: EXCHANGE GOOD PRACTICE AND DEVELOP A COMMON POINT OF VIEW

Data collection on current practices, and selection and promotion of good practice are necessary steps to enable the sharing of progress and to develop a common knowledge basis on suitable approaches and technologies. The focus here is to identify weak or underdeveloped areas of practice. Lessons already learned from examples of good practice must feed into tutorials and papers (e.g. on digital collections, trusted repositories, metadata, IPR, selection criteria, and Web archiving) and must be the basis for specifying a 'European interoperable platform'.

- 'State-of-the-art' on existing initiatives, good practice selection and promotion.
- Technology watch reports on emerging issues.
- Propose a research agenda on technological priorities and challenges about digital preservation.
- Training (skills) initiatives and programmes.
- Issues of stakeholders/users.

Co-ordinators: ERPANET and DELOS projects.

ERPANET: data collection and analysis to identify good practices: description and analysis of single and collective case studies and creation of erpaTools; promotion of erpaAdvisory services; inform the preservation research agenda (see below).

DELOS: continued review and coordination of the NSF/DELOS Digital Preservation and Archiving research agenda; studies and papers on the field; produc-

tion of recommendations.

DIGICULT FORUM: technology watch and tutorials on the field; contribution of technology specifications for a research agenda; creation of an 'Action in the Preservation of Memory' section in the *Digicult.Info* Newsletter.

PRESTOSPACE: technology specifications for a research agenda for audio and video archives; production of recommendations. Meeting in collaboration with the FIAT annual conference, on 'technical specifications and system architecture for an audio-visual integrated preservation system', October 2004, Marseille.

MINERVA: data collection infrastructure for accumulation of good practice experience; training trials on the long-term sustainability of digitised scientific and cultural objects in cooperation with ERPANET and DELOS.

ACTION AREA 3: LONG-TERM POLICIES AND STRATEGIES

Lack of clear reference policies and responsibilities today is a serious risk in the future preservation of our memory. The expert group will stimulate national/sectoral authorities in order to develop digital preservation policies, research/training programmes and trials as appropriate. In particular, the legal framework for electronic resources/records collections and deposits may be incomplete, not only for libraries and archives but also for e-government, for education, for research and other environments of our information society. The adoption of a common agenda could be instrumental in catalysing national efforts on a coordinated framework at the European level and investment in collaborative public/private projects.

- Identify reference actors, within the Member States, to act as coordinating bodies for national initiatives and to produce recommendations.



- Analyse and compare current legal and regulatory contexts leading to specifications for suitable frameworks for stakeholders, in particular for memory institutions, for permanent access to digital deposits (with particular attention to objects that are 'born digital') and online registers of trusted preservation repositories.
- Study and propose suitable business models and costs analysis and long-term funding strategies for public institutions.
- Foster cooperation with emerging international activities such as that being developed under the auspices of UNESCO.
- Validate and promote the research agenda at EU and Member State levels with special emphasis on encouraging national funding agencies to allocate resources to enabling research in line with the evolving research agenda.

Co-ordinator: Italian Presidency. The Italian Presidency will lead this area, but it is crucial to have long-term commitment by future Presidencies and Member States authorities.

ITALIAN PRESIDENCY: developing the report on risks and emergencies with at least four other cases and the report on legislation and policy frameworks; implementing Website for the initiative and ensuring continuity in cooperation with the future Presidencies.

PRESTOSPACE: bridging the gap between technical results and service implementation, meeting in Paris, December 2004.

ERPANET: improving the understanding of the interconnectedness between e-government and e-citizenship viability and the long-term authenticity and integrity of digital entities; providing an annual review of national policies and strategies; enhancing and promoting policy and legal framework tools.

[BACK TO PAGE 1](#)

NZ ELECTRONIC TEXT CENTRE

AS THE NEW ZEALAND ELECTRONIC TEXT CENTRE APPROACHES ITS SECOND BIRTHDAY, ITS FOUNDER AND DIRECTOR, ELIZABETH STYRON, SPOKE TO *DIGICULT.INFO* DURING A VISIT TO HATII (UNIVERSITY OF GLASGOW) ABOUT HOW THE CENTRE WAS FORMED, AND ITS WORK, AIMS AND SERVICES.

The New Zealand Electronic Text Centre (<http://www.nzetc.org/>) generates standards-based digital content and provides a wide range of services to both educational and commercial organisations. NZETC, based at the Victoria University of Wellington, is a vibrant institution which, despite its young age, has already won the confidence of the cultural heritage community. The Centre was first conceived by Elizabeth Styron, its current Director, while she was involved in the Early American Fiction project (<http://eaf.chadwyck.com/>) at the University of Virginia. While on a Fulbright fellowship in New Zealand, using TEI (<http://www.tei-c.org/>) and XML, and working on the development of e-text coding, Styron put a proposal to Victoria University to create a centre based on the model used at the University of Virginia. The project was funded by the Faculty of Humanities and Social Sciences at Victoria, and in December 2001 the NZETC was born.

NZETC has three major aims:

- to create a free online archive, focusing on the Pacific Island and NZ region;
- to provide resources and services for others;
- to establish and develop a community of people who have expertise in the creation and use of digital resources.



The digital archive, accessible from the NZETC Website, is a constantly expanding collection of searchable XML texts and images. The document delivery service is open source: a Java servlet called Cocoon, which is run inside the Apache Tomcat servlet container. zCocoon is a flexible environment that compartmentalises the transformation of documents from XML via XSLT, allowing the delivery of a variety of formats.

NZETC also acts as a source of expertise for the digitisation community in New Zealand. It can host collections created externally, perform document conversion and repurposing, open-source e-publishing, metadata creation and digital imaging, and provide consultation services for both the public and private sectors and project management for activities involving XML, SGML, TEI, text and image conversion, and multiple e-book delivery formats. In addition, the Centre provides institutions with advice on the digitising of their own materials. Public institutions are particularly interested in the open-source solutions made possible by this expert advice. NZETC also collaborated in the creation of the Heritage Materials Imaging Facility (<http://www.hmif.co.nz/>). Housed at the National Library, this facility opens up new digitisation possibilities for the whole country.



Through regular seminars and the training of graduate students, as well as instructional outreach activities both within the university and for the public, NZETC is gradually building up a community of knowledge about XML, electronic publishing and the creation and use of digital resources.

The policy of outreach allows the Centre to identify unexpected users of its resources and collections. It has evolved beyond the simple model on which it was based and has become a truly wide-ranging focus for digitisation efforts in New Zealand. Projects with which the Centre is currently involved include the Greenstone project at Waikato University,

who developed the Greenstone open-source Digital Library delivery system (<http://www.greenstone.org/english/home.html>), and the Online Encyclopaedia of New Zealand project (<http://www.teara.govt.nz/>). The Centre is working on a print-on-demand repository and will continue to develop its activities to meet users' demands well into the future.

[BACK TO PAGE 1](#)

METADATA DEBATE: TWO PERSPECTIVES ON DUBLIN CORE

TO PRESENT A BALANCED VIEW OF THE METADATA STANDARD DUBLIN CORE, *DigiCULT.Info* INTERVIEWED TWO CULTURAL HERITAGE PROFESSIONALS WITH DIFFERING OPINIONS: DR ANDREW WILSON, THE PROJECT MANAGER OF DIGITAL PRESERVATION AT THE NATIONAL ARCHIVES OF AUSTRALIA ([HTTP://WWW.NAA.GOV.AU/](http://www.naa.gov.au/)), AND DR WENDY DUFF FROM THE UNIVERSITY OF TORONTO ([HTTP://WWW.UTORONTO.CA/](http://www.utoronto.ca/)).

THE DUBLIN CORE AND ITS LIMITATIONS WENDY M. DUFF

The Dublin Core (DC) is a simple 16-element metadata standard for resource discovery, which aims to facilitate the creation of metadata by end-users. The Dublin Core has four basic goals: simplicity of metadata creation and maintenance, commonly understood semantics, international scope, and extensibility. The goal of extensibility acknowledges that some projects and domains may find the Core inadequate to their needs. The Core encourages metadata experts from other domains to extend the Core with additional elements as the Australian Government did when it created the metadata standard for the Australian Government Locator Service. The goals of the Dublin Core should be applauded. Information retrieval on the Web could benefit from having consistent metadata elements linked to resources; however, the question remains whether or

not the Dublin Core is sufficiently robust and usable, and sufficiently used, for *archivists* to adopt. I would suggest that it fails and that a careful examination of the assumptions underlying it points to the difficulties. These assumptions include the ability and willingness of individuals and institutions who create Web resources to supply Dublin Core metadata, the applicability of the Core to the archival domain, as well as user expectations when they search the Web.

END-USERS

As previously noted, the Core was developed as a simple metadata scheme that end-users could employ. A project to conduct five annual surveys (1998–2002) of the Web administered by the OCLC Office of Research sheds some light on whether creators of Web resources use it.³⁷ The survey found that metadata usage is on the rise, though it noted that many new HTML editors automatically supply metadata tags. The researchers

found that the average number of tags per page went from a low of 2.32 in 1999 to a high of 3.14 in 2002. They also noted the limited number of pages that used formal metadata schemes. They found that 'Dublin Core metadata appeared on only 0.5 % of public Website home pages in 1998' and the number rose to only 0.7 % in 2002. They concluded that little progress had been made in the use of metadata standards and that existing metadata tags are *ad hoc*.

UNIVERSALITY AND DOMAIN NEUTRALITY

The principle of extensibility suggests that other domains can use the Core – though they may need to augment it with domain-specific elements. The unstated assumption of this principle is that the Core is applicable across most type of resources and that all domains could use it. I would suggest that this requires a domain neutrality which DC does not have. Metadata standards involve the identification, delineation and categorisation of information. Categories of information

37 Rick Bennett, Edward T. O'Neill and Brian F. Lavoie, 'Trends in the Evolution of the Public Web: 1998 – 2002', in *D-Lib Magazine* (April 2003). <http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>



are not neutral. They reflect a particular view of the world and the Dublin Core is no exception. Bowker and Star point out that 'categories are historically situated artefacts, and like all artefacts, are learned as part of membership in a community of practice.'³⁸ Taxonomies are grounded upon, shaped by, and reflect the world view of their creators. The Core is inherently a standard that reflects a bibliographic view of the world and assumes that the entities being described are single objects that emanate from an intellectual endeavour. The sixteen elements are grouped into three categories: elements that relate to content, elements that relate to intellectual property, and elements that relate to instantiation. The elements related to intellectual property – contributor, creator, publisher and rights – describe a limited set of relationships an object can have to a person or institutions. Some view this distinction as arbitrary at best. Bearman, Miller, Rust, Trant and Weibel point out that this categorisation might meet the needs of bibliographic items, but these divisions are not sufficient for many information objects. They state that:

*'The terms Creator, Contributor and Publisher are often used as if they roughly correspond to the relationships between an Agent and a Work (he wrote this play), an Agent and an Expression (she acted in this play), and an Agent and a Manifestation (they published this script). However, the Creator, Contributor, Publisher distinction does not provide a fine enough classification to specify such vastly different roles as editor, composer, or actor, or distinguish these roles from others with a lesser creative impact on a manifestation such as typographer, foundry, or audio engineer.'*³⁹

The relationships that are highlighted are those traditionally emphasised in bibliographic description. Moreover, the relationships supported by DC are not the ones archivists stress in their descriptions. For example, various rules for creating archival metadata emphasise the importance of documenting the provenance of records. Provenance is linked to creation, but creation is defined as creation, and/or accumulation and use. The creator of an archival document might be the person who, or organisation that, acquired and used the document, not the 'entity primarily responsible for making the content of the resource,'⁴⁰ as the Core defines it.

INFORMATION RETRIEVAL

The Core strives to improve information retrieval across different domains and to enable access to various types of material. The Web provides the means to open up access in an unprecedented way. However, we need to study our users to understand better how they distinguish among different types of material. Will the users who seek published documents become confused when they retrieve business records in their search? Will users find search results with DC tags more understandable than results without these tags? Will the user be able to distinguish between a representation of a series of records consisting of photographs, textual documents and a sound recording from a description of a single report? If archivists use the creator element for documenting provenance, will they help or hinder the searchers on the Web? We need to find answers to these questions before we simply endorse or adopt the Core.

Improved access to the rich resources on the Web is a noble cause but the results of study by Bennett *et al.* may indicate that it may be a lost cause. The DC aims to support greater collaboration across

domains. Collaboration is important and the various professions should work together to build a model that supports different perspectives. Furthermore we need more research that investigates how the Core is used and its impact on those who seek information on the Web. With collaboration, research and automated tools for supplying metadata we may develop a Core that meets the needs of *all* types of users who seek information on the Web.

WHY THE DUBLIN CORE METADATA INITIATIVE (DCMI) IS IMPORTANT DR ANDREW WILSON

First, I had better make my biases clear. I have been attending DCMI Workshops and Conferences for the last five years. I am a member of the Dublin Core Usage Board, the Dublin Core Advisory Board, and am currently co-chair of the Dublin Core Preservation Metadata Special Interest Group. For the previous three years I was co-chair of the Dublin Core Governments Working Group.

In response to a number of non-specific criticisms of the Dublin Core metadata element set (DCMES) I recently articulated my own thoughts about DCMI and its achievements over the last 8-10 years at the ERPANET Workshop in Marburg in early September 2003 (<http://www.ermanet.org/www/products/marburg/marburg.htm>). My intention was to help those at the workshop understand a little better how significant the DCMI has been for us all, and for metadata generally.

I want to preface my remarks by saying that the Dublin Core metadata element set (DCMES) was never advertised as being a perfect element set, and the community had never undertaken to deliver perfection. As Tom Baker, perhaps the most

38 Geoffrey C. Bowker and Susan Leigh Star, *Sorting things out: Classification and its Consequences* (Cambridge, Mass: MIT Press, 1999), p. 287.

39 David Bearman, Eric Miller, Godfrey Rust, Jennifer Trant and Stuart Weibel, 'A Common Model to Support Interoperable Metadata: Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities' in *D-Lib Magazine* (January 1999). <http://www.dlib.org/dlib/january99/bearman/01bearman.html>

40 Diane Hillman, *Using Dublin Core* (August 2003) <http://dublincore.org/documents/usaguide/> (accessed 10 November 2003).



© HATHI (Uo) Glasgow, Scamus Ross, 2009

National Archives of Australia

significant DC theorist, has said, DC metadata is a set of signposts for digital surfers, i.e. it is there to guide people to resources, and does not claim to be a GPS system or even a road map. Those who try to fit DCMES into tasks for which it was never designed are the ones making the mistakes. It is not the fault of the DCMES that it cannot be used for record-keeping, or for database management, or for preservation metadata, except insofar as the success of DC has encouraged people to think it can be bent sufficiently to allow it to fit almost any application involving metadata. Those who try to misshape the DCMES in these ways do not understand enough about metadata or enough about the aims and purposes of DC, and DC itself should not be blamed for such misuse.

The Dublin Core community began in 1995, mostly comprised of people from the North American library and academic communities, but DCMI participants quickly became truly global in nature. This is where the major success of the Dublin Core community lies, and it is an achievement that should be taken as a

shining example for us all. The DCMI was the first truly global community, in the area we can generalise by calling information management, to develop a successful metadata standard. DCMI showed that it can be done – that a diverse group of individuals can work together to develop a standard that can be used globally by all the communities which participated in its development. DCMI demonstrated that such communities can be put together and can work cooperatively to achieve their aims and desired outcomes. The Dublin Core metadata element set, despite its flaws and shortcomings, is a viable standard for resource discovery metadata, as is shown by its very widespread implementation across a large range of communities. Without the example of DCMI I do not think there would be groups like ERPANET (<http://www.erpanet.org>), and the other global initiatives in the field of information/knowledge management.

The second area in which I believe that all metadata implementers and practitioners should be eternally grateful to DCMI is the work on other technologies

that support metadata. These tools and technologies support all types of metadata, not just DC metadata. A prime example of this is the work on registries by UKOLN (<http://www.ukoln.ac.uk>), the European Commission, and other bodies like OCLC (Online Computer Library Centre: <http://www.oclc.org/>). It seems to me that any projects on registries and any real-life implementations of registries have all been initiated as a result of the work of the DCMI DC-Registries Working Group. It is a similar case with application profiles. These are probably one of the more significant technologies for helping to achieve metadata interoperability being developed currently, and without DCMI the work of CEN (European Committee for Standardisation, <http://www.cenorm.be/cenorm/index.htm>) on application profiles, for example, would have taken a lot longer to initiate. Again, if RDF ever takes off and lives up to the promise envisaged for it in the ‘Semantic Web’ concept,⁴¹ this will be due, in no small part, to the use and encouragement of this technology by metadata communities.

So, all we information professionals have a lot for which to thank the DCMI. And the DCMES is developing to take into account the practical needs of the communities that make up the DCMI. Thus, the purpose of working groups such as DC-Government is not to develop a set of elements for government resource discovery but to look at ways to make DCMES more relevant to the task of discovering government resources. DCMI Working Groups are not about developing more metadata standards (we all know that there are already more than enough out there) but to help make the DCMES do what it does a little bit better for specific communities. I think this is a crucial point for everyone to understand – if we do not understand this, then we do not understand properly what the purpose, aims and drivers of the DCMI are.

41 Cf. DigiCULT's *Thematic Issue 3: Towards a Semantic Web for Heritage Resources* (http://www.digicult.info/downloads/thematic_issue_3_low.pdf).



Finally, let me say that, while I understand the issues archivists and record-keepers have about metadata and DC metadata in particular, one of the reasons that DCMES does not quite fit our context is because we, as a profession, are very late in coming to dine at this particular table. Librarians were thinking about these issues long before archivists/record-keepers were, and therefore had more influence over the shape of the DCMES and the

DCMI than other communities. We should not decry this and throw the baby out with the bathwater but all work together. The DCMES is not the answer to every metadata need but it was never intended to be. Instead of criticising DCMI, I believe it behoves all of us to endeavour to work together as a community of like-minded individuals intent on making the DCMES as good and as useful as it can be within its own constraints.

[BACK TO PAGE 1](#)

The papers of the 2003 Dublin Core Conference: 'Supporting Communities of Discourse and Practice – Metadata Research and Application', which was held in Seattle at the end of September 2003, are available online from: <http://www.ischool.washington.edu/dc2003/index.html>

CISTERCIANS IN YORKSHIRE: CREATING A VIRTUAL HERITAGE LEARNING PACKAGE

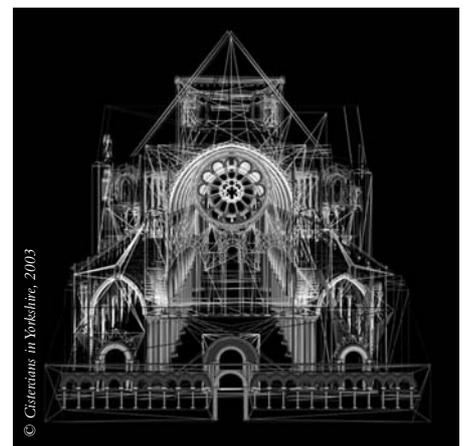
TO FOLLOW UP OUR PREVIOUS COVERAGE OF 3D RECONSTRUCTIONS OF ARCHAEOLOGICAL SITES, *DIGICULT.INFO* SPOKE TO CARL SMITH, DIGITISATION OFFICER FOR THE CISTERCIANS IN YORKSHIRE PROJECT BASED AT THE UNIVERSITY OF SHEFFIELD.

The Cistercians in Yorkshire project (<http://cistercians.shef.ac.uk/>), funded with a grant from the New Opportunities Fund's UK-wide digitisation programme, is creating a Web-based learning package centred around three-dimensional virtual heritage reconstructions of Cistercian abbeys in Yorkshire. These reconstructions are seen by the project team as providing both a way into the learning materials on the Website and an informed insight into the Cistercian mindset. The project's central ambition is to use the architecture as an explanatory teaching device, to answer questions about the nature of this expression of the monastic ideal through explicit reference to the shape and form of the buildings within which it was articulated.

The Cistercians, or 'white monks', played a major role in the religious and economic life of medieval England. They became significant land-owners in

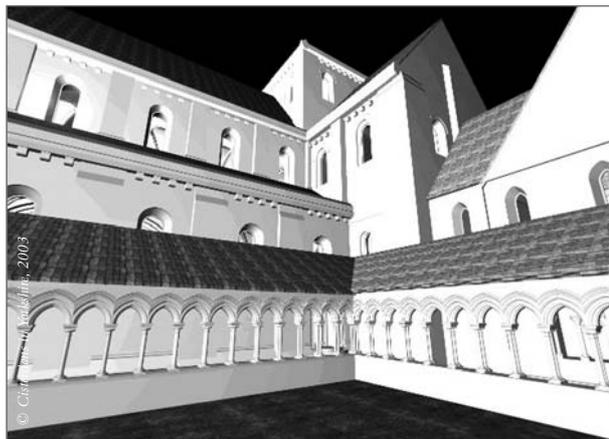
Yorkshire (and beyond) and had an enduring impact on their local environment. There are only six Yorkshire Cistercian abbeys with surviving fabric. Among the Yorkshire houses, Fountains and Rievaulx, both founded in 1132, are of notable importance and remain popular with visitors of all ages; Fountains Abbey, indeed, is a World Heritage site. Roche was chosen because of its location in South Yorkshire near to Sheffield and because of the architectural significance of the building (one of the earliest examples of 'New Gothic' architecture in northern England). We elected to work also on Kirkstall in Leeds in order to contribute to the celebration of the 850th anniversary of that monastery's foundation in collaboration with the Abbey House Museum in Leeds. Many modern communities in Yorkshire live on, or near, land that was once owned by the Cistercian order; this project will encourage potential users, such as residents and local schools, to explore the history of their local area.

Areas of humanities research that can benefit from the implementation of Virtual Reality (VR) include those that require greater interactivity in the learning process. A well-designed virtual environ-



A wireframe rendering of the exterior of Roche Abbey

ment should allow the users to regularly ask their own questions while analysing the associations between pieces of information, rather than just isolated facts. At the heart of the Cistercians project are 3D virtual reconstructions of the abbey churches and monastic buildings, which will be accompanied by pages explaining the history and the social, economic and cultural significance of each foundation. We are creating high-quality, architecturally accurate 3D models of buildings using CAD and 3D MAX technologies. These packages have been chosen on the basis that they are industry standard and maximise the chances of future file migration.



A rendered view of Roche Abbey from the cloister

The models are then rendered using historically accurate textures and surfaces. Reconstructing these ancient structures, which have not, and could not, be readily visualised before, has intrinsic value and is at the core of the project's ambitions. Even expert architectural historians can struggle to rebuild ruins of this structural complexity in their mind's eye; uninstructed lay visitors can find that imaginative leap completely beyond them. This research makes a major contribution to modern understanding of the architectural construction of these now ruined buildings and sheds new light on the Cistercian use of sacred space. By combining 3D modelling technologies with rigorous conventional historical and archaeological research and expert architectural analysis of the surviving fabric, this project is genuinely multi-faceted and

interdisciplinary. Its mission is to make the fruits of this scholarly research accessible and appealing to the widest possible audience.

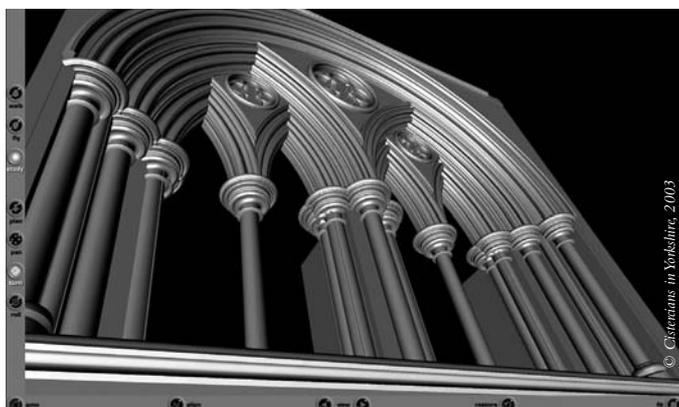
At present most VR-related projects place an overwhelming emphasis on the final presentation of the model. As a result, the developmental stages are rarely made available to researchers. However, for these reconstructions to

work as efficient mechanisms for encoding new interpretations and ideas, any decisions that have been made during the design and construction stages of the project must also be made explicit. We would like our users to become familiar with these architectural structures from the point of view of their units of construction.

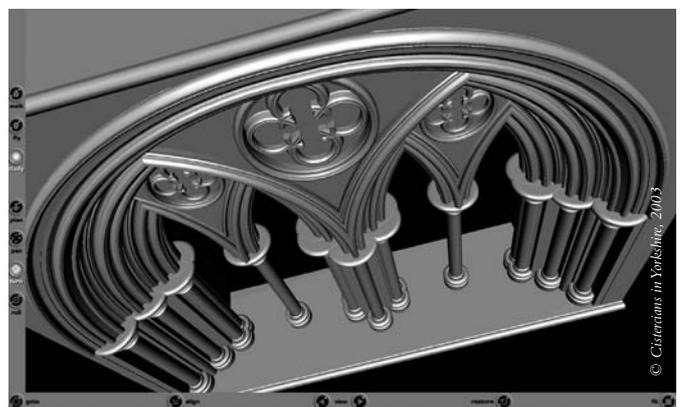
Therefore, access to the source material and supporting evidence for every reconstruction is established and maintained. The project database includes material catalogued from each stage of the creation of the resource including references to varied information: architectural 2D drawings of ground plans and elevations; reconstruction drawings; photographic surveys of the sites; photographs of the models; artistic impressions; movies and

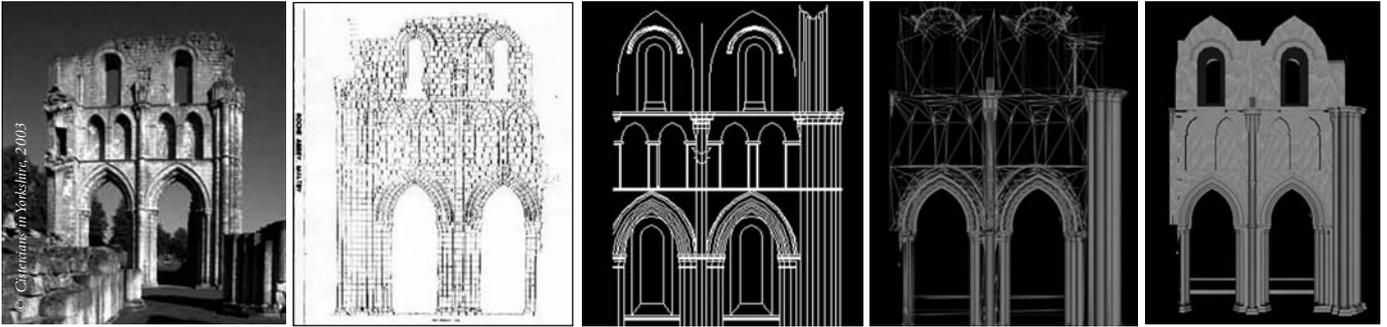
Quicktime files; VRML files; MAX files; as well as visual and textual descriptions of the methods of the 3D reconstructions. Ultimately the direct linking of the models with the database will allow queries of all the associated data. Inheritance will be used to allow the desegregation of the designs into their original elements – where the output of one phase can be seen as the input to another. The final learning packages will contain a wealth of material including images, 3D models, text and sound about the Cistercians, who they were, how they lived, what they believed, and why they were significant in the history of medieval Yorkshire. The architecture of each site, explained in the context of other local churches (and European Cistercian abbeys), will open visual and textual windows onto the Cistercian way of life as experienced by the monks, the lay-brothers and their secular neighbours.

The key advantage of this approach is that the 3D model becomes a front end for providing access to the full scope of knowledge collected on the resource at many points in time. This will provide many different ways of viewing the information and help to promote the greatest number of interpretations. Arguably, the user can learn a great deal more about a specific building or style of architecture if they are able to interact with the material in ways that are not possible on site or via traditional print media. We are carrying

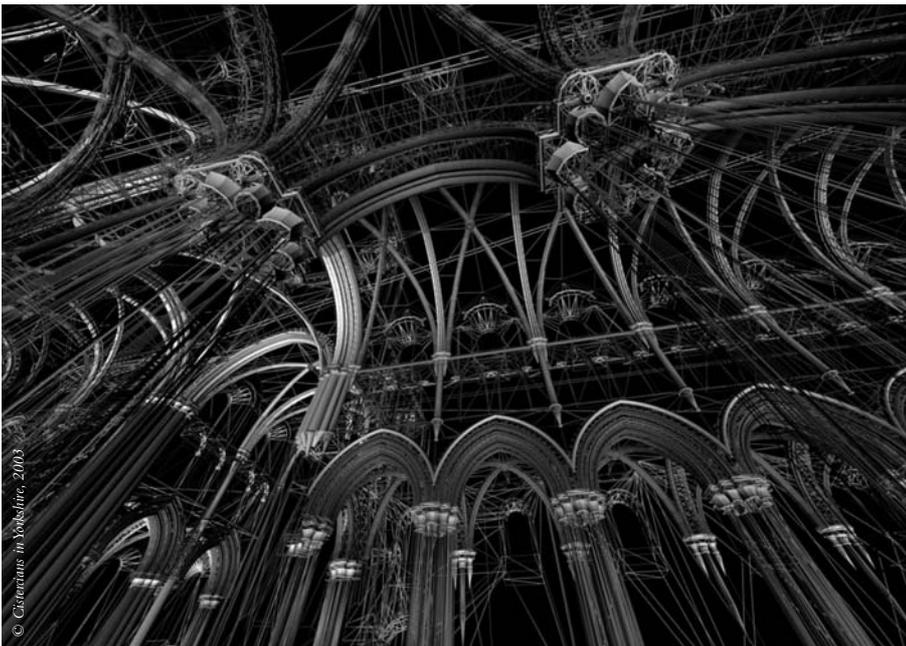


Complex arch moulding objects from Rievaulx Abbey





Summary of the stages involved in the reconstruction process: a section from the transepts at Roche abbey; the recording of the section using photogrammetry; the 2D elevation drawings of the section; the 3D wireframe reconstruction of the section; and the 3D solid reconstruction of the section.



Wire frame renderings of the interior of Roche Abbey

out a variety of evaluation studies concerning the most efficient forms of delivery and how best to integrate the models with the learning package.

While working on the Cistercians in Yorkshire project one fact has been highlighted on many occasions – that the accuracy (and subsequent academic value) of the reconstructions is entirely dependent on the quality of the source materials available. These VR models have the potential to generate a level of detail that is rarely available in the drawings or plans of the site, and there is a tendency for these visual representations to be accepted by the users as completely accurate when in fact there are limits to our historical understanding of the site. All of our models are reconstructed from existing site plans, photogrammetric surveys, early photographs and from the surviving fabric. Although we can reconstruct much of the fabric of the building from the fragments that remain, there are sections for which there is no surviving evidence. In these cases the architectural consultant formulates a reconstruction to show the most likely appearance of the building based on analogy with other more complete sites. In order to make the degree of hypothetical reconstruction involved explicit at all stages of the project cycle, it is necessary to provide a clear distinction between all the original and restored areas of the structure. In order not to distract from the overall coherence of the models we have decided



that the 2D plans and photogrammetrics, which are themselves accessible from the models, are the most efficient place to make this distinction.

As well as myself, the project team includes the director (Dr Sarah Foot, a medieval historian employed in the Sheffield History department), the project manager (Nigel Williamson, who has a background in the use of IT in the humanities and works for Corporate Information and Computing Services at the University of Sheffield), the historical researcher (Dr Julie Kerr, whose PhD was in medieval monasticism), and the archaeological consultant (Stuart Harrison, an independent scholar, recognised as the world expert on northern British Cistercian architecture). I currently work between six floors of architecture students and I am struck by the fact that they spend a lot of effort creating physical models

which generally serve only one purpose before they are put into storage or thrown away. As a result of 3D printers⁴² becoming more affordable it will be possible to take advantage of common modelling software which allows you to make slight changes to designs without having to start from scratch. I am personally very interested in recycling information, i.e. how do we re-arrange humanities data that we already possess in order to allow new forms of analysis to proliferate? Ultimately, if the entire archive of the Cistercians project could be made available then any or all of the huge range of research materials could be analysed and, where necessary, reused to provide updated or modified models. This would significantly extend the lifespan and usefulness of the project and would reaffirm one of our central aims: to allow users to create multiple interpretations from multiple angles, aided, not restricted, by the architectural VR reconstructions. [BACK TO PAGE 1](#)

CULTURAL HERITAGE EVENTS

DigiCULT's Website provides a full calendar of events relevant to the cultural heritage sector from conferences to training workshops. Events taking place in January and February 2004 range from annual conferences on standards and technologies to events on multimedia modelling, collaborative systems and mobile data management, topics which are all covered in the upcoming *DigiCULT Technology Watch Report 2*. [BACK TO PAGE 1](#)

For more details of cultural heritage events and to view the entire list, please visit: <http://www.digicult.info/pages/events.php>

NEWS FROM DIGICULT'S REGIONAL CORRESPONDENTS

DELOS NETWORK OF EXCELLENCE ON DIGITAL LIBRARIES REPORT

'Invest to Save' – the Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation (2003) – is now available from <http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalarchiving/Digitalarchiving.pdf> and presents the conclusions of the International Forum's discussion of strategies for digital preservation.

BULGARIA DIGICULT REGIONAL CORRESPONDENT: MILENA DOBREVA



DIGITAL PRESERVATION OF CULTURAL HERITAGE SYMPOSIUM

The Digital Preservation of Cultural Heritage symposium was held on the 16 and 17 September 2003 within the International Congress of MASSEE (Mathematical Society of South-Eastern

Europe) in Borovets, Bulgaria. The symposium attracted some 30 participants from Bulgaria, Greece, Denmark, Ireland, Romania, Ukraine, and Serbia and Montenegro. Nineteen papers were presented, grouped in several tracks:

- General trends in digitisation of cultural heritage
- Digitisation in Libraries and Archives
- Technology and Tools
- Education: historical issues, preservation and accessibility aspects.

The papers will be published in a special issue of the *Review of the National Centre for Digitisation* (<http://www.ncd.matf.bg.ac.yu/>) based at the University of Belgrade. The issue will appear early in 2004.

42 3D printing will be examined in detail in the next issue of *DigiCULT.Info*.



Within the Technology and Tools session, a presentation of the *DigiCULT Technology Watch Report 1: New Technologies for the Cultural and Scientific Heritage Sector* (http://www.digicult.info/downloads/twr2003_01_low.pdf) was delivered by Milena Dobrova and Nikola Ikononov. One of the symposium discussions raised the issues of digitisation of scientific and cultural heritage, where Professor Jonathan Borwein (from the Centre for Experimental and Constructive Mathematics, Department of Mathematics, at Simon Fraser University, Canada) and Professor Julian Revalski (Institute of Mathematics and Informatics) presented their experience in digitisation of mathematical heritage and the Digital Mathematics Library project (<http://www.library.cornell.edu/dmlib/index.html>).

The final discussion, entitled 'Does South-Eastern Europe need a network in digitisation of Cultural Heritage?' initiated the creation of the South-Eastern European Network for Digitisation of Scientific and Cultural Heritage; a network which will be developed in the near future and which issued its first document, *The Borovets Declaration*. Participants agreed that organisations in the region face common problems and share common scientific and cultural heritage. The knowledge and experience of individual institutions should be shared. It is very important to take measures to increase communication and the exchange of technological expertise, standards and practical skills within the region. Access to the experience of colleagues outside the region would also help. The participants agreed that additional effort should be made to respond to the Action Plan envisaged in the Lund Principles of the European Union (http://www.cordis.lu/ist/ka3/digicult/lund_principles.htm).

The following issues were defined as of crucial importance:

- To mobilise the human and material resources existing in the region.
- To disseminate scientific information as well as the results of research.
- To facilitate communication between centres having similar scientific interest.

In many cases, institutions from the cultural and scientific heritage sectors still have to be made aware of the need for digitisation work and the challenges in preservation and access. The network will seek a mechanism to support a wide range of initiatives to overcome these challenges. South Eastern European countries, including a number of associated states, risk becoming casualties of the 'digital divide' not because of lack of access to the world, but because their heritage is not accessible in the electronic space.

The next Meeting will be held in Belgrade in late May or early June 2004. The constitution of a body representing the network will be presented there for agreement.

CZECH REPUBLIC DIGICULT REGIONAL CORRESPONDENT: ADOLF KNOLL



CZECH NATIONAL DIGITISATION PROGRAMMES

Almost thirty important Czech cultural institutions have already taken part in two national digitisation programmes coordinated by the National Library of the Czech Republic (<http://www.nkp.cz/altnkeng.htm>). The calls for proposals

(launched yearly) offer institutions grants to cover up to 70% of the digitisation cost of their projects. The variation in funding levels from one year to another continues to be a problem; nevertheless, the virtual completion of dispersed collections is becoming reality. At the end of 2003, around 550,000 high-quality digitised pages of manuscripts and 1.2 million pages of periodicals will be available. Not only documents from big library collections, but rare items from regional museums and castle or monastery libraries are being digitised.

The first programme originated in 1992-1993 when the first pilot CD-ROM for the UNESCO Memory of the World programme was prepared and published in Prague. Routine production started in 1996 when the digitisation centre was built in the National Library. The centre has since substantially expanded, thanks to a long-term cooperation with AIP Beroun Ltd (<http://www.aipberoun.cz/english/index.asp>) who are considered to be a co-founder of the programme, now called Memoria (formerly known as Memoriae Mundi Series Bohemica, http://www.nkp.cz/o_knihovnach/English/LPIS6.htm). Under this programme, old manuscripts are digitised, supplemented by rare printed books and historical maps.

The second programme, Kramerius (http://www.nkp.cz/o_knihovnach/English/LPIS7.htm), started routine work in 2000. It is oriented to the digitisation of the preservation microfilm produced in order to safeguard acid-paper library materials, mostly older periodicals.

The data from both programmes are stored and made available thanks to funding from the third national programme, Digital Library (http://www.nkp.cz/o_knihovnach/English/LPIS4.htm), concerning long-term storage of, and access to, digital documents.



The current average annual production is roughly 100,000 pages of manuscripts and 400,000 pages of periodicals. The data format used for storage is now exclusively JPEG with varying compression qualities – 256 shades of grey in the case of periodicals. Since routine production began, a lot of attention was paid to the metadata container comprising the whole digitised document. The lack of available good practice at that time led to the creation of a proprietary SGML-based standard, called DOBM, in 1996. The format was in fact an enlargement of the HTML DTD with elements enabling the description of contents of objects, a rudimentary solution, now better expressed by XML. The DOBM files were HTML files containing formatting in parallel with internal mark-up of rich content elements. The format was very flexible and enabled the needed mark-up on all structural levels of digitised documents including pages, annual volumes, periodical numbers and even articles.

Nowadays, new XML-based formats are available. The move to XML was fostered not so much on the availability and universality of this platform, but rather on changes in content description rules since 1996, especially the appearance of the MASTER DTD (<http://www.tei-c.org.uk/Master/Reference/>) for bibliographic description of manuscripts (in the definition of which we also participated).

New XML Document Type Definitions have been prepared for manuscripts, as well as other types of documents. Today, metadata migration into new XML containers is coming to an end for manuscripts and is in full swing for digitised periodicals. We used the redefinition of the metadata containers to facilitate a large qualitative improvement of metadata structures on all levels of digitised documents. The new XML DTDs are available from our digitisation server

(<http://digit.nkp.cz>) for manuscripts and old printed books, periodicals and museum objects. The DTD for digitised monographs is becoming available and the DTD for sound documents is in preparation.

The digitised documents are archived on compact discs (Memoria programme) and in a robotic library on magnetic tapes (Kramerius programme), and fast access services to all of them will soon be available via special applications.

Manuscripts and similar materials will be accessed through the Memoria database (<http://www.manuscriptorium.com>). We intend that the database will become a shared catalogue for historical collections. The bibliographic description of its records should be in MASTER format, while from the beginning of 2004 UNIMARC records will also be accepted, as a tool to support conversion between these two formats will have been developed. Interest in uploading records into this database is international: the first international contract was signed with the University of Wroclaw Library, Poland. Other institutions in Poland, Slovakia and other countries have also indicated a wish to join. The database software is based on the reliable Czech National Bibliography Tornado application, developed and maintained by AIP Beroun Ltd.⁴³ Tornado has been extended to operate on MASTER records and to handle the complex metadata structures of the complete digitised documents. Furthermore, additional tools for work with images have been added. The access images will be optimised JPEGs in preview, low and normal quality, and 1-bit (black and white) GIF format. The normal quality JPEG is carefully watermarked with visible letters on the image background. JPEG 2000 (JP2) was considered for high-quality access but, owing to too visible background structure cleaning at certain compression ratios, its application was discontinued. Free access will be possi-

ble to all metadata as well as to preview-quality images. Access to higher resolution images requires a licence. For digitised maps, the MrSID technology will be implemented by the end of 2003.

The periodicals (and other types of documents in the future) will be made available in the Kramerius application, which is currently being developed and tested by the Czech Qbizm Corporation (<http://www.qbizm.com/>). Activities include migration of metadata as well as the optimisation of images for access. In this case, the access image format is DjVu (<http://www.djvuzone.org/>), into which all the images are being converted. The application is expected to be launched for users by the end of 2003. All the copy-right-free titles (from the eighteenth century up to 1880) will be accessible free of charge. More recent titles will be accessible on site, where the originals are stored.

These two access systems will enable more concentration on digital archival storage that will become physically independent. The archival system applied today is the Czech AIP Safe document system, product of AIP Safe Ltd (<http://www.aipsafe.com/>).

Our digitisation programmes have grown on the basis of substantial research and development activities. They concern both metadata and data areas including large tests of new emerging solutions. Our long-term goal is to create a virtual research environment based on historical documents. For this, other types of documents and data objects will be added as well as TEI-processed texts of selected manuscripts.

Up-to-date information about Czech digitisation activities is available from the <http://digit.nkp.cz> server in English and Czech.

⁴³ User Manual can be found at http://www.aipberoun.cz/docs/CNB/manual_en.pdf.

FRANCE
DigiCULT REGIONAL CORRESPONDENT:
MURIEL FOULONNEAU



NEW CULTURAL PORTAL:
WWW.CULTURE.FR

The French minister of culture, Jean-Jacques Aillagon, recently launched the new cultural portal <http://www.culture.fr> initiated by the French Ministry of Culture, with the aim of supporting and strengthening the dynamic aspect of culture on the Internet.

He declared: 'When we talk about culture on the Internet, we too often think of the illegal downloading of material. But more importantly, culture on the Internet means online museums, digitised works, virtual tours, documentation, information for visiting, seeing, listening, reading, making reservations and creating. [Culture] means all the Websites produced by institutions (and I am keen for all those attached to the Ministry to be particularly active in this area), amateurs, associations and businesses... I have set a very simple objective for the culture.fr portal: giving a new, decisive boost to culture on the Internet by increasing the number of visitors to high-quality cultural Websites.'

'So the culture.fr portal does not directly create new content: it provides support, visibility and recognition to all the sites that make up the quality and dynamism of the culture Web. We cannot list everything, of course: that's the job of the search engines, but the culture.fr portal, a more qualitative affair, aims to present sites, and give a chance to lesser known sites which

will benefit from an address like culture.fr, from all the information that goes with it and from sites that are already popular.'

M. Aillagon added that he wished to put up 'a simple Website that is easy to use for the many people in France who are not yet expert Net surfers, but are beginning to explore the wealth at their fingertips. A site that offers them direct, rapid access to culture, so that they can see what's on where they live, find a particular museum or library, and discover all the rich and fascinating Internet sites around.'

As far as the organisation of the portal is concerned, he said this Website was working 'like the Web itself, in a flexible, decentralised way. Today, several hundred people already contribute and update content, using material from their own sites or initiatives.'

The portal is based on an open-source software, Zope⁴⁴ and CPS (collaborative portal server) developed by Nuxeo (<http://www.nuxeo.org/cps>). Parts of the content of culture.fr are available in English and Spanish.

RÉSONANCES – INTERNATIONAL CONVENTION ON TECHNOLOGIES FOR MUSIC
OLIVIER LESCURIEUX, IRCAM

For the second year, and with a renewed affluence of music professionals, scientists and visitors, Ircam-Centre Pompidou⁴⁵ organised and hosted Résonances, which has become a major European event for those involved in the future of sound and music.

The evolution of the music industry towards concentration and globalisation, the integration of its manifold functions within the computer, the emergence of generalised Internet access as well as the obvious but powerful merging of creation

and audio tools, all these trends are defining the framework of a new technical music system, explored each year at Résonances. This continuously changing horizon involves:

- a renewal of organology and musicology concepts by reactivating ancestral musical difficulties related to writing and composition techniques, analysis, representation, listening, instrumental conception, training, interpretation and even rehearsal;
- and the increased merging of electronic fabrication with digital instruments widely used in the artistic field – as if the organological question would finally have to be extended to the whole of the trans-artistic field.

The instrument was already the central issue at the first Résonances in October 2002, which focused on two different topics: 'new ways of listening' and 'new instruments'.

Résonances 2003, which brought together 6,000 visitors, built on brainstorming on new instruments for synthesis and control, and expanded it to innovative systems for live performance. It focused on issues such as the positioning of instruments in space (particularly sound spatialisation), score following technologies, motion capture systems for installations and choreography, real-time sound/image interaction, as well as new perspectives related to real-time voice processing of actors and singers. The public discovered the extraordinary variety of 'new instruments' designed for artistic expression in live performances.

The second theme of Résonances 2003 was Tools for Music Analysis. This major trend emerging from the fantastic processing power of the digital era is fundamentally transforming personal appropriation of the musical phenomenon, from musicians to listeners. These issues were explored in an international conference on

44 Cf. *DigiCULT.Info* article on Zope, also in this issue. [LINK](#)

45 Cf. *DigiCULT.Info* article on the Ircam Sound Archive, also in this issue. [LINK](#)

musical analysis and listening based on two perspectives (electro-acoustic music and the Set Theory).

You can find the complete programme of Résonances 2002 and 2003, including biographies and abstracts, on <http://resonances.ircam.fr>. Provisional themes for 2004 are Improvisation and Interpretation, and planned events already include:

- Computer Music Days (JIM 2004)
- Symposium on interpretation analysis
- International workshop on Improvisation
- International conference on Sound Design
- 10th anniversary of the Electronic Music Foundation.

Résonances 2004 will run from 13-22 October.

GREECE DigiCULT REGIONAL CORRESPONDENT: EFTHIMIOS MAVRIKAS



PRACTICE OF KNOWLEDGE SHARING' WORKSHOP

The CIDOC CRM Special Interest Group (<http://cidoc.ics.forth.gr/>) and the Centre for Cultural Informatics at ICS-FORTH (<http://www.ics.forth.gr/isl/ci.html>) have announced a three-day International Workshop designed to bring together researchers and practitioners in information provision and exchange, with the goal of presenting and discussing knowledge sharing among museums, libraries and archives, and information

integration within cultural institutions. The Workshop, entitled 'Practice of Knowledge Sharing', will be held in Heraklion, Crete, on 20-22 April 2004. The deadline for submission of papers, presentations and position statements is 30 January 2004. Interested participants are invited to attend the CIDOC CRM Special Interest Group working meeting that will take place on the third day of the Workshop.

The Centre for Cultural Informatics is part of the Information Systems Laboratory at the Foundation for Research and Technology Hellas (FORTH). The Centre pursues a comprehensive, cross-disciplinary approach to supporting the entire lifecycle of cultural information and documentation procedures for the benefit of the preservation, study and promotion of cultural heritage.

INTRODUCTION TO KEY PLAYERS AND PROJECTS

The Informatics and Telematics Institute (ITI, <http://www.iti.gr/>) and the Centre of Research and Technology Hellas (CERTH) are key participants in REACH: New Forms of Distributed Organisation and Access to Cultural Heritage, a project funded by the General Secretariat for Research and Technology (GSRT, <http://www.gsrt.gr/>). The objective of REACH is the development of a system of unified access to, and management of, information concerning the Hellenic Cultural Heritage. This system will interconnect all the Greek databases of cultural content, offer a central point of access to existing, scattered collections, and promote the worldwide study of the Hellenic Cultural Heritage. REACH complements the extensive nationwide effort of digitisation of collections under way in the CULTURE Sector of National Priority.

The Research Academic Computer Technology Institute (CTI, <http://www.cti.gr/>) Research Unit 2 and

the University of Patras are leading the IST project BALKANET: Design and Development of a Multilingual Balkan WordNet. BALKANET (<http://www.ceid.upatras.gr/Balkanet/>) will develop a multilingual lexical database comprising an individual WordNet for each Balkan language, and attempt to represent the semantic relations between words in each language, linking them together to form an online multilingual semantic network. Semantic relations will be classified in each individual WordNet according to a shared ontology. BALKANET is an excellent opportunity to explore less-studied Balkan languages, and to combine and compare them cross-linguistically.

EXODUS S.A. is the coordinator of the eContent project M-Guide: Cultural Location Based Information Services. M-Guide (<http://www.exodus.gr/m-guide/>) will develop and deploy a system for providing location-based information services for the cities of Athens (GR) and Turku (FI). Information content will consist mainly of tourist information on specific monuments and sights of downtown areas, neighbouring information and public services hot spots. M-Guide services will be provided using existing GSM +/2.5 G mobile technology.

The Foundation of the Hellenic World (FHW, <http://www.fhw.gr/>) participates in the IST thematic network **art-nouveau**: on the Transition to the Digital Era of Arts and Culture. The network has been established as a platform for information exchange and discussion on opportunities for advanced information technologies in the artistic and cultural application domains. The FHW contributes to the network its wide experience in the use of state-of-the-art information and computer technology in the research, awareness and understanding of Hellenic history and culture.



The Prisma Centre for Development Studies (<http://www.prismanet.gr/>) coordinates the Culture2000 project **ACRINET**: European Acritic Heritage Network. Since its inception in 2001, ACRINET has been investigating the acritic tradition across Europe, establishing its symbolic and social meanings, and collecting material from archives and popular culture in the form of manuscripts, publications and films to represent the acritic heritage. ACRINET will make this material available to a wide public by creating a virtual museum.

The Cultural Informatics Laboratory (CILab) at the Department of Cultural Technology and Communication (DCTC, <http://www.aegean.gr/culturaltec/>), University of the Aegean, initiated the Socrates-Minerva project **WebDANCE** using Advanced E-Learning Tools. WebDANCE (<http://www.aegean.gr/culturaltec/webdance/>) promotes the use of new technologies in dance education and, in particular, the use of interactive multimedia technologies (3D animation, MPEG-7 annotated video) in traditional dance information representation, and the use of the Internet as a dance-learning medium. To this end, WebDANCE will develop a platform that can be used both as a Web-based resource for dance teachers and as a Web-based explorative learning environment for dance students in formal (school) and informal contexts.

CONFERENCE UPDATES

The Greek Computer Society (EPY), in cooperation with the Aristotle University of Thessaloniki, the University of Macedonia and the Technological and Educational Institution of Thessaloniki, organised the *9th Panhellenic Conference in Informatics* (PCI 2003, <http://epy9.csd.auth.gr/>), the *1st Balkan Conference in Informatics* (BCI), and the *MultiMine* 'Knowledge from Multimedia' parallel workshop (<http://www.multimine.gr/>).

Conference events took place at University of Macedonia facilities in Thessaloniki in November 2003.

The University of the Aegean, Department of Cultural Technology and Communication (DCTC), is organising the *2nd International Conference of Museology* 'Technology for the Cultural Heritage: Management – Education – Communication' (<http://www.aegean.gr/culturaltec/museum/>), under the auspices of the Ministry of Culture and the Ministry of the Aegean. The conference will be held at University of the Aegean facilities in Mytilene, Lesvos, from 28 June to 2 July 2004.

FOCUS ON EDUCATION:

Graduate studies programmes currently offered include:

- MSc in Cultural Informatics – PhD (<http://www.aegean.gr/culturaltec/msc/>) at the University of the Aegean, Department of Cultural Technology and Communication;
- MSc in Cultural Policy, Management and Communication – PhD (<http://panteion.criticalpublics.com/cmcc/pages/postgraduate/>) at Panteion University, Department of Communication, Media and Culture;
- MSc in Cultural Information Systems and Cultural Heritage Management at the University of Crete, History and Archaeology Department (<http://www.history-archaeology.uoc.gr/>) – Computer Science Department (<http://www.csd.uoc.gr/>);
- MSc in Language Technologies – PhD ('Technoglossia', <http://www.ilspr/tecnoglossia/>) at the Institute for Language and Speech Processing – University of Athens, Department of Linguistics – National Technical University of Athens, School of Electrical and Computer Engineering (<http://www.ece.ntua.gr/>).

ITALY
DIGICULT REGIONAL CORRESPONDENT:
ANNA MARIA TAMMARO



EUROESCO CONFERENCE REPORT:⁴⁶ DR MILENA DOBREVA

In early September, from a picturesque mountain in Tuscany, the international centre Il Ciocco hosted the European research conference 'Philological Disciplines and Digital Technology – Computational Philology: Tradition versus Innovation'. The event was organised with the support of the EUROESCO conference programme of the European Science Foundation, the French CNRS, and the Regione Toscana (Tuscany Regional Government). It brought together about 70 experienced and young researchers from 22 countries. This created an exceptionally good mixture of experiences.

Papers examined digital technology for conservation, restoration, dissemination and management of library materials; the relationship between the digital documents and linguistic/philological research; philology and bibliographical databases; critical editing and hypertextual and hypermedia tools; and the philology of non-alphabetical sources (e.g. demotic alphabet, music), philology and archaeological research.

The diversity of questions that should be addressed in the years to come were summarised by the Euroconference Chair, Dr Andrea Bozzi from the Institute of Computational Linguistics in Pisa, Italy, in his opening address:

⁴⁶ Based partly on the presentations made at the European Research Conference (EUROESCO) on 'Philological Disciplines and Digital Technology – Computational Philology: Tradition versus Innovation' (Il Ciocco, Italy, 6–11 September 2003) organised by the European Science Foundation. This information is the sole responsibility of the author and does not reflect the ESF or Community's opinion. The ESF and the Community are not responsible for any use that might be made of data appearing in this publication.



‘In my opinion, today computational philologists (in the strict sense of the term) have superior technological resources available compared with their colleagues working during the pioneering phase. The resources I refer to include digital technology, computer programs and in particular interconnectivity offered by the Internet.

We can only hope to have contributed to the development and advancement of a multi-secular discipline if, on one hand, we get rid of the usual prejudices that go hand in hand with novelty especially if it is closely tied with technological development and, on the other hand, if approaches and strategies that are truly considered useful for third millennium philology are clearly indicated. Therefore this means that we must answer a series of questions, including:

- To what point is digital image reproduction capable of offering an alternative reading to that of the original or facsimile editions?
- To what point can critical edition information recorded in a database contribute to facilitating *constitutio textus*?
- What sectors, traditionally considered part of the philological discipline (epigraphy, papyrology, medieval philology, ancient printed text philology, musical philology, genetic criticism, etc.), can effectively be enhanced by computer elaboration techniques?
- In what situations can the electronic edition be considered preferable to the paper edition and what are the situations where an electronic edition *must* substitute a paper one?
- How can Web-based textual archives be of effective assistance to a critical editor?
- How can hypertextual instruments be used in philological disciplines for research and/or didactics?
- How can clustering techniques and visualisation of information assist an

editor to evaluate the relationships of traditional hand-written text and to represent it using bi- or tri-dimensional graphic technologies?

- How can linguistic analysis (morphology, lexicon, etc.) together with the assistance of computational systems (morphological and morpho-syntactic analysers, lemmatisers and lexical databases, etc.) effectively contribute to the study of text?’

These and other questions, that could be raised by any of us, are in fact a stimulating research path for a scholar who has, on the one hand, consolidated tradition and methodology and, on the other, a series of surprisingly efficient technological possibilities.

One common concern, especially for the younger participants in the forum, was how they could receive systematic and regular training in this field where frameworks change at an incredible pace. Another important question raised is what actually shapes this community of scholars working in the field of digital philology. The current diversity of problems and approaches should somehow form clearer frameworks in the community, and such forums have the important role of supporting this process.

The excellent work of the conference chair, Andrea Bozzi from ILC (Pisa), and Vice Chair, Jean-Louis Lebrave from ITEM (Paris), and the financial support of the conference sponsors made possible the balanced participation of experienced and young researchers with various interests in the field combined with specialists from associated countries of the EC. This made of the conference an inspiring forum, which will hopefully be the first in a sequence of events. Let us express our wish that this idea will find further support and will grow as a regular meeting place for the community working on

applications of digital technologies to philological disciplines, possibly accompanied by training courses for the beginners in the field. Without any doubt the event increased the feeling of attendees that they belong together and face similar problems. It also gave practical and working answers to some of the questions raised, which will reach a larger group of scholars through a special volume of *Linguistica Computazionale*, a publication of ILC in Pisa, where some of the subjects discussed at the Euro conference could be developed.

CONFERENCE TO PREPARE NEW INFORMATION PROFESSIONALS

First International Conference of employers and LIS school teachers (University of Parma, 24-25 November 2003)

In an international context, this First International Conference collected employers’, staff’s and students’ perceptions and expectations of education and training needs for the digital environment.

It aimed to reflect upon and share ideas and best practices related to preparing information professionals for the hybrid digital library and in an international labour market. How can employers afford the re-qualification and continuing education needs of information professionals in the digital environment? How are Library & Information Science (LIS) schools, after the Bologna Declaration, planning for better employability of new information professionals? What are the best practices of LIS schools worldwide to prepare them in the digital environment?

POLAND

DigiCULT REGIONAL CORRESPONDENT:
MARIA SLIWINSKA



POLISH INTERNET LIBRARY

The Polish Internet Library (<http://www.pbi.edu.pl>), established in 2002, contains about 260 scanned books and since its opening over seven million visits have been recorded. An average visit takes about half an hour. It is planned to develop the collection, making available over 6,000 items. At the end of 2008 literature without copyright, handbooks, some scores and cartographies should be available from the Internet Library as well as graphics, photographs and periodicals. A list of currently available titles can be found at: <http://www.pbi.edu.pl/lista.php>.

INFOBIBNET

Polish public library directors have formed the INFOBIBnet consortium in order to prepare public libraries for their new roles and include them and their users in the global information society.

INFOBIBnet will work on:

- the introduction of compatible library management systems and Internet connection at all public libraries;
- making Internet access points available for local users;
- contributing to the Polish information resources;
- the creation of new information services for users;
- active participation in government and local government information systems.

These tasks will be realised in cooperation with public libraries that (following the academic libraries' example) have established a Standing Conference and Committees.

INTERNATIONAL CONFERENCE

The annual IATUL Conference 2004 devoted to 'Library Management in Changing Environment' will be held in Krakow (Poland) from 30 May to 3 June 2004. The conference sessions have been subdivided into four themes:

- Elements of Library Management
- Traditional Collection and e-Resources – Policy Dilemma
- Public Relations: User-oriented Services
- Regional Library Cooperation.

More information can be found at: <http://www.biblos.pk.edu.pl/IATUL/programme.php3>.

SERBIA AND MONTENEGRO

DigiCULT REGIONAL CORRESPONDENT:
ZORAN OGNJANOVIC



NATIONAL CENTRE FOR DIGITISATION

Numerous institutions from Serbia are in possession of a large number of scientific journals and rare books, archive documents, textual and image data, as well as the other cultural and scientific values. Recently, important efforts have been made to digitise, preserve, describe and index the cultural and scientific heritage. All these attempts are interesting and have concrete outputs, but on the other hand they only last as long as they are financially supported, and in almost every new project

people start from the beginning and find their own ways of performing digitisation. This results in ad hoc approaches, incompatibilities, no clear notion of future use, and wasted time and money.

The idea behind the Nacionalni Centar za Digitalizaciju (National Centre for Digitisation, NCD, <http://www.ncd.matf.bg.ac.yu>) is to overcome these shortcomings by organising continuous and systematic work. At present, the consortium includes the Mathematical Institute and Archaeological Institute of the Serbian Academy of Science and Art, Mathematical Faculty Belgrade, National Library of Serbia, National Museum Belgrade, Archive of Republic Serbia, Serbian Institute for Monument Protection and Yugoslav Film Archive. Representatives of these leading cultural and research institutions form the Main Committee of the NCD.

The main aims of the cooperation are:

- coordination of efforts of institutions involved in cultural and scientific heritage digitisation;
- establishing and promoting a national strategy for cultural and scientific heritage digitisation;
- exploration, adaptation and implementation of international standards and protocols for cultural and scientific heritage digitisation and preservation at the national level; development of new standards in areas where they do not yet exist;
- launching cultural and scientific heritage digitisation and making plans for a possible migration process to new formats and technologies for already digitised data.

These goals should be enabled through the activities of the NCD. Some of the activities are:

- forming expert groups for specific fields of interest (groups for museums,



- libraries, archives, monument protection, standardisation, data protection, etc.);
- establishing, at the national level, a basic set of standards in the area of digitisation of cultural and scientific heritage;
- developing models of the appropriate distributed information systems and specific software for the realisation of digitisation standards;
- connecting cultural and research institutions involved in the digitisation project into a particular Internet subnet and developing a network of information and WWW servers;
- securing long-term storage and maintenance of digitised heritage;
- organising continuous, persistent and systematic work on cultural and scientific heritage digitisation;
- training collaborators from cultural and scientific institutions;
- collaboration with similar institutions from abroad.

The activities illustrate the multidisciplinary character of the NCD. For example, the issue of standards in the area of digitisation of cultural and scientific heritage concerns data and metadata formats and structures, access protocols, rights management, thesauri, search procedures,

and the tools for storage, processing and presentation of the digitised data. During the digitisation process, models of a distributed information system for digitised cultural and scientific heritage and specific software that follows the above standards should be developed. The models should allow dissemination and wide access over the Internet to cultural and scientific values that at present are not easily accessible. These systems should be as compatible as possible. Collaborators from cultural and scientific institutions ought to use the software both as end-users and as experts who define metadata for corresponding objects and enter information into databases.

NCD established a journal, called *Review of the National Centre for Digitisation*, which is published both in electronic and in printed form (<http://www.ncd.matf.bg.ac.yu/?page=publishing&lang=en>). The journal publishes papers (mostly in Serbian) concerning all aspects of digitisation of cultural and scientific heritage. Two national conferences under the name 'New Technologies and Standards: Digitisation of National Heritage' were organised in 2002 and 2003 (www.ncd.matf.bg.ac.yu/?page=conferences&lang=en), presenting local experiences with new technologies, standards and the other fields in the area of digitisation. [BACK TO PAGE 1](#)

DIGICULT REGIONAL CORRESPONDENTS:

Warm thanks go to all of our Regional Correspondents:

Bulgaria	Milena Dobрева
Czech Republic	Adolf Knoll
France	Muriel Foulonneau
Greece	Efthimios Mavrikas
Italy	Anna Maria Tammamo
Poland	Maria Sliwinska
Serbia and Montenegro	Zoran Ognjanovic

DigiCULT are still seeking Regional Correspondents for countries not mentioned above. If you would like to contribute, please contact the *DigiCULT.Info* content editor, Daisy Abbott, at D.Abbott@hatii.arts.gla.ac.uk

SMIL EXPLAINED

CONTINUING ON FROM DIGITAL OBJECT IDENTIFIERS AND HUMAN LANGUAGE TECHNOLOGIES IN *DIGICULT.INFO'S* TECHNOLOGIES EXPLAINED SERIES, DAISY ABBOTT CONSIDERS THE SYNCHRONISED MEDIA INTEGRATION LANGUAGE.

WHAT IS SMIL?

The Synchronised Media Integration Language (SMIL, pronounced 'smile') is a mark-up language which supports the

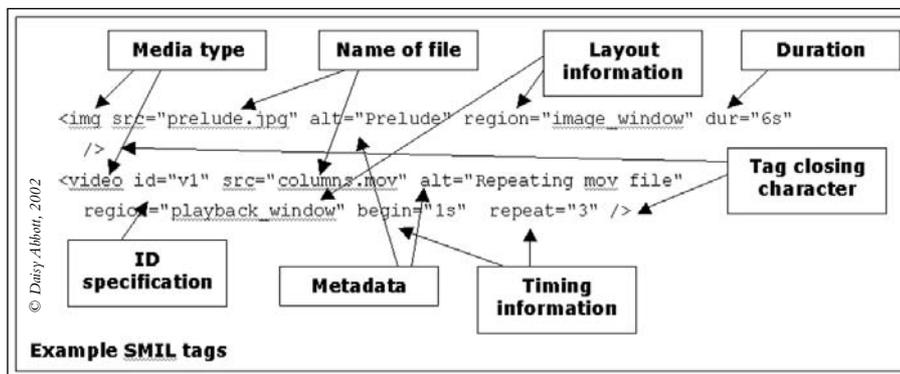
co-ordination of different media displayed on a Website. SMIL can be thought of as similar to HTML but with additional temporal definition, controlling placement of multimedia in both space and time and therefore allowing both the synchronisation and integration of various media.

SMIL was designed as a declarative format for expressing media synchronisation on the Web as an alternative and complementary approach to scripting languages. In March 1997 the first Working Group was established by

W3C, leading to SMIL 1.0 becoming a W3C recommendation in June 1998.

HOW DOES IT WORK?

SMIL, an XML specification, defines a number of elements and attributes useful for controlling presentation, synchronisation and interactivity. Media components accessible by URLs are named (tagged) and their display can be positioned in space and scheduled in time. Positioning can be defined similarly to HTML. Controlling display of media in time is achieved in two main ways: parallel (e.g.



show ImageA at the same time as playing AudioB) or sequential (e.g. play VideoC then show TextD). Synchronisation can be further controlled (again using methods familiar from HTML) by defining times as relative or absolute. An example of using relative and absolute temporal definitions could be: show ImageE three seconds after VideoF starts playing (absolute), stop showing ImageE when VideoF finishes (relative). As parallel and sequential display is defined by tags, these two temporal display types can be nested to create very complex multimedia displays.

A SMIL EXAMPLE:

Multimedia definition in SMIL can be extremely simple; only a very few tags are required to display most media. Two typical pieces of SMIL code pertaining to one media file each are shown below. The first tag displays an image in a previously specified area, `image_window`. The image is shown for 6 seconds. The second tag plays a video in a different area of the screen defined by the region `playback_window`. The video has an intrinsic length, therefore the duration tag can be omitted. However, the `begin` time specified causes the video to wait 1 second before beginning playback. The video will repeat three times. This file is given an id number that can be referenced by later tags and used to specify behaviour such as 'start v2 playing 10 seconds after v1 has finished'.

Exactly how these two files are displayed depends on whether the tags

are defined as sequential or parallel. If contained within a pair of `<seq>` `</seq>` tags, the tags will produce a display as follows: `prelude.jpg` is shown for 6 seconds in the image window, then there will be a second of blankness, followed by the video `columns.mov` looping three times in the video window. However, if the tags are contained within `<par>` `</par>` tags, the image will be displayed as before, but after 1 second the video will also start. Five seconds later the image will disappear, and the video will stop when it has completed three repetitions. The `<par>` and `<seq>` tags can be nested, thereby facilitating any complex synchronisation of media, as long as the individual tags are placed within the correct combination of parallel and sequence tags.⁴⁷

SMIL allows interactive control of multimedia by supporting the definition of buttons such as Play, Stop, FastForward and so on, and includes features of HTML such as linking to separate files. This means that a user can interrupt a presentation in order to follow an interesting link and then resume the display when finished browsing. Recent versions of SMIL have included more elaborate control mechanisms such as slow motion or a random starting point.

HOW CAN SMIL BE USED?

SMIL can be further compared to HTML in that it was designed specifically to be easy to use – a simple presentation can be authored in a few lines of SMIL just as a simple HTML page only

requires a few lines of mark-up. No specialised tools are required to author SMIL; it can be written in a text editor and many free tutorials have been published on the Web making it accessible to new users. To play SMIL presentations, your browser must be SMIL compatible – several free SMIL browsers exist (e.g. X-Smiles (<http://www.x-smiles.org/>), GriNS (<http://www.oratrix.com/GriNS/>), HPAS (<http://www.research.compaq.com/SRC/HPAS/>), and SOJA (<http://www.helio.org/products/>)) and most of the more common browsers now support the language.

SMIL's flexibility means that SMIL presentations can be as varied as Web pages. A simple example of a multimedia presentation could be a slide show: a picture is displayed with an audio commentary (in parallel); once the first commentary is finished, the presentation moves on to the second picture and commentary (in sequence). Another example might be a musical analysis application: 'Peter and the Wolf' plays as an audio file while text and images appear at set points in the piece to identify each animal and to describe the story. The possibilities of cultural heritage applications of synchronised multimedia are huge and SMIL offers a simple, flexible and accessible way to achieve them.

More information and SMIL resources can be found at the Synchronised Multimedia homepage, <http://www.w3.org/AudioVideo/>, at the dedicated Web page JustSMIL <http://smw.internet.com/smil/>, and the upcoming *DigiCULT Technology Watch Report 2*, chapter on the XML family of specifications, due for publication early in 2004. To see the potential of SMIL in action, although not in a cultural heritage context, look at <http://www.people.hbs.edu/lbouthillier/smil/>, or for an easy tutorial, try <http://www.helio.org/products/smil/tutorial/toc.html>.

BACK TO PAGE 1

⁴⁷ Example taken from Daisy Abbott, *Documentary Previsualisation: Multimedia Presentation Languages* (September 2002).

FINDING NAMES WITH THE AUTONOM PARSER



KARINA VAN DALEN-OSKAM &
JORIS VAN ZUNDERT
NETHERLANDS INSTITUTE FOR SCIENTIFIC
INFORMATION SERVICES (NIWI-KNAW)
DEPARTMENT OF DUTCH LINGUISTICS AND
LITERARY STUDIES

INTRODUCTION

One of the research topics of the Department of Dutch Linguistics and Literary Studies of the Netherlands Institute for Scientific Information Services (NIWI-KNAW) is 'names in literary texts'. Names can be studied from different perspectives: from linguistic, literary or onomastic ('name studies') points of view. Areas for investigation include: how authors use names to distinguish between entities (persons, places, etc); which types of names are used and what their frequencies are; how authors make use of names to give their story a certain atmosphere or to describe the characters or places in their work; and how authors differ from each other in their use of names. Central to this kind of research is to consider *all* names in the texts under investigation in the analysis.

Without the help of a name parser this would be a time-consuming and error-prone job, which may explain why this method has not been used much until now. The name parser we have developed and the Web site through which it can be used will make this type of name research a lot easier. We expect it will eventually yield much more concrete and reliable results

than onomastic research of this type has done to date. Here we will describe the first steps we have taken towards a Website in which researchers from all over the world can analyse names and other features in the texts of their own choice.

TECHNICAL BACKGROUND

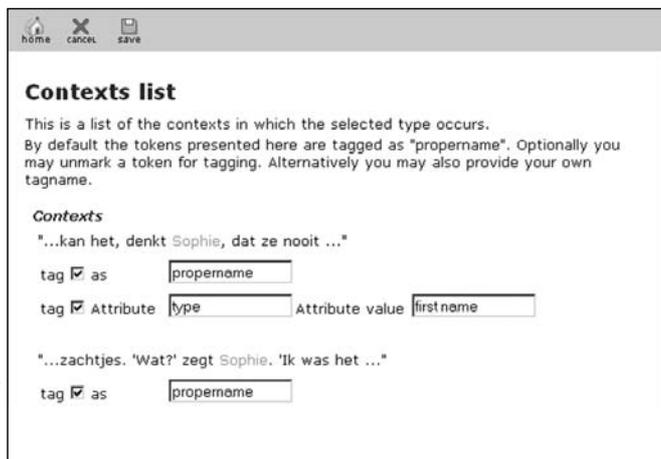
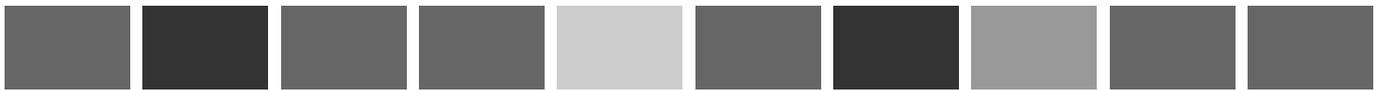
Autonom (<http://autonom.niwi.knaw.nl>) is a Web application that provides a framework for computer-assisted analysis of texts. Its special focus is *literary* texts. At the moment Autonom does not implement any machine learning statistical approaches, because such approaches require large training corpora that may not always be available or appropriate to the literary field in question. Rather we adopted a rule-based (or finite state) shallow parsing techniques as a pragmatic solution. We do, however, intend to strengthen this technique by adding probabilistic feature parsing. We will use the growing number of texts that are researched within the Autonom framework as a training corpus.

First, we developed the name parser. The rule we use within the parser to identify proper names is our knowledge of the fact that in Dutch it is customary to write proper names with an initial capital,

combined with our expectation that, if those words really are names, they will (with a certain probability) not appear in the same text without the first character capitalised. So we expect to find 'Simon' in a text, but not 'simon', but we do expect to find 'Table' as well as 'table'. Therefore, to identify all probable proper names in a text under investigation, we simply return all tokens that have an initial capital but do not appear in the same text uncapitalised. With the Autonom framework we intend to provide (and in part already do provide) an online platform for individual researchers to express such rule-like features or phenomena in texts. The platform will convert formalised textual features described by the researcher into rule-based algorithms for text parsing. Such algorithms may then be applied to any text available to the researcher within a repository.



First page of the novel *De Middelburgsche Avanturier* (1760, author unknown), one of the eighteenth-century novels that will be analysed with [the] help of Autonom.



© Autonomi, 2003

© Autonomi, 2003

In this presentation all tokens by default have been marked as 'proper name', so the researcher only has to unmark the wrongly suggested words.

Although at the moment we only provide an algorithm for parsing proper names, the contours of the framework are beginning to form. A researcher may upload a text file to a personalised and private repository. From the repository the text may then be analysed by applying an algorithm of the researcher's choice. The parsing algorithm returns an XML tree representing the tokens in the text identified as 'probable' proper names. The researcher can approve or reject the suggestions made by the algorithm and annotate (i.e. tag) analysed items. The 'authorised' results are stored in a private analysis repository. From this repository result trees may be regained for future reference, editing, or for rendering a view on the combined result trees and analysed text. For these latter purposes we rely on a 'just in time trees' model (JITT)⁴⁸ that allows us to preserve the original digital text and to store multiple result trees in a concurrent versioning system (CVS). By realising concurrent versioning for analysis results we intend to provide the researcher with the possibility to recursively refine the formal definition (and thereby the algorithm) for the textual phenomena sought within texts. The researcher should in this way be able to reproduce, refine and combine any prior algorithms used for analyses. In addition, we want to provide collaborative means for researchers to share

their algorithms, texts and analysis results. As such, researchers will be able to combine and cross-examine their individual result trees for further cooperative analysis.

HOW DOES AUTONOM WORK?

To analyse the names in a (literary) text a researcher provides scanned and OCR'd text. As the goal is not to present an ideal and 'clean' digital version of the text, the researcher can leave the OCR text uncorrected. The .txt version of the digital text must then be uploaded to the repository; this is very easy to do and all procedures are carefully explained on the Website. To avoid problems with copyright we have taken care to ensure that only the researcher who submits a document will be able to analyse and view it.

The next step is to analyse the uploaded text. The name parser presents the researcher with a list of all types in the text that might be names. In this list, which also shows the number of occurrences, the researcher can view all tokens per type. Each word that is probably a name is presented with some context, so that it is very easy to check whether all of the parser's suggestions are correct.

Clicking the save button fixes the chosen interpretation of the token(s) as a name or not. To check whether the

parser has missed any names the researcher can request a view of the text, in which the tagged names are highlighted.

It will be possible to add names that were not found by the parser by marking them in this text view. This method proved to be efficient: our first test resulted in a percentage of 79.4% accuracy in tokens that the parser suggested were names. There were very few names that were not recognised as such by the parser, only five from a total of 61,152 in the test text.

When the researcher has decided the tagging is sufficient, s/he can turn to the tag view and take the set of tagged types and tokens as input for further analysis. In this respect, little has been done so far. One of the first things we plan to do is to add the possibility for the researcher to group name types by normalised word-forms, for instance to get nominatives and genitives together (e.g. Albert and Albert's). Furthermore, we would like to add results of our deeper knowledge about the names in the text, for instance mark certain names as 'speaking', when quotes from the text under investigation can prove that the author has made use of the etymological meaning of the name to 'build' the character or the place he or she is describing. We also want to be able to get results for the

48 For more information about Just in Time Trees, see <http://www.sbl-site2.org/Extreme2002/>.



TRY OUT THE NAME PARSER

If you would like to try the name parser as it is, please visit <http://autonom.niwi.knaw.nl> and register as a user (this is necessary to keep the texts safe from copyright infringement; registering and making use of all tools is free of charge). We would be interested to hear about your experiences and we welcome any suggestions you might have.

frequencies of different categories of names. This means that first we will have to elaborate on the tag set. Up until now, we have worked with proper names only. For our research we will have to enlarge the set of attributes for this tag: we want to separate first names from surnames, personal names from place names, and distinguish different types of place names from each other, and so on, but keep all of them in the large group of 'proper names'. Each researcher will have to be able to create his or her own tag set and attributes. We are currently developing tools for this.

As far as possible, Autonom has been developed in respect of good practices: we intend Autonom to be open-source software under a GPL licence (as we think all software developed within an academic context should be); we only rely on platform-independent programming (Java), on open standards (i.e. HTML,

XML, XSL, CSS), and on open-source software (e.g. TomCat, Xalan, Dom4J etc); and we use a lightweight development protocol (eXtreme Programming or XP), which we feel is flexible enough for use in academic and research environments.

NEXT STEPS

The first tool we have added supports a 'nearness analysis'. This tool presents a table of all the tagged name types in the text, and all other names that are located between one and ten words away from each instance of this name. The number of names in the 'neighbourhood' is given as well as the mean distance. When, after elaborating on the tag set and the attributes, the results of this nearness analysis are more promising and valuable, we will try out several ways to visualise them.

The name parser is developed for modern Dutch texts. We are current-

ly testing it on other texts and trying it out on older texts as well. The first older texts we are analysing are some eighteenth-century novels. In this time period, authors capitalised not only names, but many nouns as well. We will test whether the different use of capitalisation will significantly change the usability of the parser. If so, we will try to adapt the name parser for eighteenth-century Dutch texts and give the user the choice of different versions of the parser. Autonom can also be used for other languages that are presented with the Roman alphabet, namely for those languages that make use of the same capitalisation rules as Dutch. A lot of languages have the same system and can therefore make use of Autonom. (For modern German, where all nouns are capitalised, Autonom actually functions as a *noun* parser.) We are thinking of developing the interface to enable the researcher to tag a text completely in any language (in the Roman alphabet) and of any time period, thus to tag each word in it with a dictionary entry and with a tag denoting part of speech.

[BACK TO PAGE 1](#)

For more information on Autonom, please visit <http://www.niwi.knaw.nl/en/neerlandistiek/onderzoek/toon>.

THE OBJECT OF LEARNING: VIRTUAL AND PHYSICAL CULTURAL HERITAGE INTERACTIONS

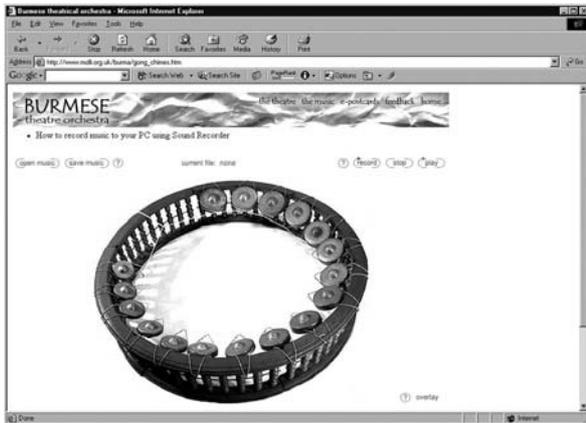
AFTER THE PUBLICATION OF *DIGICULT THEMATIC ISSUE 4: LEARNING OBJECTS FROM CULTURAL AND SCIENTIFIC HERITAGE RESOURCES* ([HTTP://WWW.DIGICULT.INFO/PAGES/PUBLICATIONS.PHP](http://www.digicult.info/pages/publications.php)), DOMINIC PROSSER, FROM THE UNIVERSITY OF EXETER TELEMATICS CENTRE, AND SUSAN EDDISFORD, TIVERTON MUSEUM AND THE TELEMATICS CENTRE, SPOKE TO *DIGICULT.INFO* ABOUT THE THEORY AND PRACTICE OF LEARNING THROUGH VIRTUAL EXHIBITIONS.

At the Telematics Centre, University of Exeter, we have just started our 12th digital heritage education project. A full description of all our projects can be found on our Website at <http://telematics.ex.ac.uk>. Although we cannot describe them all here, exploring the following projects will demonstrate our work:

- **RealCornwall** (<http://www.realcornwall.net/>): in partnership with Cornwall

County Council, the Telematics Centre is creating an open-access Website to stimulate and support involvement with Cornish Culture. It is designed to support the adult literacy core curriculum. (Supported by the New Opportunities Fund).

- **Virtual Victorians** (<http://www.victorians.org.uk/>): working in partner-



Playing the Burmese Gong Chime Circle

ship with Tiverton and Mid Devon Museum Trust, the Centre is researching and developing a Web-based resource built around museum artefacts, newspaper archives and photographs. This resource is designed for use by schools (primary and early secondary), and life-long learners. (Funded by the Department for Education and Employment Museums and Galleries Education Programme and the New Opportunities Fund).

- **EVE** (Everyone's Virtual Exhibition, <http://billdouglas.ex.ac.uk/eve/>): we are creating tools to allow users to create their own virtual teaching resources and exhibitions from the collections of items encompassing the development of popular entertainment housed at the Bill Douglas Centre. EVE is designed for use in Further and Higher Education. (Funded by the Arts and Humanities Research Board).
- **Burma** (<http://www.molli.org.uk/burma/>): based around the Burmese Theatrical Orchestra, or 'saing', part of the reserve collection at the Royal Albert Memorial Museum, Exeter, this project allows users to see and hear the Burmese instruments and express their own musical talent by making a recording of themselves playing the instruments. This project is designed for primary education. (Supported by the re:Source Designation Challenge Fund).

In our work there is a symbiotic relationship between 'virtual visiting' and physical museum-going. We recognise that each has its strengths and limitations, and would argue strongly that the virtual should be designed to promote, complement, extend and enhance the physical experience rather than seek to replace it.

Too many heritage

Websites fall into the trap of simply trying to create a surrogate virtual copy of the physical exhibition, rather than creatively embracing the potential of the ICT medium to provide exciting and meaningful learning experiences, such as being able to record yourself playing on a Burmese Gong Chime Circle, asking a 'real' Virtual Victorian a question via the Web or building your own virtual exhibition.

The relationship between the real and the virtual museum visit appears to be mutually beneficial. Active engagement with the objects in the virtual environment contributes to the users' conceptual understanding of the objects that they see in the real museum, but cannot touch. Inadequacies in the digital objects, such as loss of ability to appreciate and 'read' scale and textural quality, are compensated for by the users' own prior knowledge, experience and beliefs, constructed in the real world and, of course, the opportunity to participate in meaningful, exciting and rewarding virtual interactions.

The well-designed online museum site will provide a meaningful addition to the formal and informal learning experience and enjoyment of the museum visit. It will enable the visitor to explore more deeply the context in which the objects need to be placed in order to understand their meaning. The virtual visit should

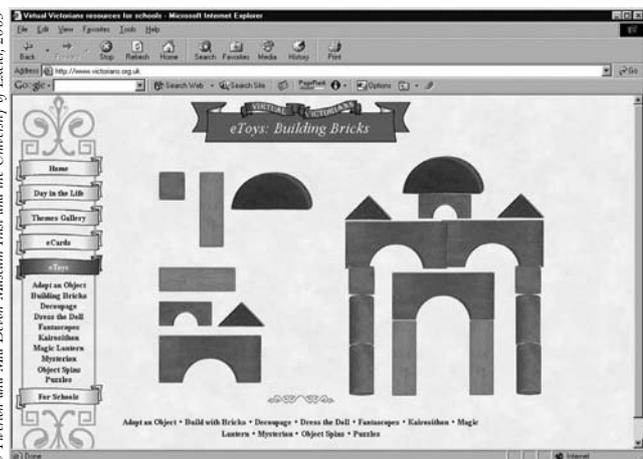
facilitate deeper engagement with objects and thus provide time for reflection and the building of personal meanings and interpretation. This virtual contextualisation could also be made available within the gallery, to provide this additional dimension to the interpretation of artefacts, and to enhance the visitor experience.

We have already gathered evidence through our work that virtual visits to our sites have prompted subsequent physical visits, and vice versa. The findings from Burma and Virtual Victorians show that the virtual objects promoted an appreciation of the value of physical objects and of displaying them in museums, and increased the desire to see them for real.

The virtual museum can help to attract visitors from non-traditional museum audiences. An engaging and enjoyable virtual visit can dispel negative preconceptions of physical museums and their collections and encourage new audiences to become museum visitors.

It goes without saying that well-designed museum Websites can build a worldwide audience and engage them in contextualised and meaningful virtual visits. Virtual Victorians, for example, has had strong take-up in classrooms around the world, including Italy, New Zealand, Australia, Dubai, USA, South Korea and even Scotland!

Each of our projects has a social science research methodology underpinning it. Primarily, this is concerned with the analysis of user data gathered through a range of means including electronic and paper-based feedback forms, analysis of user contributions and interactions prompted through the site such as 'ask a Victorian questions', registrations to build exhibitions, or exchange of e-cards. We also employ a lot of qualitative face-to-face techniques, such as focus groups, interviews and observation.



Using the building bricks e-toy.

At the heart of all our work is a learner-centred approach. We endeavour to make the learning experience personal to each learner. In achieving this we offer various opportunities for the learner to engage with objects, originating contexts and peoples in a meaningful and enjoyable way, either through hands-on virtual activities such as the Virtual Victorians e-toys, through opportunities to interact with other people, through publication of users' own material on the site, by participation in a series of experimental Webcasts or simply through sending your friend an e-card of the Looe Fishermen's Choir.

Central to our work is the creation of the 'value added' experience, rather than merely content and information delivery. The virtual representations of the objects need to be more satisfying and illuminating than mere two-dimensional representations, which cannot replicate the quality of the physical experience of the original artefact. Technology is used both to provide an opportunity to explore the functionality of the artefacts and to add a personal dimension, which extends engagement with the object through person-to-object interaction and person-to-person interaction.

The virtual environment increases learning by allowing the user to explore and manipulate the object and engage with both its originating context and its owners. In Virtual Victorians it is possible to give a magic lantern show, while photographs, archive material and the daily journals of the 'Virtual Victorians' themselves help to place the

objects in a wider context. In Burma, users can watch a shadow puppet play, or create their own, accompanied by music composed and recorded on the Burmese orchestra. It is essential that these activities are enjoyable and provide a rich user experience. Such elements capture the imagination and interest of the user, promote fuller engagement with the subject, and stimulate and aid learning.

Although the projects provide resources for lifelong learning, where appropriate they are carefully tailored to the learning requirements of the English National Curriculum and QCA schemes of work. Not only does this maximise the learning potential of the projects within formal education, our research has gathered evidence to suggest that it also encourages the use of the resources by pupils and parents to support formal learning and facilitates shared family learning.

Samuel Taylor Coleridge argued that our response to drama is characterised by a 'willing suspension of disbelief'⁴⁹ and thus involves the very same ingredient of belief that is essential to everyday emotion. Our work has led us to believe that in certain digital interactive exchanges, children, and adults, are *willing to suspend disbelief* in order to engage with the virtual world. In so

doing, they temporarily suspend their focus on the physical world and enter the virtual world as willing participants, dispelling expectations of the normal conventions and logic of reality, and creating a belief in the virtual.

In order to create and sustain a believable virtual world, the digital learning domain must be sufficiently credible, compelling and indeed enjoyable. Personalisation is an important design element. The opportunity that the virtual world provides for the creation of belief in the Victorian Possett family, for instance, contributes to engagement with the objects from the collection and the development of the key historical skills of enquiry and empathy, and increasing the users' historical understanding.

If the site is fun, as well as educational, it has the chance of allowing users to become immersed and deeply engaged with the cultural materials. The intention is that the users get so caught up in the site that they do not realise they are undertaking learning activities.

We encourage people to reflect on their cultural experience and to share their feelings, beliefs, knowledge, creativity and ideas with others. This is achieved in a number of ways, though each places an emphasis on the 'C' in ICT. For example, in RealCornwall, our users become creators of cultural artefacts through the contribution of memories, tales, sayings and so on. The users then are both consumers and producers of culture in a true learning community. Within the site design we have attempted to blend user contributions seamlessly into the site without differentiation from in-house content. As with most of our sites, learning pathways are not overt, though all the user contribution forms are in fact writing frames designed to support the adult literacy core curriculum at entry levels 1 and 2, and ICT literacy.

49 S. Coleridge, *Biographia Literaria* (London: Rest Fenner, 1817).



We have found, through researching schoolchildren's use of The Burmese Theatre Orchestra project, that parents and children alike derived learning gains through this work. It is the cultural content and the nature of the learning activities that are fresh and interesting to both. Who could resist composing a tune on the Gong Chimes or Burmese clappers online?

PUBLICATIONS

We are publishing our work in a range of national and international journals. The first in the series is: P. Dillon & D. Prosser, 'Educational Transactions in Museum Online Learning Initiatives' in *International Journal on E-Learning*, vol. 2, no. (1), 2003, pp. 4-20.

Details of future publications will be available from <http://telematics.ex.ac.uk/pubs.htm>.

[BACK TO PAGE 1](#)

NEW PROJECT HIGHLIGHTS THE GROWING USE OF DOIs

A major new project will use the Digital Object Identifier system to persistently identify scientific data. The German National Library of Science and Technology (TIB, <http://www.tib.de>) has recently joined the International DOI Foundation (<http://www.doi.org/welcome.html>) and will act as a contact for this project, which will develop DOI implementation for data sets based on the recommendations of a 2002 report by CODATA, the Committee on Data for Science and Technology of the International Council for Science (<http://www.icsu.org>): 'Concept of Citing Scientific Primary Data'.

This pilot will provide the means of registering data sets for long-term scientific reference; initially geo-reference data will be used, but the principle is equally applicable to all scientific data. Each primary data set, persistently identified with a DOI, will have a structured description, facilitating its study, citation and continued use. As well as working towards a more standardised method of storage, the robust nature of DOIs also allows the data sets to be linked with other material, for example, scientific articles that have used the data.

The project leader, Dr Michael Lautenschlager of the World Data Centre for Climate (<http://www.mad.zmaw.de/wdcc/>), said: 'This development will ameliorate current shortcomings in data provision and interdisciplinary use, where data sources may not be widely known and data are archived without context. It will enable citations of data in a standard manner, and facilitate links to more specialised data schemes. The DOI system offers us a proven well-developed system that is already widely deployed and enables us to focus our efforts on the scientific data aspects of the project.'

The data were described by TIB as a 'vital resource' for the scientific community and their inclusion in the TIB's collection of digital sources is an important step in improving access to information in the largest library of science and technology in the world. The project is funded for one year by the Deutsche Forschungsgemeinschaft (<http://www.dfg.de/>), and coordinated by the World Data Center for Climate (WDCC) at the Max-Planck-Institut für Meteorologie in Hamburg.

A recent issue of online technology report Release 1.0 (<http://www.edventure.com/release1/>) is

entitled 'Online Registries: The DNS and Beyond...' and deals with digital object architecture. Release 1.0 is itself powered by the DOI through Content Directions, Inc. (<http://www.contentdirections.com/>).⁵⁰

There are now over 10 million DOIs assigned. 'An introduction to the Digital Object Identifier System' begins on page 20 of *DigiCULT.Info Issue 4* (http://www.digicult.info/downloads/digicult_newsletter_issue4_lowres.pdf). More information can be found in the detailed document by IDF Director Dr Norman Paskin: 'Digital Object Identifiers and Digital Preservation of the Record of Science' at http://www.doi.org/topics/020210_CSTI.pdf.

DIGITAL RESOURCES FOR THE HUMANITIES 2003

DAISY ABBOTT ATTENDED DRH2003 (31 AUGUST – 3 SEPTEMBER 2003) ON BEHALF OF DIGICULT.



Park Campus, Cheltenham, site of DRH 2003.

The Digital Resources for the Humanities conference, held annually, is a major forum for everyone affected by the digitisation of cultural heritage, from the creators of digital resources for access or conservation to the information's or collection's varied users and audiences. DRH2003 was hosted by the School of

⁵⁰ Recent issues also include coverage of issues related to DigiCULT's upcoming *Technology Watch Report*: social software, blogging and virtual community software; and Radio Frequency Identification technology (RFID) which was covered in the chapter on Smart Labels and Smart Tags in *Technology Watch Report 1* (http://www.digicult.info/downloads/twr2003_01_low.pdf).



The busy poster sessions in the exhibition hall.

Humanities at the University of Gloucestershire. Presentations, poster discussions and plenary speeches were held within the beautiful Park Campus, 25 acres of landscaped parkland a short distance from the centre of Cheltenham, which contained the conference venues and accommodation for all participants.

There were over 160 participants in attendance, from across Europe, Australasia and the Americas, over 100 of whom also presented a paper or project. The ages and backgrounds of the attendees varied widely, representing a wide cross-section of the people in the digital cultural heritage community.

The programme was packed with interesting projects and presentations, making it difficult to choose which sessions to attend – although many papers were subsequently discussed among participants around the dinner table, allowing for an informal (and often animated) insight into presentations which may have been missed by individuals. Papers were grouped into linked sessions, allowing the presentation of a particular topic from several different angles. This grouping encouraged a refreshing combination of theoretical, proj-

ect-based and technical papers and gave a well-rounded and detailed view of distinct areas of the sector. Although thematically linked sessions were deliberately organised, the extremely efficient timekeeping of session chairs allowed attendees to pick and choose individual papers within each group. Themes that arose from linked groups of papers were varied and offered a comprehensive view of the digital humanities

sector. Many themes were wide-ranging and some more specific; from the conference as a whole, many pertinent issues emerged, often posing more questions than were answered!

Specific papers on how to digitise the cultural past and present it in a meaningful and accessible way included discussions on maximising learning potential for children, re-presenting historical data (in projects such as 3D Virtual Buildings and the Cromwell in Ireland ‘time-maps’ project), and the digitisation of community, cultural and religious resources. There were also many presentations devoted to issues surrounding digital publishing and the changing of texts and access to them. Papers ranged from technical projects such as using computers to assist in analysis of fictional texts,⁵¹ to presentation of E-Books, to more abstract dialogues on the changing nature of publishing, and a panel of publishers discussed the approaches to and effects of digital publishing as a whole.

Naturally, many of the papers and posters concentrated on digital collections: libraries, archives, museums and brand new resources. Digital archives from the UK National Archive to small projects

devoted to archiving endangered languages were examined, as were many aspects of image archives and visual resources. Structure and content of digital libraries and other collections were debated too, several papers introducing very specific new resources such as the Forced Migration Online Portal (<http://www.forcedmigration.org/>) and the TV Times Digitisation Project. Other themes emerging from the grouped papers were: encoding and mark-up, XML, and data standards; changing scholarly communities; rehousing historical data; places and spaces; new approaches to cultural preservation; and an excellent series of papers and exhibits on resource discovery services for the arts and humanities (Artifact (<http://www.artifact.ac.uk/>), The Arts and Humanities Data Service (<http://www.ahds.ac.uk/>), EDINA (<http://edina.ac.uk/>), Humbul Humanities Hub (<http://www.humbul.ac.uk/>), and the Resource Discovery Network (<http://www.rdn.ac.uk/>)).

The conference plenary speeches reflected the overall themes of the four days’ activities. Meg Bellinger, Associate University Librarian for Integrated Systems and Technical Services, Yale University Library, opened the conference with a speech considering technology related to the creation and use of knowledge objects, and the implications of transformative technologies on preservation of information. She went on to provide an overview of digital repository developments. The two engaging and entertaining speakers who closed the conference, Theodor Nelson (founder of Project Xanadu, the original hypertext project, and Senior Fellow of the McLuhan Institute, Toronto) and Kim Veltman (Scientific Director of the Maastricht McLuhan Institute and founder of E-Culture Net), discussed the past, current and future state of the Web and current and new modes of access to the information it presents.

51 Cf. *DigiCULT.Info* article on Autonom, in this issue. [LINK](#)



© DigiCULT, Daisy Albert, 2003

Sculpture at Park Campus

Overall, DRH2003 justified the excellent reputation this series of conferences has built up over the years. Most material presented was both accessible from a position of little familiarity and challenging enough to engage the interest of the audience well beyond each session. The organisation of both the programme and the social activities was outstanding, creating a friendly atmosphere of intellectual discourse and allowing both new and regular attendees an insight into current theoretical and practical work in the field of Digital Resources for the Humanities. I anticipate that DRH2004, to be hosted by the University of Newcastle, will be a similar success.

[BACK TO PAGE 1](#)

NEW UK LAW FOR PRESERVATION OF ELECTRONIC PUBLICATIONS

At the end of October, a Private Members Bill became law with the result that the UK Legal Deposit Libraries Act 2003 now provides for secondary legislation to ensure that non-print materials

will be deposited in archives just like printed materials. A campaign, led by the British Library on behalf of the UK's six legal deposit libraries, brings the law up-to-date. Since the law was first passed in 1911, the six legal deposit libraries have been able to collect copies of all printed material published in the UK. However, the increasing volume of important material published in electronic and other non-print formats fell outside the scope of the 1911 Act and was not therefore being comprehensively collected.

The adjusted law ensures that non-print publications can be saved as part of the published archive, becoming an important resource in their own right, and is indicative of the changing attitudes towards contemporary publishing and its preservation.

[BACK TO PAGE 1](#)

TURNING THE PAGES

Have you ever wished that reading digitised materials online was more like reading a real book? The British Library has developed an interactive application called 'Turning the Pages', which allows visitors to virtually 'turn' the pages of rare books and manuscripts. The pages are animated and appear to obey the laws of gravity, increasing the realism of the reading experience. Users can zoom in on

the texts, read or listen to notes explaining each page or use other features specific to the individual text.

This application allows access to many more pages than could be viewed if the manuscripts were simply displayed in a glass case and, crucially, without the risk of damage to the original.

Turning the Pages was launched in April 1998 and won the BIMA Interactive Display Award, followed by a Millennium Product Award and the National Heritage/NP Museum of the Year Multimedia Award. Originally based on touchscreen technology, since 2002 the Library, together with its commercial partner Armadillo Systems, has developed a Web version of Turning the Pages. When it was released using this software, Sultan Baybars' Qur'an was an immediate success. A selection of 40 pages from the Lindisfarne Gospels, with superior functionality, was added in May 2003.

Turning the Pages at the British Library (<http://www.bl.uk/collections/treasures/digitisation.html>) currently includes the Lindisfarne Gospels, the Diamond Sutra, the Sforza Hours, the Leonardo Notebook, the Golden Haggadah, the Luttrell Psalter, Elizabeth Blackwell's Herbal, the Sherborne Missal, and Sultan Baybars' Qur'an.

[BACK TO PAGE 1](#)

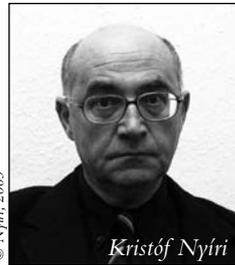


Screenshot showing the Lindisfarne Gospel viewed with Turning the Pages.

Copyright © The British Library

TOWARDS A KNOWLEDGE SOCIETY

PROFESSOR KRISTÓF NYÍRI
DIRECTOR, INSTITUTE FOR
PHILOSOPHICAL RESEARCH,
HUNGARIAN ACADEMY OF SCIENCES



© Nyíri, 2003

Kristóf Nyíri

New information and communication technologies herald the promise of, and indeed have to a significant measure already brought about, changes that can be meaningfully discussed under the heading of *knowledge societies*. However, the term ‘knowledge societies’ by now seems to have acquired two distinct, albeit related, meanings. In its first meaning it refers to the trend classically analysed by Daniel Bell in his *The Coming of Post-Industrial Society*.⁵² In post-industrial society – Bell also uses the terms ‘knowledgeable society’ and ‘knowledge society’ – knowledge and technology have become the central resources of society and economy. Perhaps it is the currently oft-used circumlocution ‘knowledge-based society’ that best expresses this state of affairs; and perhaps it is useful to stress, as Manuel Castells does, that what characterises this society ‘is not the centrality of knowledge and information, but the application of such knowledge and information to knowledge generation and information processing/communication devices, in a cumulative feedback loop between innovation and the uses of innovation’.⁵³

In its second meaning, the term ‘knowledge society’ is connected with the idea of universal access to knowledge, emphasising

that, while such access is becoming increasingly attainable today, it was never a possibility in earlier ages. In a trivial sense, all societies are knowledge-based; however, much of the knowledge both in traditional societies and throughout modernity was possessed by a minority, and much of that knowledge was not knowledge at all, but rather myth, superstition, lethal error or, at best, spurious learning. It does a disservice to progress to deny that we are today in fact witnessing a historical turning-point, and that past societies were ignorance-based societies rather than knowledge-based societies.

As a result of the Internet and mobile telephony, radically new forms of knowledge are emerging. It is philosophy and the social sciences that help us view this phenomenon from a historical perspective; help us understand that the new forms are indeed forms of genuine knowledge; and help us realise that the feeling of a cultural loss, voiced by so many contemporary intellectuals, is unfounded.

Now, of course, the term ‘knowledge’ refers to a family of concepts. There is knowledge as opposed to mere information, or indeed to raw data; multisensorial knowledge as opposed to textual knowledge; linearly and hierarchically organised knowledge as opposed to knowledge organised in network structures; and

indeed *practical* knowledge (skill) as opposed to theoretical knowledge. This last distinction is the most momentous – and the least tenable.

It was perhaps the main discovery of twentieth-century philosophy that *all* knowledge, ultimately, is based on practical knowledge, on *knowing how* rather than *knowing that*, as Gilbert Ryle has put it.⁵⁴ Practical knowledge is largely autonomous. There is a layer, or dimension, of practical knowledge that can in no sense be cast in the form of theoretical knowledge, and all theoretical knowledge represents merely an articulation of knowledge that is invariably reducible to practice. It is this sense that Ryle flatly states: ‘theorising is one practice amongst others’.⁵⁵

‘Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?’

T.S. Eliot

Throughout the ages, the knowledge possessed by humankind has primarily been practical knowledge: the knowledge of how to *do* things, and the *material culture* in which that know-how was embodied. Certainly there existed theoretical knowledge too, in the primordial pictorial form of cave paintings for example⁵⁶ and in the form of verbal knowledge: handed-down sayings, myths, and the like. However, much of that knowledge was, indeed, not knowledge at all. Prior to the rise of modern science, the progress of humankind was simply not due to any knowledge preserved in texts, oral or written.⁵⁷ Furthermore, one of the crucial factors in the rise of modern science was

52 Daniel Bell, *The Coming of Post-Industrial Society: A Venture in Social Forecasting* (New York: Basic Books, 1973).

53 Manuel Castells, *The Information Age: Economy, Society and Culture*, vol. 1: *The Rise of the Network Society* (Oxford: Blackwell, 1996), p. 32.

54 Gilbert Ryle, *The Concept of Mind* (London: Hutchinson's University Library, 1949).

55 Ryle, *The Concept of Mind*, p. 26.

56 Cf. John E. Pfeiffer, *The Creative Explosion: An Inquiry into the Origins of Art and Religion* (Ithaca, NY: Cornell University Press, 1982).

57 As Lynn White, Jr puts it in his *Medieval Technology and Social Change*: ‘Since, until recent centuries, technology was chiefly the concern of groups which wrote little, the role which technological development plays in human affairs has been neglected’ (*Medieval Technology and Social Change* (Oxford: Clarendon Press, 1962) p. vii).

precisely the early-modern encounter between traditional scholarship on the one hand, and the crafts and arts on the other.

It is essential to realise that the original task of the nascent *humanities* disciplines, too, was a thoroughly *practical* one. The emergence and development of the humanities were initially bound up with the spread of alphabetic writing, and subsequently with the development of printing. Entirely practical is, also, the emerging scholarship of the European Middle Ages. It amounts to no more than the mere exercise of the (still rare) ability to write. For centuries the issue is simply the conservation of texts by laborious copying; the learning conveyed by the University of Paris around the twelfth century culminates in the skill required for composing legal documents. After the invention of printing, the later humanists are taking an active part in the technical production of classical editions; it is printing which subsequently leads to new developments in the domains of grammar and letters, for instance the need to elaborate unified standards in orthography, syntax and vocabulary. On the basis of practice there arises theory – which however need not be far removed from life, as long as the practice itself is a living one.

Certainly by the nineteenth century, humanities scholarship had largely lost its touch with the real concerns of everyday life. Friedrich Nietzsche was the first philosopher to recognise this state of affairs. As he put it: 'Modern man finally drags a huge crowd of indigestible rocks of knowledge around inside him ... our modern culture is not alive ... it is really no true culture, but only a way of knowing about culture. There remain in it thoughts of culture, feelings of culture, but no cultural imperatives come from it.'⁵⁸ It appears, however, that Nietzsche's diagnosis is still unpalatable to a great many contem-

porary opinion-makers, who mistake the new practical culture of the digital age for the sickness for which it is a cure. Culture, learning and theoretical knowledge are humankind's collective instruments in the struggle to overcome its vital, practical problems. It is entirely obvious that digital networked communication is a vastly bet-

NOMEN HERBAE ASPARAGI
AGRESTIS.



"*Asparagus agrestis*" from the *Pseudo-Apuleius*, a printed version of a ninth century botanical manuscript, published just after 1480 in Rome. It contains woodcuts that are careless copies of the manuscript illustrations and could, of course, not be of any practical use.

ter-suited medium for dealing with the global problems facing us today than communication through channels inherited from the Gutenberg age. But then let us ask: do we have at our disposal, in those struggles today, real knowledge – or merely information?

Echoing T. S. Eliot's famous lines from the early 1930s – 'Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?' – John Naisbitt in his popular book *Megatrends* (1982) bemoans the phenomenon that the world is 'drowning in infor-

mation, but is starved for knowledge'.⁵⁹

The notion that 'information' is somehow inferior to 'knowledge' is not of recent origin, and modern definitions can be summed up by saying that knowledge might be usefully regarded as information in context. Now there is no reason at all to suppose that information looked for, and found, in the world of digital networked communication is more devoid of context than information gathered in the world of the printed page. The opposite is the case. Web pages offer a variety of links whereas books stand alone. Information encountered on the Internet can be more easily checked than information found in books. And as to *mobile* communications, it is a standard observation that information sought through cell phones is, characteristically, location-specific and situation-specific. It seems, then, that mobile communication, too, tends to engender not just information, but information in context: that is, knowledge *per se*.

Knowledge possessed by individuals will always represent but a fragment of what society at large knows; knowledge is dispersed in society. Also, knowledge is not independent of the medium in which it is embodied, preserved and communicated. Communication technologies, from cave paintings to the printed page, have always influenced the very nature of the knowledge they communicated. In particular, the rise of rational thinking is not independent of the spread of alphabetic literacy, first among the Greeks, and then in early-modern and modern Europe.

In his paper 'Visualisation and Cognition' Bruno Latour points to 'writing and imaging craftsmanship'⁶⁰ as the ultimate ground of modern science. Through the technologies of writing and pictorial representation the objects of cognition become mobile, and at the same time immutable; they can be collected, presented and combined with one another.

58 Friedrich Nietzsche, 'On the Use and Abuse of History for Life' (1874), translated by Ian C. Johnston, see <http://www.mala.bc.ca/~johnstoi/Nietzsche/history.htm>.

59 See also Vartan Gregorian's address at <http://www.cni.org/docs/tsh/Keynote.html>.

60 Bruno Latour, 'Visualisation and Cognition: Thinking with Eyes and Hands' in *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, vol. 6 (Greenwich, CT: JAI Press, 1986), p. 3.



Picture printing was invented around 1400 AD. This was arguably a much more revolutionary invention in the history of communication than that of typography half a century later. Prior to printing, pictures could not become aids to the communication of knowledge, as their inevitable distortion in the course of the copying process meant that information could not be reliably preserved. Printing allowed pictures to become more or less exactly repeatable; tangible knowledge became easier to disseminate. Of course, woodcuts, etchings and engravings were still a long way from being faithful copies of what they attempted to depict. Until the age of photography, there existed no technology of exactly repeatable pictorial representations of particular objects.

In the mid-1980s, there began a veritable iconic revolution, made possible by the graphical capabilities of computer software. The ease with which one can produce pictures and the everyday possibility of pictorial communication bring about a world in which people are becoming familiar with pictures, and are acquiring a rich experience of dealing with them, to an extent unprecedented throughout written history. The fact that we are increasingly able to communicate via pictures is of fundamental significance, since not words but perceptual symbols and, in particular, images are the primordial stuff of consciousness. To this age old view – never doubted by everyday thinking but practically forced underground by the psychology and philosophy of the first half of the twentieth century – science increasingly appears to return today. Human consciousness is a network of multisensorial, multimodal representations reflecting a world of movements, shapes, colours, sounds, and so on. Oral communication, rich in metaphors, embedded in here-and-now situations, and accompanied by non-verbal metacommunicative signals, can successfully cope with

the task of translating multimodal contents into the single modality of spoken language. Written language is a rather more limited medium. Thus digital multimedia documents, adding sound and image to text, can be truly liberating instruments of communication.

Let us note also that, whereas written language is poor at conveying practical knowledge, pictures, especially animated pictures – by themselves, or in combination with words – are quite effective.⁶¹ Pictures can show what texts can merely tell about, and pictures can summarise, in a way that can be grasped in a single glance, complex information that may be unintelligible when propositionally expressed.⁶²

The emergence of digital graphics is of course only one aspect of the profound change in the course of which the computer (as part of the interactive multimedia global network) has become an everyday element of knowledge production. Those patterns of mobility, immutability, compoundability and demonstrability analysed by Latour gain an entirely new meaning in the medium of the Internet. Science based on the book is replaced by science based on the global network. The barriers separating different specialties seem today to become fluid once more. A new, transdisciplinary mode of science emerges. This change is not independent of the fact that, as Gibbons *et al.* put it in their book *The New Production of Knowledge*, ‘the density of communication among scientists through various forms of mobility has been greatly increased in recent decades’, resulting in the ‘linking together of sites in a variety of ways – electronically, organisationally, socially, informally – through functioning networks of communication.’ This new mode of science is characterised by problem solving ‘organised around a particular application’, rather than by problem solv-

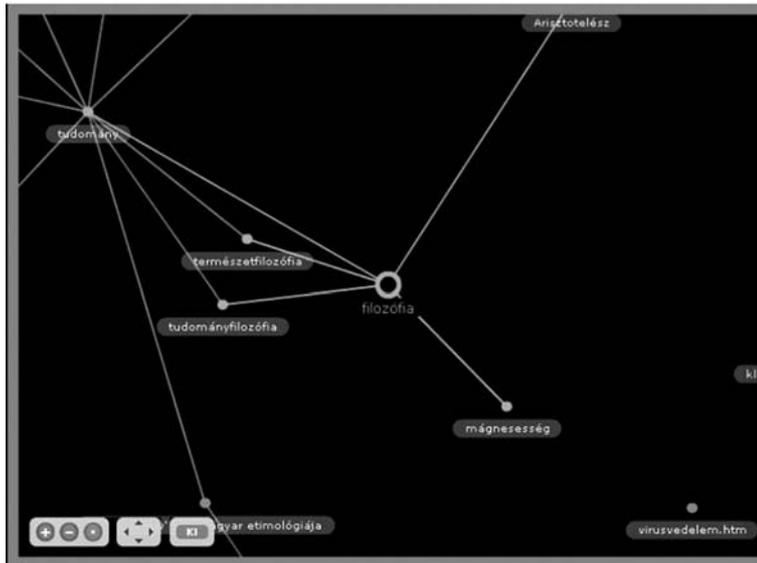
ing which is ‘carried out following the codes of practice relevant to a particular discipline’.⁶³ When the relative weight of applied research as compared with basic research is growing, the experience of coherence in everyday life overrides the image of fragmented scientific specialties. The idea of a non-fragmented, unified knowledge need not imply the possibility of a single harmonious vision of reality. It suffices if we can demonstrate the possibility of transitions from one field of knowledge to another; the possibility of conceptual bridges, passages, interactions – transitions which become easier when a word is enhanced by image. The digital network environment promises the emergence of a world of knowledge less fragmented than that of the past centuries.

Bearing in mind that human consciousness is itself a network of mental representations, the linear order of written language necessarily has a constrictive, indeed distorting, effect on thinking. Hence from the point of view of cognitive psychology the trend of supplanting, on the Web, extended linear texts by clusters of interlinked short documents is an unquestionably progressive one. Hypertext is a more natural form of organising ideas than the linear text is, and hypermediality, the interlinking of multimedia documents, is – given the multisensorial character of consciousness – an even more natural form. Supported by increasingly powerful search engines, the World Wide Web has the potential to become a truly inalienated communicational environment. The Web – now enhanced by the miracle of mobile communications – is still vastly underrated, and indeed misrepresented, even by leading opinion-makers, not least since headstrong views characteristically lack the foundation of hands-on experience. Hypermediality creates an environment in which fragmented theoretical knowledge, and also practical knowledge dispersed among the members of society, becomes more easily accessible than ever before.

61 For some references see Kristóf Nyíri ‘Pictorial Meaning and Mobile Communication’, in Nyíri (ed.), *Mobile Communication: Essays on Cognition and Community*, pp. 175 f. and 179.

62 A brilliant book on the subject is Colin Ware, *Information Visualization* (San Francisco: Morgan Kaufmann, 2000).

63 Michael Gibbons, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott and Martin Trow, *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies* (London: SAGE Publications, 1994), pp. 38, 6, 39, 45, 44, 3.



© Nyíri, 2003

Screenshot demonstrating the planned Hungarian Virtual Encyclopedia ('Philosophy')

Although there exists a real inequality in terms of the visibility of the different nodes and hubs on the Net,⁶⁴ the new ICTs are still very conducive to the building of bottom-up knowledge. Publishing on the Web is infinitely easier and more democratic than publishing in the world of the printed page ever was.

By way of concluding, I will present an illustration of networked knowledge in action: the first plans of a Hungarian Virtual Encyclopaedia, the construction of which has been entrusted to the Institute for Philosophical Research of the Hungarian Academy of Sciences (<http://www.mta.hu/>), with numerous institutes of the Academy participating in its development. We envisage an encyclopaedia in the classical sense: a full circle of learning – a densely interconnected web of entries, with specially designed software to map and display the structure of links. The mapping software will place unlinked entries far towards the peripheries of the map, while densely linked entries are moved towards the centre.

BACK TO PAGE 1

LATE BREAKING NEWS

As we went to press we selected the following reports, surveys, and working groups from those that crossed our desks as worthy of mention on 10 December 2003.

AMERICAN MUSEUMS DIGITAL IMAGING SURVEY

Professors Roy S Berns and Franziska Frey of the Rochester Institute of Technology have begun a study of direct digital image capture of works of art within US museums, archives and other cultural heritage institutions. To give their study a good grounding they are conducting a survey. For further information and forms see: <http://www.cis.rit.edu/museumSurvey/intro.php>.

PUBLICATION OF 'IT'S ABOUT TIME: RESEARCH CHALLENGES IN DIGITAL ARCHIVING AND LONG-TERM PRESERVATION'

At the start of December, The National Digital Information Infrastructure and Preservation Program (NDIIPP) at the U.S. Library of Congress announced the publication of 'It's About Time: Research

Challenges in Digital Archiving and Long-term Preservation.' The report presents findings from a joint Library of Congress and National Science Foundation workshop on research challenges in digital preservation. The report can be downloaded from: <http://www.digitalpreservation.gov/index.php?nav=3&subnav=11>.

An example of one of the types of research needed is emulation, as discussed in the Digital Preservation Testbed White Paper "Emulation: Context and Current Status" available from: http://www.digitaleduurzaamheid.nl/bibliotheek/docs/White_paper_emulation_UK.pdf.

PRESERVATION METADATA WORKGROUP

OLC's and RLG's sponsored Preservation Metadata: Implementation Strategies (PREMIS) workgroup has two main aims. One goal is to establish a core set of preservation metadata with wide applicability. A second objective is to understand how the scientific, cultural heritage, and public sectors develop and run digital repositories, as well as developing an appreciation as to how they are encoding, storing and managing their preservation metadata. If you have a chance to help out with the The Implementation Survey before 16 January you can find out how to do so at: <http://www.oclc.org/research/pmwg/>.

NEW DUBLIN CORE WORKING GROUP ON PRESERVATION METADATA

Drs Andrew Wilson (National Archives of Australia) and Heike Neuroth (University of Göttingen) announced the formation of a Dublin Core Working Group on Preservation Metadata. Information about the workgroup can be found at: <http://dublincore.org/groups/preservation/> and those wishing keep up to date with or to contribute to discussions about preservation metadata can subscribe to: <http://www.jiscmail.ac.uk/lists/DC-PRESERVATION.html>.

64 Cf. Albert-László Barabási, *Linked: The New Science of Networks* (Cambridge, MA: Perseus Publishing, 2002); Steven Strogatz, *Sync: The Emerging Science of Spontaneous Order* (New York: Hyperion, 2003); Mark Buchanan, *NEXUS: Small Worlds and the Groundbreaking Science of Networks* (New York: W.W. Norton & Co., 2002).



CULTURE 2000 PROGRAMME AND AUDIOVISUAL POLICIES

Just after we closed this issue we received a report of discussions on the future of the Culture 2000 programme and of the European audiovisual policies were the major issues on the agenda of the Culture Council on 24 November. During their Council meeting ministers exhaustively examined the Commission's mid-term reports on Culture 2000 and Media programmes. During the debate on the Culture 2000 programme, ministers agreed to pursue closer cultural co-operation within Europe and with third countries, and to enable access by new Member States' to the programme through specific actions.

Based on the mid-term report Ministers concluded that the Media and Media training programmes showed how a lack of coherence has undermined EU audiovisual policies. Ministers agreed that it should be made easier for new Member States to access funding and that digital technology should be included in the programme through pilot projects. During the first quarter of 2004, the Commission will present a draft proposal for the 2007-2013 Culture and Media programmes. For more information on the programmes see Media: http://europa.eu.int/comm/avpolicy/media/index_en.html and Culture 2000: http://europa.eu.int/comm/culture/eac/index_en.html

THE GLASS WALL

Thanks to one of our contacts we recently saw a printout of document produced by the BBC travelling the web under the filename of `theglasswall_bbc_design.pdf`. The document describes how the BBC went about revising the design of its website before its relaunch last year. Rumour has it that 'the document was never officially released, instead simply leaked and it's difficult to get hold of, since whenever it appears online, BBC bosses request that it be removed.' If you can find a copy, it is worth a look.

THE AMAZON.COM 'NOT A DIGITAL LIBRARY'

The December (2003) issue of *Wired* carries a piece by Gary Wolf on 'The Great Library of Amazonia'. *Search Inside the Book* makes it possible for registered users to search and then view pages from 120,000 volumes representing some 33 million pages. Amazon claims that it is not attempting to build a digital library, but that its aim is to sell more books by giving users more effective access to content. The technical design, delivery mechanisms, and business model that underpin Amazon's approach are worthy of evaluation.

LEADERS TOOLKIT

A new toolkit has been developed that delivers transcripts and images of archive documents over the Internet together with appropriate contextual material. The LEADERS toolkit is built on standard methods for encoding texts (TEI), digital images (NISO Mix), finding aids (EAD) and authority records (EAC). A demonstration of the project using test data is online at the LEADERS Project home page (<http://www.ucl.ac.uk/leaders-project>).

EVALUATION OF DIGITAL CULTURE CONTENT

The Cultural Content Forum has just made available a new report, 'Evaluation of Digital Cultural Content: Analysis of evaluation material' evaluating audiences for online cultural content. The analysis is wide-ranging and reflects the experience of 90 different evaluation studies from across the world. The study emphasises the lack of standardised approaches. This study and others are available for download from <http://www.culturalcontentforum.org/publications/> Evaluation issues will be covered more fully in the next issue of *DigiCULT.Info*

[BACK TO PAGE 1](#)