

# DigiCULT .Info

Issue 5 A Newsletter on Digital Culture

ISSN 1609-3941

October 2003



## INTRODUCTION

An extra day in Berlin after IFLA2003 provided an opportunity to visit the Museumsinsel, one of the finest museum complexes in the world. Currently many of its buildings are undergoing extensive renovations. As on my right I passed the reopened Alte Nationalgalerie I noticed one of the buildings currently wrapped for restoration was covered by a massive poster. In the words

*Continued next page*

### Technology explained...

- p. 11 An Introduction to Application Service Providers
- p. 13 Value Chain for Digital Learning Content
- p. 14 Human Language Technologies Explained
- p. 26 WebKit – Intuitive Physical Interfaces for Information Retrieval and Management
- p. 27 Stewardship Resources on the Web: Raising awareness of the role stewardship plays in ensuring the preservation of cultural material
- p. 36 Open Source and Free Software Solutions: Reliable standardised, free

### News, events...

- p. 13 Trials Support for European Projects
- p. 16 The Second Conference of the National Digitisation Centre of Serbia and Montenegro: Digitisation from a Balkan perspective
- p. 17 Cultural Heritage Events
- p. 18 DigiCULT Seeks National Correspondents
- p. 18 Netarchive.dk
- p. 21 News from UK Archives
- p. 26 Survey of Cultural Websites
- p. 26 Intellectual Property and Technology Forum
- p. 33 UNESCO Convention for the Safeguarding of the Intangible Heritage of the World



© Steinhilber 2003

### Technology applied...

- p. 8 Enhancing Learning with Online Games
- p. 8 Human Language Technologies in Europe: The EUROMAP Language Technologies Project
- p. 18 Preserving electronic records from database-driven information systems
- p. 22 The Open Video Project: Building an open source video archive
- p. 39 Introducing Digitisation into Turkish Libraries: Current attitudes and the way forward

### Challenges, strategic issues, new initiatives...

- p. 5 Launching the Digital ARK
- p. 32 JODI Issue on Economic Factors for Digital Libraries
- p. 29 Semantic web Challenge
- p. 30 DELOS Network of Excellence
- p. 33 E-Culture Net Thematic Issues Network (2002–2003)
- p. 41 Digital Access to Aerial and Astronomical Archives

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



Weltkultur beflügelt (world culture gives wings) it asks all who pass to consider the liberating power of the cultural heritage. Articles in this issue of DigiCULT.Info all show the power of new technologies in helping cultural heritage institutions in achieving their objective of improving the care, understanding and benefits of cultural heritage to individuals and society. Reflecting on the phrase *Weltkultur beflügelt*, we are reminded that technology is an enabler and not an end in itself.

We often think of our heritage mainly as the materials held by archives, libraries, and museums, but our natural, environmental and biological heritages have equal, if not more primary, merits.

*Launching the Digital ARK* describes a Bristol based project that is making use of digital technologies to select, collect, make accessible, and preserve in perpetuity a digital record of the Earth's biodiversity beginning with the endangered species of the British Isles. The project has had to address formidable organisational, intellectual property rights, and technical challenges as it strives to build a descriptive, audio, still and moving image record of these species. An accompanying interview with the Senior Education Officer at ARKive introduces us to the thinking that lies behind their use of games to promote learning. The team at ARKive started with the recognition that play had value as a learning medium. They then took their digital asset base and built a series of interactive educational materials on top of it. The educational team at ARKive recognised that maximum value comes not from the content itself but from the ways technology enables us to increase the value of the content and make it work.

Demand for digital video materials for learning, teaching, and research purposes continues to grow. A team led by Gary Marchionini at the Interaction Design Laboratory at the School of Information and Library Science at the University of

North Carolina (Chapel Hill) are gaining practical experience addressing the structural, rights, and technical complexities that make development of video archives difficult. Meng Yang and colleagues provide an introduction to the work of *The Open Video Project*. Their paper investigates the obstacles that make development and management of video archives challenging, and sketches the practical approaches (e.g. video segmentation technologies) that can improve the usability of archives of this kind. The work of the project provides a fundamental building block for further research and activity in this area.

While heritage institutions aim to increase their use of technology they are looking at ways of reducing the costs associated with doing so. Thomas Finholt from the Collaboratory for Research on Electronic Work (CREW) at University of Michigan's School of Information examines how Application Service Providers (ASP's) might offer institutions a way of doing this. Their research enabled the team at CREW to draw a number of conclusions that are of value to the heritage sector. Among them is the recognition that the public sector institutions most likely to benefit from ASP's are the medium-sized organisations as they are, as Finholt explains, 'big enough to need efficient information technology and support but are not big enough to carry the costs of the equipment and staff'. It is widely recognised that heritage institutions remain poorly supported by information technologies. Many heritage institutions find themselves struggling to keep pace with the opportunities offered by technology and visitor expectations (see for example, [http://www.hlf.org.uk/dimages/needs\\_ict.pdf](http://www.hlf.org.uk/dimages/needs_ict.pdf)).

Another way to reduce technology costs is to move to open source and free software programmes (OSS and OFS), and Andrew McHugh investigates these

alternatives. They offer technical solutions to users that are low risk, standardised, reliable, and low cost (if not free). This introduction was written using Open Office rather than the more commonly used proprietary alternative from a company based in Washington State. Open Office is easy to use, friendly, feature rich, and cost effective. Moreover, it has a low learning curve. After reading Andrew McHugh's article a useful followup resource might be *The IDA Open Source Migration Guidelines (Version 1.0)* (<http://europa.eu.int/ISPO/ida/export/files/en/1603.pdf>). This provides IT managers and practitioners in public administrations (within Europe) planning or doing migrations to open source software with a comprehensive introduction to experience based best practices. The authors, from **netproject Ltd**, report that the use of OSS in the server environment 'is well understood and is extensively deployed' but 'deploying OSS on the desktop for most organisations offers the largest cost savings' (October 2003, p. 12). The decision in May 2003 of the Munich City Council to migrate its 14,000 desktop and notebook computers from Microsoft Windows products to Linux and open-source software was widely reported (see for example, <http://www.computerweekly.com/Article122160.htm>). Richard Seibt, the CEO of SuSE Linux AG, commented at the time that the 'city clearly sees Linux not just as providing cost savings over costly, proprietary software, but also as the best tool for the job bringing security, stability, flexibility and privacy not available to them before.'

Darryl and Judy Mead, of M Squared Consulting, report on a detailed study they produced on the twenty best stewardship information sources they identified on the Web. Stewardship in this context includes collections management, care and access as applied to portable cultural heritage. The report drew a number of conclusions among them that a substantial



number of 'stewardship sites were equally relevant for libraries, archives and museums', that developers of these resources did not pay sufficient attention to the objectives and needs of potential users of the resources they were constructing, and that the longer term sustainability of some valuable stewardship resources was questionable.

Questions of stewardship of portable objects arise when we consider those elements of our heritage that are stored on digital media. Awareness of the technical difficulties posed by CD and DVD media, such as the variable and seemingly often unpredictable life span of CD-Rs is becoming more widespread. We look for CD-Rs produced with phthalocyanine rather than cyanine die layers when longer-term storage is an issue, we store the CD-Rs in optimal humidity and temperature conditions, we minimise how often we handle the media and when we do we avoid scratching it or even depositing oil or other dirt from our hands onto it, we do not write on (except on the very inner hub area and then with solvent free felt markers) or stick labels on the media itself, and we keep the media out of the sunlight. We recognise that even with these precautions, if we are managing stores of digital media we will need to establish and use strategies for regular assessing the status of our CDs. Some types of CDs are more durable than others, but none are immune to degradation. The National Institute of Standards and Technology (US) recently released a draft report on the *Care and Handling of CDs and DVDs – A Guide for Librarians and Archivists* (NIST Special Publication 500-252, October 2003), which provides an authoritative, accessible, and comprehensive introduction to the topic. <http://www.itl.nist.gov/div895/carefordisc/CDandDVDCareandHandlingGuide.pdf>

More recent developments, such as self-destructing DVDs, pose new kinds of challenges to long term access to

cultural heritage resources. Disney has just released movies on EZ-D DVDs that self-destruct forty-eight hours after the purchaser opens the airtight package containing the DVD. The underlying technology licensed by Flexplay Technologies (<http://www.flexplay.com/>) creates a chemically restricted 'viewing period or window' during which the DVD can be played as often as the user wishes, but after which the disc is rendered unreadable. The chemical reaction starts once the DVD is exposed to air. These chemical processes combined with the software designed to make copying difficult (if not impossible) create new difficulties for those institutions which aim to collect and preserve our heritage. It is easy to imagine future scenarios where heritage institutions having acquired materials with a view to ensuring they could be made accessible decades down the road are unaware of the risks associated with opening the packaging or of the degradation of the packaging itself. How will institutions verify that the contents of digital objects they have acquired are complete if they can not open the packaging to check the media at point of acquisition or how will we warn future curators about the key role that packaging plays in maintaining the readability of such digital objects?

Access to and preservation of digital resources, whether the product of digitisation or born digital, have become crucial activities of memory institutions worldwide. Much of this effort has originated in national libraries and archives. In a very few countries this has resulted in new laws to promote such preservation. The National Library of New Zealand (NLNZ) Act 2003, which came into force in May, has mandated the Library to collect, preserve, and make available New Zealand's electronic documents. The Act responds to changes in how our documentary heritage is created, disseminated, and used. 'The diversity and complex nature of electronic objects whether websites, digital manuscripts, pack-

aged publications (such as those on CDs), or products of in-house digitisation can only be handled by new ways of working and enhanced curatorial and technical services.' To fulfil this mandate NLNZ needs to craft adequate procedural and policy guidelines, as well as robust technical and staffing infrastructures that will enable it to discharge its new responsibilities. Numerous other institutions are confronting similar problems. The work that has been done by the Digital Library team at the NLNZ to lay the foundation for a digital library capable of enabling the Library to respond to the requirements laid out in the Act is worthy of review by institutions considering similar developments. The work of and challenges facing the National of New Zealand are examined in a review of their digital library development activities published at the end of July 2003 ([http://www.natlib.govt.nz/files/ross\\_report.pdf](http://www.natlib.govt.nz/files/ross_report.pdf)).

Maria Sliwinska, of The International Centre for Information Management Systems and Services (ICIMSS) in Torun (Poland) draws our attention to the work of the Fifth Framework funded DELOS Network of Excellence on Digital Libraries, which brought together sixty-eight academic and industrial organisations to examine digital library related issues. Her report provides a solid introduction to the contributions of DELOS in general and to its impact on digital library activity in Central and Eastern Europe in specific. It reminded us that we have not called the attention of readers of DigiCULT.Info to the work of the joint working groups established by DELOS and the National Science Foundation (NSF) to study the digital library landscape and define research agendas surrounding a number of core challenges facing digital libraries (<http://delosnoe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/joint-wgs.html>). Among the eight working groups were those defined research agendas for Spoken-Word Digital Audio Collections, Digital



Libraries Information Infrastructures, Personalization and Recommender Systems in Digital Libraries, and Digital Archiving and Preservation.

The report of the last of these working groups, *Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation* (2003), (<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/JointWGs/digitalarchiving/Digitalarchiving.pdf>) charts the research that is needed if digital libraries are to have access to the experience, methodologies, practices, and technologies necessary to ensure the long term accessibility and usability of the digital assets that they acquire. While this report has identified work to be done, it acknowledges that some high quality research and experimentation in digital preservation that is delivering solutions on the ground. In this issue Filip Boudrez of the City Archives of Antwerp describes in 'Preserving electronic records from database-driven information systems' the recordkeeping system developed by the DAVID Project to enable them to investigate the preservation of authentic and durable electronic records.

Daisy Abbott of the DigiCULT team has provided an introduction to Human Language Technologies (HLT). In future issues we aim to build on the foundation she has laid by investigating in more detail how these technologies will improve access to and understanding of the cultural heritage. The uses are varied from devices which can be controlled through speech input, to support for cross-lingual search engines, to intelligent labelling in museums and other heritage institutions.

Jean-Pierre De Cuyper, of the Royal Observatory in Belgium, introduces us to a project that aims to digitise astronomical and aerial photographic plates. The project, a collaboration between content holders, academic institutions, and the commercial sector, is

creating the expertise, including an understanding of how the plates themselves were created, what hardware and software will enable the project to extract the maximum amounts of historic-scientific information from the photographic plates. What becomes obvious in reading this article is that understanding how information is represented on the medium is essential in selecting and establishing adequate digital imaging systems. This conclusion, has sadly been overlooked by many digitisation projects as they rush to purchase scanning equipment and to get underway with digitisation. In Minerva Europe has an activity which is attempting to promote best practice in digitisation and it has worked hard to foster the take-up of the Lund Principles within the European Union Member States. It is now embarking on developing training programmes. Good digitisation practice is an international problem as is evident from the report of Selenay Aytac (Isik University, Istanbul) on a survey she conducted to investigate the level of knowledge about digitisation issues among library professionals in Turkey. Among her conclusions is the argument that librarians and other information professionals in Turkey need access to professional development opportunities in both digitisation and digital preservation. There is an international need for programs of continuous professional development more generally in cultural and heritage informatics.

DigiCULT.Info may too often concentrate on the more classical applications of information, communication and technology (ICT) to the cultural heritage sector. An exhibition which just closed at the Cooper-Hewitt National Design Museum (NYC), offered an exciting glimpse of how combinations of technologies are delivering new materials for architectural spaces. The progress of the exhibition can be followed in a series of photos accessible at: <http://ndm.si.edu/SOLOS/>. Stephen Kieran and James Timberlake demonstrated materials that they believe will be one of

the 'building skins of the future.' SmartWrap combines plastics with devices that insulate, store energy, heat, and provide access to power and light. For example, within SmartWrap solar panels collect energy, flat batteries store it, and light emitting diodes can take advantage of the energy to illuminate, change colour, or to display patterns such as advertisements or textures, say the 'brick look' on its surface. The combination of plastics, printing technologies, phase change materials, and organic light emitting diodes (OLED) will provide heritage institutions with a wide range of new opportunities (e.g. the ability to create micro-environments for exhibitions).

As in previous issues we have highlighted key events. DigiCULT's website supports a database of upcoming events which is more comprehensive than the information published here. We hope that the cultural heritage community will contribute information about events to this database and help us to make it a richer source of information.

We have included some shorter articles or pointers to recent reports or new sources of information that the readers might find beneficial. For example, in this issue we have highlighted the launch by the UK and Ireland branch of the International Association of Music Libraries, Archives and Documentation Centres (IAML) of Cecilia, an online database of music resources and collections throughout the UK and Ireland. To increase the diversity in the types and sources of information that we are able to make accessible to readers DigiCULT continues to develop its network of correspondents. At the same time we hope that if readers have ideas for articles or wish to call materials to the attention of DigiCULT readers they will contact Daisy Abbott at [d.abbott@hatii.arts.gla.ac.uk](mailto:d.abbott@hatii.arts.gla.ac.uk).

Seamus Ross & John Pereira  
Editors, DigiCULT.Info

## LAUNCHING THE DIGITAL ARK

The launch of ARKive in May this year by Sir David Attenborough was the realisation of a vision first conceived nearly twenty years ago – before the technology was available to even attempt making the project a reality.

Described as a 21st century Noah's Ark, ARKive is leading the virtual conservation effort. It is harnessing the latest in digital technology to bring together, for the first time, the world's most important nature films, photographs, sound recordings, and memories, then using them to build vivid and fact-backed portraits of Earth's endangered plants and animals. The project's aim is to provide a permanent safe haven for the records, and the insights they offer, and to make the best of the collection globally accessible, for pleasure, learning and research, via a free, click-and-view Website – <http://www.arkive.org>

ARKive is an initiative of The Wildscreen Trust, an educational charity which organises the biennial Wildscreen Festival and works to raise conservation awareness by encouraging and applauding excellence in wildlife filmmaking and related media.

### A VISION BEFORE ITS TIME

When natural history film-maker Christopher Parsons first put forward his vision for an environmental record archive along with Sir Peter Scott, he was already used to discovering that technology was not necessarily up to the same speed as the thoughts in his head. From his entry into wildlife television in 1955 to his death, aged seventy, last year, he was forever thinking up impossible ideas, then finding the kit and the kindred spirits to make them work. Despite Christopher's success at enabling innovation in wildlife

filmmaking, there were many years when his proposal for a central library of wildlife films, photographs and sound effects seemed likely to remain on the shelf.

Certainly, there was wide international support for the project. It was hailed as brilliant, essential, the Noah's Ark of the electronic era. Then a government-funded feasibility study revealed the one snag.

There simply wasn't the computing power anywhere in the world to copy, hold, index, layer, sort, retrieve and relay the amount of information involved. However, as Sir David Attenborough, a long-time friend and colleague, explains, Chris Parsons was nothing if not determined. "Once he took up a project – a series, a book, a revolutionary way of popularising the natural world – I know of no-one more persistent, ingenious, diplomatic and, ultimately, more successful." And, as Parsons had proved many times previously, technology changes.

### ARKIVE BECOMES A REALITY

So it was that on the 20th of May this year, almost twenty years since its conception, David Attenborough combined a memorial tribute to one of natural history's "most influential, most imaginative and most self-effacing champions" with the launch of his last gift to it: a digital safe haven for sounds and images celebrating and explaining Earth's biodiversity – ARKive.

Phase one of the £2.5m ARKive project is a Website <http://www.arkive.org> – funded primarily by Hewlett-Packard Laboratories (Europe), the Heritage Lottery Fund and New

*"Over the past few decades a vast treasury of wildlife images has been steadily accumulating, yet no one has known its full extent – or its gaps – and no one has had a comprehensive way of getting access to it. ARKive will put that right. It will become an invaluable tool for all concerned with the well-being of the natural world."*

Sir David Attenborough

Opportunities Fund, and offering free global access to multimedia portraits of almost 1000 plants and animals.

### EVOLVING ARKIVE

Initial work on project design centred around the building of taxonomies, and the idea of creating a species record. At the time there were no taxonomies for media describing animal species that were considered adequate for ARKive's preservation remit. So ARKive created its own, in consultation with academics concerned with endangered species, and with a research team of experts in digital media management and preservation at Hewlett-Packard Laboratories (HP Labs). This allowed the creation of a flexible taxonomy that would allow for expansion, while still being manageable enough for use on a day-to-day basis by a team of media researchers.



Sir David Attenborough at the ARKive launch in May 2003



*Last known individual thylacine. Moving images available from ARKive: <http://www.arkive.org/species/speciesOverview.do?id=4758>*

Having designed a way to tag the data, the project needed a Digital Asset Management (DAM) and workflow application. HP Labs, having investigated the DAM and workflow markets, undertook a major development project to build an accessioning system for ARKive as part of their \$2 million donation of technical professional services. Videotapes and slides move in and out of the building regularly. These pieces of media are in many cases irreplaceable should they be damaged, so it is essential to have a robust and reliable system in place.

Tracking down the records on the site at present has been made easier by the unprecedented levels of co-operation the ARKive team has encountered from record-holders. Most of the world's leading broadcasters, picture libraries, conservation bodies and academic institutions have set aside their initial legitimate concerns about rivalry and copyright protection to share information. Donations have also come from many individual camera operators, scientists and amateur nature observers. Once the records began to come in, a key challenge for ARKive's researchers was to decide on the essential ingredients of each multimedia profile. What - they had to ask - would people a century or more from now want to know about a species that might no longer exist?

the song of the dodo was like. Therefore it was agreed the profiles needed to feature sounds as well as visuals wherever possible, along with factuality illustrating appearance, movement, feeding habits, social behaviour, mating, threats, special adaptations and interactions with other species, including humans.

Finding the pictures that matched the checklist was the next challenge. Chief media researcher, Richard Edwards, estimates he and his colleagues viewed and logged more than 12,000 slides to select the 5000 on the Website at launch, and watched over 2000 hours of videotape.

Compiling the records was only part of the story. Next, data had to be edited, copied and stored - and in such a way that it could be readily indexed, cross-referenced and retrieved, and cope with future changes of technology.

At ARKive's heart is a vast storage infrastructure linked to the latest optical networking running at 2Gb/s. It can store and access up to 74 Terabytes of data - roughly the same capacity as fifty million floppy disks - and can output moving footage in a range of formats, from dial-up modem-sized videos to broadcast quality clips for professional TV production.

For inspiration, the team turned to the dodo - the flightless bird hunted to extinction more than 200 years ago. Current knowledge of it comes from drawings, observer notes, the diaries of sailors, and a handful of preserved remains. But none of the records reveals what

The data capture and Web publishing systems are built from off-the-shelf modules, but their speed and capacity are far from run of the mill. Storage is on a robot-served LTO Ultrium-1 tape library - the largest tape library made by Hewlett-Packard and which, with a tape access time of less than ten seconds, is one of the fastest available. This connects to state-of-the-art Web servers via a 2Gb/s optical network. The whole system is highly robust and, for added long-term security, makes use of double UPS, doubly redundant disk arrays and multiple servers.

IT Systems Manager, Rob Curtis, says: "The durability of the system, and the back-ups we have in place are a crucial part of the project. After all, ARKive will be storing some of the most valuable wildlife footage on the planet and we have to be sure we can keep it safe for many, many years to come." The value of the project - to wildlife conservation and education, and to film and television history - has won widespread support for ARKive from many quarters, including the BBC, Discovery Channels International, Granada Television, National Geographic, BirdLife International, Fauna and Flora International, the United Nations Environment Programme, World Conservation Monitoring Centre, the World Conservation Union (IUCN), and the World Wildlife Fund (WWF).

#### WHAT'S ON BOARD THE ARK?

The information is arranged in two chapters. Chapter One is dedicated to British animals, plants and fungi - from the familiar to the rare and obscure. Chapter Two, for globally endangered species, is only just beginning but currently covers 250 representative examples from the IUCN's Red Lists.

Each portrait comprises video footage and stills photographs, backed by authenticated fact-files about the lifeform's



More ARKive material about otters:  
<http://www.arkive.org/species/speciesOverview.do?id=4484>

lifestyle, habitat, reproduction, survival strategies, threats and more. The downloadable information is layered, so users can choose the format which best suits their level of expertise and research needs, and is linked to sources of further information.

There is a separate zone for teachers, offering worksheets, project ideas and lesson plans. For children, there is Planet ARKive, where interactive games convey learning points about habitats, predation, and life-cycles, making learning an enjoyable experience.

Project Manager, Harriet Nimmo, says: "The Website is a monumental undertaking, but it's important to stress that ARKive is much more than an online wildlife reference library. Essentially, the Website is a window into a vast and expanding vault of vital wildlife records. The films, photographs, sound recordings, books, facts and memories we are collecting are not just powerful tools, which help people to understand nature and why it needs conserving. The value of what they reveal increases every time another depicted species is threatened or disappears. Yet,

without ARKive, there is no easy-to-reach centralised repository for the information. The source material is scattered all around the world, sometimes in places where access is restricted, or where its value is not fully appreciated."

## SUPPORT

ARKive is endorsed by many leading scientists, among them Sylvia Earle, Richard Leakey, Tom Lovejoy, Robert May, and Professor E.O. Wilson. Andrew P. Dobson, the professor of ecology and evolutionary biology at Princeton University, USA, spoke for many when he gave his reasons for wanting ARKive to happen: "ARKive provides a unique and crucial opportunity to create a virtual museum of natural history. It will document for the world the vast range of inhabitants who have shared the planet with us, and the habitats in which they once lived. In many cases it will provide the only documentary evidence of how these species moved, ate, reproduced, and interacted with each other. They are images we should conserve with an intensity equal to that devoted to our greatest works of art. They are images that will haunt our grandchildren."

## PLANNING FOR THE FUTURE

ARKive's manager, Harriet Nimmo, is the first to admit ARKive is only just beginning to become what Chris Parsons envisaged. "We need to add many more portraits to the Website, and we need to carry on building our complementary physical library of films, photographs, books and oral memories, and find the funding to make more of the collection available to the public. But it's a magnificent start and we believe Chris would be proud of how far we've come."

To help with the rest of the journey, the team is keen to hear about any specialist collections of wildlife films, photographs or sound recordings, especially of rare or less studied species. A list of what is

currently classed as 'Most Wanted' appears at [http://www.arkive.org/news\\_mostwanted.html](http://www.arkive.org/news_mostwanted.html). As the quest for further records continues, a story from Dr Galen Rathbun, a Wildlife Research Biologist (Emeritus) at California Academy of Sciences, encapsulates the urgency of the effort.

*"In the late 1970s I deposited my footage of the golden rumped elephant shrew with Byron Motion Pictures in Washington, DC. They had a film vault where the National Zoo also kept much of their valued footage. In the late 1980s I wanted the footage, and tried to make contact with Byron, only to find that my material (along with the entire zoo collection) had totally disappeared. Apparently Byron went out of business and everyone's cine footage was lost..."*

To date, only one other short sequence of the shrew has been located.

## RECOGNITION

In June this year ARKive was assessed by the National Grid for Learning, founded by the UK Government in 1997 to promote excellence in educational Websites. The project became the 1000th Website to be accepted into their learning portal <http://www.ngfl.gov.uk>. This acceptance recognises not only the standard of content and research which forms the heart of this Website but the educational value and importance of the project as a whole.

The launch is only the beginning with around 600 species (half British) on board. This will rise to 1500 by the end of the year, with work continuing to include the 11,000 animals and plant species threatened with extinction. This will ensure that ARKive not only has its work cut out over the coming months and years, but that there will always be a comprehensive and permanent digital record of our life on Earth.

BACK TO PAGE 1



## ENHANCING LEARNING WITH ONLINE GAMES

**P**lanet ARKive (<http://www.planetarkive.org>) and ARKive Education (<http://www.arkiveeducation.org/>) are specialised sections of ARKive's Website for learning and teaching. **Daisy Abbott** spoke to **Karen McDonnell, ARKive's Senior Education Officer**, about using games as an educational tool.

### **How did you become involved in the ARKive project?**

I trained as a teacher and worked in primary, secondary and special schools before leaving teaching to work as an Assistant Education Officer at Jersey Zoo where my love of wildlife and passion for conservation combined with teaching could be thoroughly indulged. This led to several years in lecturing on zoo education and devising education programmes for various zoos and wildlife sanctuaries before taking up the post of Senior Education Officer for the ARKive project. It was (and is) my job, along with my colleague Jo Canning, to devise a way of taking the information and media created for the main ARKive site and 'repurposing it' for a different audience.

### **The information presented on the ARKive Website is layered to better suit different users. Were the ARKive education facilities (aimed at teachers) and Planet ARKive planned from the beginning, or did they evolve later in the project?**

From the very beginning ARKive was going to have an education remit and education was an important part of our funding bids. Before I joined the project, it had been decided that the initial focus would be on Key Stage 2 primary pupils. The reason for this was that children aged 7 to 11 have an innate interest in wildlife and there

is a great deal within the National Curriculum at that particular Key Stage that concerns living things. I thought it was very important that the children should have their own Website, with its own identity and a separate URL. Children like to have ownership of sites; they like to see that it has been created especially for them and no one else. That meant that we also had to have a separate site for educating adults. We started from scratch, devising a plan for the two education sites listing the aims, objectives and content and I was responsible for delivering those sites in time for launch in May 2003.

We don't like to think of ARKive Education as purely for teachers; it is for any adult who wishes to use the sites with children – teachers, parents and specialist education officers such as those in zoos, wildlife parks or natural history museums. By having two separate sites our job was made a great deal easier in terms of design and content; we did not have to make any of the sites appeal to a wide audience, they could all be specifically targeted.

### **Why did you decide to use games to facilitate learning?**

Jo and I are trained teachers and we both know the value of play as a learning medium. Also, we spent weeks looking at the variety of sites aimed at children on the Web and they all utilised games – although some were much better than others. We were very conscious that the games we created had to have learning outcomes; we did not want to have gaming for the sake of gaming. No matter how 'soft' the educational objective, it had to be there. It was therefore extremely important for us to find the right programmers to help us to create the games and we were

lucky to get Simon Gurr (<http://www.simongurr.com/>) and CompleteControl (<http://www.completecontrol.co.uk/>) who made our ideas for the games work.

### **How did you go about designing specific educational games?**

One of our major objectives was to create games that are not all question and answer based. Children are not stupid – they can see education coming a mile away! While it is true that these Q&A games make evaluation of learning much easier, they just aren't as much fun and, due to their perceived educational content, may not be as popular.

Once our research had been completed we prepared a list of possible games, just ideas, rough storyboards that could hopefully be made a reality. The game ideas were based on information held on the site, topics that would be appropriate to the National Curriculum, a certain level of game playing familiarity and, most importantly, a learning objective, be it information, a skill or a greater understanding of scientific principle.

### **Were there any problems or obstacles that hindered the design and implementation of the games?**

Thankfully no! We had very clear ideas about what we wanted and both Simon and CompleteControl worked very closely with us, discussing every stage of development so it was an amazingly painless process.

### **Naturally, the accuracy of the information presented is critical to an educational game. How was it decided what information to include in the games and how did you handle presenting information in a way suitable for children without 'dumbing it down' or risking its accuracy?**

Obviously, as the educational aspects of the games were as important to us as the fun of playing the game, they had to be

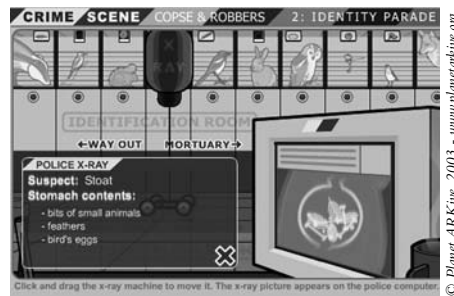


accurate in every aspect, not just the textual information but also the way the animals looked and behaved. All of the information had to be presented in a fun and interesting manner but without losing its integrity, which isn't an easy balancing act.

The species texts were the basis of all of our games. We spent time discussing which species to use and then how to present the information. For a heavily text-based game such as Animal Survival (<http://www.planetarkive.org/games/animalSurvival.html>), we had to condense the information and make it accessible. For Design A Habitat (<http://www.planetarkive.org/games/designAHabitat.html>) we had to get all of the information about black-footed ferrets across to the user so that they could play the game but without giving them reams of text to read – during evaluation sessions we noted that children often will not take the time to read instructions or information, preferring to click away and get on with it. This gave us the idea for the old slide show with voice over; children are used to listening to information and stories, it's a really great way of getting your point across.

Once a game was in its final stages of development, it was circulated around the rest of the team for comment. The text authors and the media researchers are all really familiar with the species and they would be the first to tell us if we got something wrong! In other games such as Copse and Robbers (<http://www.planetarkive.org/games/copseAndRobbers.html>), we let the environment of the game give the information initially and then gradually revealed more information as the game went on. For example, the X-ray machine gives children an insight into a species' diet.

**The games currently available on Planet ARKive range from online versions of simple word games and puzzles to investigative or 'God' role-playing. How do these different approaches affect the learning experience?**



*Users can drag an X-ray machine which describes an animal's stomach contents, helping children decide which 'suspects' are carnivorous*

When we were looking at the type of games we wanted to include on the site we were aware of the need for a certain level of familiarity for children. For some users, being confronted with an unfamiliar game can be off-putting. That is why we included word searches and sliding puzzles – every child is familiar with them. The word searches were also included as we knew teachers might find these useful when teaching certain topics. While using a word search a child may come across a word that s/he is unfamiliar with and go on to find out more about it. The words are grouped, demonstrating the relationship between the words and implanting that relationship in the mind of the user.

The animal survival game is designed as a quiz – it is a game that is easy to evaluate from a learning point of view as you can ask the children the questions before they use the game and again once they have completed the game, thus finding out what they have learned – if anything! There is only one correct answer and an explanation of that answer, which reinforces the learning objective.

Evaluating the learning experience of a game such as Design a Habitat or Copse and Robbers is much harder. Copse and Robbers was designed to show children that you do not need to see an animal to know that it is there. All of the clues in the woodland scene could be found in a typical woodland area, but you need to be observant. Observation is the basis of science and therefore a key skill to develop. It

also demonstrates how an ecosystem can support many different types of animals and even touches on food webs – there are several possible suspects in the case, all animals that could have killed a wood pigeon. At the end of the game, the 'murderer' is released because they didn't actually do anything wrong, they were only following their natural instincts, hunting and eating in order to survive, which is learning on a different level.

*Design a Habitat* was intended to demonstrate the practical aspects of animal conservation and to tell a conservation story from start to finish. It is based on fact; black-footed ferrets were brought to the brink of extinction by the actions of humankind, but were also saved by humans. The learning objectives were to show how something as simple as farming could endanger a species and to show that something can still be done to save species. A conservation message shouldn't be all doom and gloom!

**You mentioned evaluation of learning objectives – have you conducted any other evaluation of how the games are used, or how effective they are?**

As we are in the very early stages of Planet ARKive, so far the only evaluations carried out were prior to launch to ensure that we were heading in the right direction. We tested games with three schools in the local area, beginning by testing the texts with students. We evaluated readability and style, to ensure that all of the students could read and understand the texts and then looked at how the texts would be presented to the users. Most children go to an information source with a specific question in mind, so the texts were written as answers to general questions such as: What does [the animal] look like? Where does it live? How does it have babies? And so on.

Once we had got to a working stage, we went back to the schools to evaluate design and navigation of the site. The kids were very honest about what they liked



© ARKive 2003 - www.arkive.org

*Some of the children who helped test the site at the ARKive launch*

global project after all. The design was exceptionally important and it took a long time to get it right. Navigation has to be almost intuitive – we have all visited sites where the information is great but it is so hard to find that you end up on an easier-to-use site even if the information is inferior.

Content is ultimately what matters. The main ARKive site has scientifically correct and authenticated texts, but these are written in an inclusive way – you don't have to be a zoologist to understand them. If you do come across a word that you don't understand there is an integral glossary so you don't have to go out and buy a zoological dictionary to access the information.

On Planet ARKive and ARKive Education, the information has been aimed at a very specific target audience, so it is easy for them to digest and hopefully to learn.

More information about ARKive joining The National Grid for Learning's Internet portal can be found at: [http://www.arkive.org/news/ngfl\\_public.html](http://www.arkive.org/news/ngfl_public.html).

**Games are just one part of Planet ARKive – could you describe how the other features tie in with the education policy?**

The main aim of Planet ARKive was to reflect the main ARKive site – to provide information and media on endangered species, both those in Britain and worldwide. Therefore the education policy was centred around the repackaging of the main site. Fact Files presents facts, figures and media giving a range of information on endangered species; Fun Facts was included to try and pique users' interest in a species so that they would go to the fact files to find out more. Knowing that children love the messier side of animals – blood, guts and bodily functions – that's what we look for in compiling the facts, snippets that we know children will love. The Creature Feature is designed to high-

and didn't like and we took one of the designers with us so that they could hear what the kids thought about their work. Watching the children use the games is a great way of finding out where you have gone wrong – it is also very interesting to see the differences between the way boys and girls use the games. For instance, when we took the first working version of Cope and Robbers into schools, the boys couldn't wait to get to the police station and didn't bother to collect any clues! Simon had to change the game so that you could only go to the police station when you had at least three clues. The girls, however, worked very methodically and were keen to ensure that they had all of the clues before heading off to the station.

We are now looking at some more formal evaluation of the sites and how they are used and hope to start conducting this work, in conjunction with a local university, very soon.

**Throughout the ARKive Website and publicity, parallels are drawn between the extinction of species and the extinction of data. Does the educational material contain references to the importance of archiving and preservation or does it concentrate on wildlife issues?**

At the moment Planet ARKive deals with science and wildlife issues. As it is aimed at a relatively young audience, we felt it was important to deal with information that the students would find useful for school or their own interests. However, we are planning another section of Planet ARKive, which will give more general information and will include articles on the importance of the collection and preservation of data in all of its forms there. After all, we only know that these species are endangered because someone has bothered to go out and collect data and then others have collated that data over many years to give us an accurate picture of what is going on in the environment around us.

**ARKive was recently recognised by the National Grid for Learning as its 1000th site of educational excellence. What particular educational features do you feel stand out in terms of facilitating learning at all ages?**

For me, all aspects of the site from design and navigation through to content have an impact on education in its broadest sense. For people to use the site on a regular basis and to gain information from it the design has to be attractive across age, gender and cultural background – this is a



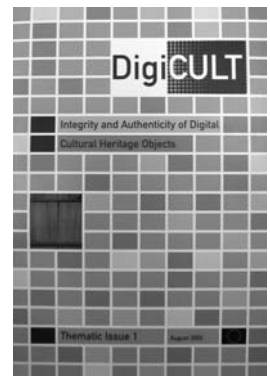
light some of the lesser known but no less interesting species on the site. Megavertebrates such as the elephant and the tiger will be looked at immediately as they are so familiar and popular. However, we want children to be equally as enthusiastic about some of the smaller beasts found on the site – ant lions may not be as well known as an African lion but in some ways they are far more fascinating.

A very important section is *What Can I Do?* As ARKive is not a physical place users can visit, nor a campaigning organisation as such, we felt it was important to

give children the opportunity to do something positive to help wildlife, even if it is just something as simple as feeding birds. So this section gives them practical activities to do and organisations they can contact if they would like to do more.

Planet ARKive is continually developing. New games which will be added soon include Egg and Spawn Race – help an Atlantic Salmon to migrate and breed, Tripwire of Terror – be a male ladybird spider and convince your mate you are not her next meal, and Plainkiller – the Copse and Robbers sequel.

[BACK TO PAGE 1](#)



*A discussion of computer games within the cultural heritage sector begins on page 149 of the first DigiCULT Technology Watch Report: [http://www.digicult.info/downloads/twr2003\\_01\\_low.pdf](http://www.digicult.info/downloads/twr2003_01_low.pdf)*

## AN INTRODUCTION TO APPLICATION SERVICE PROVIDERS

**Dr Thomas Finholt**, Director of the Collaboratory for Research on Electronic Work (CREW) at the School of Information, University of Michigan, introduces the concept of Application Service Providers (ASPs) and explains how they can benefit non-profit organisations.

Application Service Provision is the use of computer networks to access applications (e.g. database, electronic publishing or office productivity software) and resources (e.g. data storage, Web servers, computers) on a subscription or ‘per use’ basis rather than through direct ownership. This results in the overhead costs of expensive information technology and personnel being borne by an independent entity (the ASP) while the services are distributed to many customers – theoretically providing higher levels of service at a lower cost than owning the software to each customer.

Non-profit organisations typically have scarce resources for information technology equipment and personnel. As a result, most are unable to exploit advanced

systems and applications and are therefore forced to use obsolete software and hardware. For example, many non-profits struggle simply to maintain current licensed versions of basic office productivity applications (e.g. Microsoft Office). The ASP model suggests that non-profits can obtain access to current applications, expert advice and technical support by contracting with an ASP – and the ASP can distribute the costs of purchasing and maintaining software and systems across many customers. In some ways, this is similar to the ‘service bureau’ model – few organisations maintain an extensive print production capacity but, when needed, this facility can be purchased from a bureau. Hence, even small organisations are able to produce professional quality publications at a cost below ownership and maintenance of dedicated printing facilities.

The more standard the application requirements, the more likely it is that an ASP can identify a sustainable market (i.e. enough subscribers doing sufficiently similar work so that the ASP can realise economies of scale). For instance,



*Dr. Thomas Finholt*

*“An ASP attuned to the special requirements of non-profit organisations will be better able to bring the benefits of application service provision to the non-profit sector where there is a need and a desire to apply modern information technology but resources are often strained.”*

many cultural heritage non-profit institutions (e.g. museums) are likely to require some kind of donor management system – a large and fairly standard application niche. The larger and more standard the application niche, the more customers can be served with that application and the more efficient the Application Service Provision solution will become.

However, most ASPs are profit-making ventures and therefore orient towards the biggest and most lucrative markets. These companies may have little



interest in or sensitivity to the specific needs and tight budgets of non-profit organisations. This situation has led to attempts to create non-profit ASPs to exclusively serve other non-profits. One notable example is the experiment supported by NPower (a non-profit IT consulting organisation founded in Seattle: <http://www.npowerseattle.org/tools/aspower.htm>).

The hope is that an ASP attuned to the special requirements of non-profit organisations will be better able to bring the benefits of application service provision to the non-profit sector where there is a need and a desire to apply modern information technology but resources are often strained.

The Collaboratory for Research on Electronic Work (CREW: <http://www.crew.umich.edu>), a research unit within the School of Information (http://www.si.umich.edu) at the University of Michigan, has conducted two studies on the use of ASPs in the non-profit sector. The first study, completed in 2001, examined attitudes toward ASPs by executives of grant-making non-profits (e.g. philanthropic foundations) in Michigan. The sample included several hundred organisations and found interest in the ASP concept but also concerns about the costs of infrastructure (e.g. high-

speed network links) and data security. A second study, completed in 2003, examined the deployment of applications through the NPower pilot ASP to a dozen social service non-profits in the Detroit area. This study involved the intensive interview and collection of observational data of non-profit employees using the NPower ASP applications. Results suggested that the most successful ASP product was access to an administered file server while mission-specific software was less successful, primarily due to poor training and a lack of correspondence to the idiosyncratic needs of the target non-profits.

Our research suggests that medium-sized non-profits, with staff of between twenty and fifty people, are probably the best target for application service provision. Medium-sized non-profits are big enough to need efficient information technology and support but are not big enough to carry the costs of the equipment and staff, whereas larger non-profits frequently have the resources to maintain an in-house information technology operation. Smaller non-profits often must put their mission focus ahead of everything else (where the modernity or efficiency of information technology can be the least of their problems).

The main theoretical impact of ASPs is the use of more modern applications, with more expert advice, by more organisations. This will lead to a corresponding increase in the ability of organisations to focus on their missions, clients and customers, while simultaneously reducing overall costs. When combined with open source software movements (e.g. Linux), the ASP model becomes an ideal way for previously disparate non-profits to pool their purchasing power to obtain a level of service and performance that none could approach individually. The rub, however, is the assumption that a population of non-profits can identify a set of strongly over-

lapping needs that can be met by an ASP. That is, the ASP will not be viable if it must absorb the individual variance of each client organisation, such that all costs are transferred to the ASP without any corresponding economies of scale.

During the dot.com bubble, there was great enthusiasm for the ASP model as a general approach to application and resource provision. Since the collapse of the bubble, thinking about ASPs has become less exuberant and more rational. A key stumbling block remains the need to have high-speed, high-reliability network connections to enable the ASPs to deliver services and resources seamlessly to customers. For some organisations, it is possible that the potential savings represented by the ASP model may be lost through increased networking costs. Where such network links can be provided easily and at low cost, the ASP option remains viable, assuming a sufficiently large and generic market exists. Whether non-profit organisations meet these criteria is still unknown.

ASPs in the cultural heritage sector will be the subject of a full investigation in the next DigiCULT Technology Watch report, due to be published early in 2004.

[BACK TO PAGE 1](#)

The reports from these studies are available from the CREW technical reports Website at [http://www.crew.umich.edu/technical\\_reports.htm](http://www.crew.umich.edu/technical_reports.htm).

Finholt, T.A. & Lu, M., 2001:

*"Requirements analysis for application service provision among Michigan-based non-profit organisations."*

Shah, V. & Finholt, T.A., 2003:

*"Evaluation of application service provision pilot in Southeastern Michigan non-profit and social service agencies."*



## VALUE CHAIN FOR DIGITAL LEARNING CONTENT

A new document, “The Value Chain for Digital Learning Content in England”, is now available for downloading from: [LINK TO PDF DOCUMENT](#)

The Value Chain was created by Simulacra (<http://www.simulacrame.com>) for the Department for Education and Skills in the UK, and is a conceptual model that is intended to help organisations to work out how they fit into the process of creating, publishing, managing, distributing and using digital content. It aims to provide a commonly understood model and language for the Digital Learning Content market and represents the value that is added to digital learning content through the key stages from conception to use. The key roles undertaken by organisations involved in these stages are also examined. [BACK TO PAGE 1](#)

## TRIALS SUPPORT FOR EUROPEAN PROJECTS

TRIS (Trials Support) co-ordinates and promotes the results of twenty-five recently completed EC-funded trial projects which tested the application of new technologies to enhance access to collections held in museums, libraries, and archives. TRIS has published material to publicise the results of these projects on its Website at: <http://www.trisweb.org/tris/trisportalpro/material/default.asp>. The Website is regularly updated with news and events related to cultural heritage, and has a new section on European and international projects which can be found at: <http://www.trisweb.org/tris/trisportalpro/related/default.asp>. Additions to this section are welcome. Information on TRIS can be accessed in English, French, German, and Italian from: [http://www.trisweb.org/tris/trisportalpro/tris/tris\\_project\\_sum.asp](http://www.trisweb.org/tris/trisportalpro/tris/tris_project_sum.asp). [BACK TO PAGE 1](#)

previously restricted development and research in this area. However, the computational demands of language processing can now be met in many parts of the world and the full potential of language technologies is finally being realised.

HLTs can be classified in various ways but are generally divided into two broad areas: natural language processing (NLP) and speech processing. Broadly, speech processing refers to the medium of language – the human ability to hear or produce spoken language – whereas NLP refers to the message, that is, comprehension of the meaning of what is being said.

### SPEECH PROCESSING

Speech processing can be further subdivided into different ways of recognising and producing speech. For a computer to convert spoken language into text, it needs to isolate the words used (e.g. from background noise, speaker elision) and ‘recognise’ them, that is, acoustically analyse vocal sounds and produce the appropriate signifier of written language which represents it. This requires the machine to not only ‘know’ all of the words being spoken, but also to be able to ‘hear’ correctly each individual word and, in the case of homonyms, understand which version of that sound (e.g. right, write, rite) is being used. In addition to converting speech into text, however, there are additional analyses that can be performed on vocal input, such as speaker recognition/verification, identification and classification.

**Speaker recognition/verification:** This technology allows a computer to verify the identity of a person based on his or her voice. The user typically repeats a certain phrase and the computer compares this with saved sound files either to identify which of its saved speakers is most likely to be speaking (recognition) or to confirm whether the speaker is a specific individual (verification). Research continues into

# HUMAN LANGUAGE TECHNOLOGIES EXPLAINED

**DAISY ABBOTT** DigiCULT Forum Researcher

In our global village, the convergence of information society technologies has created a demand for both universal communication and the infrastructure that can support it. Human Language Technology (HLT) is an umbrella term for the innovation, research and development which can help human beings to interact with and through computers in a more natural way, that is, using their own languages. These technologies can increase ‘understanding’ between humans and machines and can help people communicate with one another, removing the barriers to communica-

tion inherent in disparate languages while preserving the cultural significance of our multilingual world.

*“Human language technologies will eventually allow anyone to use a computer simply by talking and listening to it.”*

The requirements of HLT (powerful computer processing capacity, high speed and widespread connectivity) have



text-independent speaker recognition where users could even speak in a different language and expect to be recognised.

**Speaker identification:** this technology overlaps with speaker verification, but instead of a binary state (the voice either is or isn't the person the computer is comparing it with) speaker identification can isolate from a group of enrolled speakers which one is currently speaking.

**Speaker classification:** this is the ability to perform analyses of the voices of a group of unknown speakers and to perform generic tasks based on their voices. Tasks could include identifying similar-sounding speakers (e.g. regional accents), highlighting all speech segments made by the same person, or detecting when the speaker changes.

Speech processing also includes speech output and its many applications. Obviously, there are no issues of background noise when a computer converts digitally stored text to speech; however, pronunciation of heteronyms (words that are spelt the same but can have different pronunciations and meanings, such as bow, tear) becomes an issue.

## NATURAL LANGUAGE PROCESSING (NLP)

This technology will allow a computer to understand the meaning inherent in language, that is, not only to be able to convert spoken words into text but to convert the linguistic meaning of the sentence into a form it can understand and on which it can perform actions. This transformation of natural language into formal language is known as parsing and is one of the most technically challenging aspects of HLT.

Like speech recognition, the more limited the number of possible meanings, the easier it is to achieve a good success rate with a natural language processing system. For example, a travel booking system

which uses NLP would probably only have to deal with a small variety of requests ("Give me information", "Book", "Change", "Cancel") and be able to recognise destinations. A system similar to this one is being developed by IBM (details can be viewed at: <http://www.research.ibm.com/hlt/>) which can process travel bookings to 9000 destinations given verbal commands in either English or French.

## HOW DOES HLT WORK?

HLT requires certain components of language in order to function. A lexicon or corpus provides a large amount of raw linguistic data – a vocabulary – then grammatical rules or statistical analyses are applied to determine the likelihood of a variety of meanings of the piece of language. To build an accurate language model, computers use this huge amount of data both to recognise individual sounds and to predict the most likely version of the word being used. For example, based on a statistical analysis of samples, a computer will know that "There is a ..." occurs much more commonly than "Their is a..." or "They're is a..." and can use this information to identify which spelling is most likely to be correct. (In fact, as the previous sentence was typed, the word processing package automatically corrected the two 'wrong' sentence segments.)

One current method of natural language parsing is to produce a language model comprising each original 'natural' sentence linked to its corresponding meaning and parsing. This allows computers to identify correlations between more complex linguistic structures (word orders, irregular grammatical constructions) and intended meanings.

## HOW CAN HLT BE USED?

Ideally, HLT will eventually allow anyone to use a computer simply by talking and listening to it. There is huge potential for the applications made possible by reliable

HLT. Some search engines can use a level of linguistic intelligence to retrieve results that do not exactly match the search parameters but include extensions of words or even synonyms. Computer translation is one of the most used (and most difficult) applications for natural language processing. Language 'pairs' are used to match vocabulary in two different languages; however, it is clear that understanding word context and grammatical idiosyncrasies is of utmost importance in this area. One example of a cross-lingual application is a search engine that retrieves and translates pages that are not in the language of the search parameters. Cross-lingualism can be applied to speech technology as well as NLP and there are now several prototype cross-lingual speech interfaces which allow speakers of different languages to understand one another. Speech recognition is already used in telephone booking and transaction systems and works effectively in low-vocabulary contexts. Speech synthesis has been used for a number of years to ease communication from and to people with sensory impairments (e.g. voice synthesis for people who have difficulty speaking, text being 'spoken' by the computer for the partially or un-sighted) and is now being used in applications such as listening to e-mail.

Speaker recognition or verification has multiple applications, from assuring a person's identity in telephone banking systems or cash machines to increased security on locked workstations, PDAs.

As some of these examples demonstrate, the result of improvement in HLT is that the interface between human and computer will begin to blur. Instead of having to learn a programming language to tell a computer exactly what you want it to do, the computer may understand your language. Machines will become more transparent, allowing humans to interact with them in a more natural way.



## HLT IN EUROPE: THE EUROMAP LANGUAGE TECHNOLOGIES PROJECT

The applications of HLT to an advanced economic yet linguistically diverse area such as Europe are obvious. There are already eleven official languages in the EU<sup>1</sup> – not counting regional languages (e.g. Catalan), non-official national languages (e.g. Welsh) and immigrant languages (e.g. Urdu in Britain, Turkish in Germany) – and with thirteen more ‘applicant’ countries, this number will continue to grow. Multilingualism has become almost a necessity, particularly for international business.

European countries, while generally willing to integrate into a multinational Europe, are nevertheless keen to preserve cultural and linguistic diversity, to maintain their own languages alongside communication at an international level. The transparency of language is therefore vital to enable all citizens to access information provided at a continental level.

Developing HLT for the many different European languages is an extremely complex task; however, this difficulty itself encourages research and development of one of the most important and influential communication technologies of the future. Several European projects have begun research programmes that will improve HLT, including the recently completed EUROMAP project. (For many more projects dealing with HLT issues and technologies, please visit DigiCULT’s page of natural language processing links: <http://www.digicult.info/pages/links.php?t=16>).

### EUROMAP

EUROMAP (<http://www.hltcentral.org/page-56.shtml>) investigated the take-up and use of state-of-the-art HLT in European countries. As part of this work it compared current research and take-up

on a country-by-country basis. The study identified European HLT ‘leaders’ and suggested how other countries could develop in the future. Key to the report’s conclusions was the relationship between the European Research Area and investment in HLT and the developments made.

Recommendations included the transition from (currently) somewhat geography-specific, inward-facing studies into a pan-European technology level of language equality, and the creation of a Language Technology Agency to moderate the funding and

EUROMAP came to an end in April 2003 and the final report is now available online from: <http://www.hltcentral.org/page-243.0.shtml> (you can obtain a free printed copy of the report by contacting the National Focal Point of your country details from <http://www.hltcentral.org/page-59.shtml>).

provision of language resources.

The report found that it is increasingly common to find a multi-language focus in European HLT research and that, as this community becomes more integrated, individual language expertise can be shared across the EU, encouraging multinational ownership without severing ties with national language communities. Those nations who have benefited from consistent long-term funding (from public bodies, higher education institutes or national research organisations) have become leading innovators in HLT with approaches often mirroring local priorities. However, to create a truly multilingual HLT policy across the EU, there must be significant further public investment in bringing HLT for all languages to a similar level – market forces alone are unlikely to provide the necessary resources to achieve this relative equality of development.

The Language Technology Agency (LTA) would

identify those areas of research that need to be accelerated and help to ease the transition of current knowledge and technologies across linguistic borders to prevent duplication of effort. HLT has increasingly refocused on engineering solutions to HLT problems, therefore a central agency would enable technical solutions to be expanded, maximise the effect of contributions from the investment community, ease the process of transferring HLT to the marketplace, and allow these solutions to encourage new theoretical research.

It is hoped that an EU-wide marketplace for HLT research and development may encourage the identification of ‘best practice’, transferable technical solutions, models and architectures that can be applied to any national language. Only with a robust infrastructure such as this can the research and development already performed be fully exploited to eliminate barriers to communication across Europe.

### HUMAN LANGUAGE TECHNOLOGY CENTRAL

HLTCentral (<http://www.hltcentral.org/>) is a dedicated server providing a gateway to speech and language technology opportunities and related topics of interest to the HLT community. The Website includes breaking news, new developments in research, technology and business in the field of speech, language, multilinguality, automatic translation, localisation and related areas. Its coverage of HLT news and developments is worldwide – with a unique European perspective.

[BACK TO PAGE 1](#)

<sup>1</sup> Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, and Swedish.

The DigiCULT Forum will consider issues relating to the use of Human Language Technologies in future publications.

## CONTACT THE PEOPLE BEHIND DigiCULT FORUM

Salzburg Research  
Forschungsgesellschaft m.b.H

salzburg|research

HATII – Humanities Advanced Technology and Information Institute



### DigiCULT Project Manager

John Pereira, [john.pereira@salzburgresearch.at](mailto:john.pereira@salzburgresearch.at)

### DigiCULT Thematic Issue Scientific Editor

Guntram Geser, [guntram.geser@salzburgresearch.at](mailto:guntram.geser@salzburgresearch.at)

### DigiCULT Network Co-ordinator

Birgit Retsch, [birgit.retsch@salzburgresearch.at](mailto:birgit.retsch@salzburgresearch.at)

### Director, HATII & ERPANET

Seamus Ross, [s.ross@hatii.arts.gla.ac.uk](mailto:s.ross@hatii.arts.gla.ac.uk)

### DigiCULT Forum Technology Assessor

Martin Donnelly, [m.domnelly@hatii.arts.gla.ac.uk](mailto:m.domnelly@hatii.arts.gla.ac.uk)

### DigiCULT.Info Content & Managing Editor

Daisy Abbott, [d.abbott@hatii.arts.gla.ac.uk](mailto:d.abbott@hatii.arts.gla.ac.uk)

### DigiCULT Web Developer

Brian Aitken, [b.aitken@hatii.arts.gla.ac.uk](mailto:b.aitken@hatii.arts.gla.ac.uk)



# THE SECOND CONFERENCE OF THE NATIONAL DIGITISATION CENTRE OF SERBIA AND MONTENEGRO: DIGITISATION FROM A BALKAN PERSPECTIVE

**DR MILENA DOBREVA**

(Institute of Mathematics and Informatics, Sofia, Bulgaria) and

**DR. ZORAN OGNJANOVIĆ**

(Mathematical Institute Belgrade, Serbia and Montenegro)

The Second National Conference on New Technologies and Standards: Digitisation of National Heritage in Serbia and Montenegro was held on 29-30 May in Belgrade. Thirty-two papers dealing with current achievements of museums, libraries, archives, as well as the new technologies and their possible adoption in the digitisation of cultural and scientific heritage sector were presented over two days.

The last decade brought many challenges to people in Belgrade – a city with turbulent history. Its cultural heritage, which has survived numerous destructive events, is treated with much care in Serbia and the advent of digital methods was met here with significant interest in the 80s and 90s. As in many other places, the first applications were targeted to organising in better ways the knowledge on the content of the collections.

The conference demonstrated that the interest to various applications of information technologies in the cultural and scientific heritage sector already provides real results. Papers presented library/archive applications, work done in museums, audio archives, work on archaeological sites, encoding standards and new technologies with ideas how to apply them in the cultural and scientific heritage sector. Two CD-ROMs were presented, as well as one book. The paper on rights management presented by Jovan Krstić was met with much interest.

<sup>2</sup> Although the first saved document is dated from 1022, successive bookkeeping started in 1278.

## ONE EXAMPLE: DIGITISATION OF THE MICROFILMED ARCHIVE DOCUMENTS FROM DUBROVNIK

One of the most interesting local projects, Digitalisation and processing the Dubrovnik archive documents from Serbian Academy of Sciences and Arts archive and private collections, was presented by Alexandra Fostikov and Tibor Zivkovic from the Institute of History at Belgrade. Archive documents from Dubrovnik, now in Croatia, are important resources for understanding Balkan and Mediterranean national history and their relations in the period from 1200<sup>2</sup> to 1500 AD.

Dubrovnik archive documents in Serbian Academy of Sciences and Arts archive are kept on microfilms made during the 1950's and since the records are old, they are in great danger. The clarity of projected images is getting worse day by day, pictures are fading and the recorded material loses readability. In order to preserve the precious documents, catalogue them more systematically and increase accessibility for researchers, the Institute of History in Belgrade (<http://www.hi.sanu.ac.yu/>) started the long process of digitisation of the archive documents.

The project which is still under development, comprises the following steps:

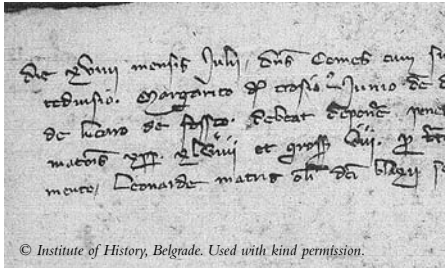
- scanning old microfilms and transferring them into CD and DVD format;
- transferring old microfilms on diazo films;

- cataloguing and archiving;
- transliteration, translation and commenting the documents;
- creation of database and internet presentations;
- further co-operation and exchange.

Canon MS800 Microfilm scanner was used to digitise the microfilm collection. The software development, the database with digitised documents and following presentations will be done in co-operation with the Mathematical institute in Belgrade.

The digitisation of the Dubrovnik archive documents and creation of the database and Internet presentations will enormously increase accessibility to this archive. It also opens up possibilities for all researchers interested in Balkan and Mediterranean history to use these documents.

This project is quite typical for the region; presenting to a wider audience an endangered collection with regional importance. It is a good example of co-operation of two institutions, one of them serving as a content provider (The Institute of History), with another supplying the necessary software support (The Institute of Mathematics). Presenting such work boosts other organisations to consider digitalisation options. We can definitely expect that next year's conference will bring together more examples of real life projects.



© Institute of History, Belgrade. Used with kind permission.

#### Archive sample

Our colleagues from Serbia and Montenegro demonstrate a strategy suitable for smaller countries. Instead of building digitisation teams in every organisation in the cultural and scientific heritage sector, they are aiming at forming one national institution which would support initiatives undertaken by different bodies. Recently, experts from the Belgrade Mathematical institute, National Museum, Faculty of Mathematics, and Archaeological Institute, along with the Archives of Serbia, National Library of Serbia, Serbian Institute for Monument Protection, and Yugoslav Film Archive formed the National Center of Digitization (NCD, <http://www.ncd.matf.bg.ac.yu/>). Bringing together these institutions means that they can share experience. For example, Mathematical Institute and Faculty of Mathematics Belgrade have launched sev-

eral projects aimed at collecting and presenting data on cultural content in digital form. This experience should contribute to building realistic strategies and standards within the National Digitisation Centre.

The NCD still expects to become a governmental centre; the group of people behind it work to promote best practices on national level and to bring people from various institutions, who have to solve similar problems, together.

The main activities of NCD are as follows:

1. Co-ordinating efforts of institutions in Serbia and Montenegro involved in the cultural and scientific heritage digitisation.
2. Establishing and promoting a national strategy for the cultural and scientific heritage digitisation in Serbia and Montenegro.
3. Exploring, adapting and applying international standards and protocols for the cultural and scientific heritage digitisation and preservation, as well as developing the new standards in the areas where none exist.
4. Launching the cultural and scientific heritage digitisation in the country and making plans for possible migration

process to new formats and technologies for already digitised data.

Methods of disseminating knowledge include publishing a journal (Review of the NCD), organising an annual conference and maintaining a website.

This experience, unique for the Balkans, can make a positive change in the country, and the interest of conference participants is clear. This conference also hosted two Bulgarian papers. Next year, the event is expected to become truly international. The Lund principles, which define the priorities in the digitisation field on European grounds, put the emphasis on better organisation and access of collections considered to be already available in a digital form. In this respect, most associated and other non-EU states in Europe still have to come a long way, since most local collections are not digitised. Co-operation and dissemination of knowledge in this field are vital in order to boost local initiatives. We hope that the work of the NCD will lead to a significant change for the better in the long run.

[BACK TO PAGE 1](#)

## CULTURAL HERITAGE EVENTS

DigiCULT's Website provides a full calendar of events from conferences to training workshops. Anyone can use this list to publicise conferences relevant to the cultural heritage sector; to submit an event, view all upcoming events and see more details, please visit <http://www.digicult.info/pages/events.php>

Upcoming events of particular interest include:

The EVA series of conferences has two upcoming events: Berlin on 12-14

November 2003 (<http://www.gfai.de/pinboard/eva/>) and Moscow on 1-5 December 2003

([http://www.evarussia.ru/eva2002/eng\\_str/EVA2003MoscowE.html](http://www.evarussia.ru/eva2002/eng_str/EVA2003MoscowE.html)).

#### DRH2003

Digital Resources for the Humanities 2003 took place from the 31st August – 3rd September on the beautiful Park campus of the University of Gloucestershire in Cheltenham. This year's conference bene-

fited from a packed programme of diverse papers, panel discussions, and poster presentations and the social programme added to the friendly atmosphere. Final plenary speakers Dr. Kim Veltman and Prof. Theodor Nelson were both entertaining and intellectually exciting, providing an excellent close to an enjoyable and beneficial conference.

A full report on DRH 2003 will appear in Issue 6 of DigiCULT.Info.



## DIGICULT SEEKS NATIONAL CORRESPONDENTS

DigiCULT.Info would like to expand our coverage of cultural heritage projects, issues and research in countries where English is not the native language. We would therefore like to appoint several honorary national correspondents who would seek out relevant news and projects, and write or source articles in languages other than English. If you are interested in a position like this, please contact Daisy Abbott at [d.abbott@hatii.arts.gla.ac.uk](mailto:d.abbott@hatii.arts.gla.ac.uk)

[BACK TO PAGE 1](#)

## NETARCHIVE.DK

Netarchive.dk have recently made available an English version of their report detailing the project's work in preserving the Internet. Various strategies for collecting and archiving of Internet material were tested and the Netarchive.dk report recommends a strategy based on a hybrid of selective harvesting and bulk harvesting. The project was a collaboration between the Royal Library, Copenhagen, The State and University Library, Aarhus, and Centre for Internet Research at the University of Aarhus and the Danish

Electronic Library initiative. Many articles discussing issues surrounding preservation of the Internet are included in the proceedings from the 2001 conference "Preserving the Present for the Future - Strategies for the Internet" <http://www.deflink.dk/arkiv/dokumenter2.asp?id=695>

[BACK TO PAGE 1](#)

The full report of the netarchive.dk project can be downloaded in either Danish or English from the project Website: <http://www.netarchive.dk>

## PRESERVING ELECTRONIC RECORDS FROM DATABASE-DRIVEN INFORMATION SYSTEMS

**FILIP BOUDREZ**  
**researcher DAVID project**  
City Archives of Antwerp

The DAVID project (<http://www.antwerpen.be/david/>) developed an electronic record-keeping system for dynamic and interactive information systems. This article describes the basic principles and concepts of this record-keeping system and the way it has been implemented by the City Archives of Antwerp (<http://stadsarchief.antwerpen.be/default.asp>), the archival partner in the DAVID project.

### THE DAVID PROJECT

The DAVID project (DAVID is the Dutch acronym for Digital Archiving in Flemish Institutions and Administrations) is the first Flemish research project on electronic record-keeping. The aim of the project is to examine how electronic records created by Flemish institutions and administrations can be archived within their context in a durable and authentic way. The DAVID project runs over a period of four years and will prepare a manual

on electronic record-keeping by the end of 2003. The project is aimed at electronic office documents (e.g. word processing files, spreadsheets, e-mails) as well as archiving records which were created in dynamic and interactive database-driven information systems. A record-keeping system has been developed for both kinds of documents.

### THE INFORMATION SYSTEM AS A STARTING POINT

One of the initial goals of the DAVID project was to design a typology of electronic records that could serve as a basis for record-keeping strategies. Mapping of all the information systems that are used within the administration of the city of Antwerp led to the conclusion that a perfectly sound typology could not be formulated to suit every situation; such a generalised system would be effectively useless. Important questions as to what is being archived and how this is done can

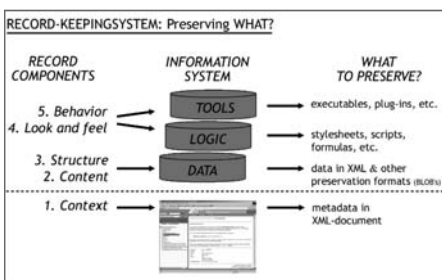
only be answered when the archivists have a lot of information about the information systems in which electronic records are created and managed. Functionalities, architecture, links to other information systems, and organisation of data are some of the important parameters that can influence the record-keeping strategy. Consequently the starting point for a record-keeping system is the information system itself.

Therefore, the archivist needs data on the information system before s/he can undertake any record-keeping action. However, metadata on information systems are not always stored in a systematic or structured way. Important information is often considered to be irrelevant or is only implicitly present. At the time of record creation the archivist may discard data or attempt to store it only in a non-permanent way, leading to the discovery, too late, that this information was important. It



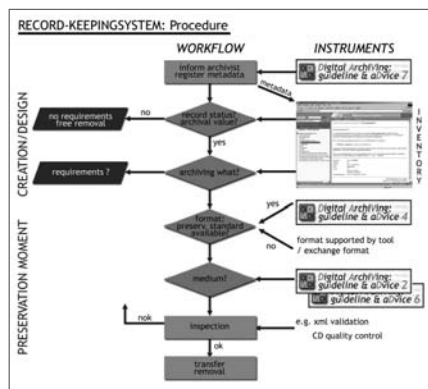
goes without saying that this is too flimsy a basis for important decisions such as the identification of the records, appraisal, and the describing of a record-keeping system.

To counter this problem, a new archival instrument was created: the information system inventory. In this information system inventory, civil servants, system administrators and the archivist collect the metadata on the electronic information system starting from the moment of creation. The form in which this information system is kept can vary from a word processing document to a database. For example, the information system inventory of the city of Antwerp is represented in a relational database with Web interface and dynamic data model. Such an information system inventory can achieve other goals as well, for example, a helpdesk function or the management of the IT infrastructure. This information system inventory offers an added value for the whole organisation, meaning that the archivist is not the only party who is interested in and benefits from keeping the inventory up-to-date.



## PROCEDURE: FROM CREATION TO ARCHIVING

From the moment that a new information system is being developed or an existing information system is being adjusted, the information system inventory has to be updated and the archival services need to be informed of these changes. This is a compulsory part of the IT procedure in the city's administration. The archival services examine whether any records with



archival value are being created or managed within the information system. It is important that this is done while the information system is still active.<sup>3</sup> It is at this moment that the decision is made as to what will be archived in the long run and if any special quality demands apply for the information system. These quality demands can refer to the file formats, creating necessary structured and explicit metadata, ensuring reliability, encoding, and will ensure the creation of electronic records which can be 'well archived' over the long term. It is important to know these qualities before the new or adjusted information system goes into operation. The archival service should be consulted preferably when evaluating the new information system or when the quality manual for the new or adjusted information system is being edited.

For the identification of the records and the appraisal, information systems are regarded as a combination of three active components: the content, the logic, and the tools. The appraisal should lead to identification of the components that give the document the status of record. Important considerations are whether the complete database, a part of the data or only the generated output forms the records of the information system. For example, applied to a GIS application, this can mean that the data is archived as GML (Geography Markup Language) documents or that only the maps are archived as image files (for

instance GeoTIFF, SVG). From the application in which the population register is kept, only the population data will be archived as XML documents in the long run. The boundaries of the database are also determined at that time. After all, more and more information systems are linked and borrow information from each other. Whether this 'external' data will be archived alongside information from the system depends on whether the linked database is archived, and on the frequency with which this happens.

The frequency of record-keeping actions depends largely on the way the database system deals with changes. If adaptations are registered without the old data being overwritten, then the archiving frequency will depend on the size of the database and the performance of the information system. Several options are possible with databases where old data are not kept separately but is being overwritten. It is possible either to archive all the basic versions and subsequently all the changes, or to archive 'snapshots' frequently. This last option risks losing the intervening versions. A combination of these methods is probably most appropriate when all versions need to be preserved.<sup>4</sup>

Typical for electronic records is the need for hardware and software to reconstruct the records in the future. So, at the moment of appraisal one should not focus all the attention on the content of the database alone. Logic elements and tools have to be considered for preservation as well. This will usually be the case when the original 'look and feel' and functionalities or behaviours of records need to be preserved. The logic layer of an information system refers to all elements that relate to the handling of the input and the way the output is generated. The tools are the instruments for input and output. So, the identification of the components that will be archived does not depend only on

<sup>3</sup> INTERPARES 1, Appraisal task force report, p. 9-10. (<http://www.interpares.org>)

<sup>4</sup> N. BÜTIKOFER, Archiving snapshots or transactions: extracting the right data at the right time from temporal databases, presentation given at Erpa Workshop 'Long term preservation of databases', Bern, 10 April 2003. (<http://www.erpanet.org>)



the archival criteria, but also on the technological demands to reconstruct the records in an authentic way in the future.

The decision on the file format and the medium that will be used is made at the moment of record-keeping. The city archive of Antwerp has set its archiving standards to the file formats, media and file systems recommended in official guidelines (which are based on the general guidelines and best practices of the DAVID project). The creator and the IT department of the city prepare the transfer of the records together. The records that have been deposited and their carriers are inspected on arrival at the archival service. To validate large XML documents, a validation parser programme was developed. CD quality is tested with the aid of a diagnostic tool; if the deposition does not meet the quality demands set by the city archives, then the electronic records are sent back to the creator so that they can be brought up to the appropriate level. Two copies of each carrier must be deposited: one copy is kept at the archival service while the safety (backup) copy is stored at a separate location. Only once the transfer has been approved is it permissible to remove the records from the original information system.

## XML PRESERVATION OF RECORDS

The eXtensible Markup Language (XML) is being used as much as possible as the preservation format when archiving databases with textual data. XML offers interesting benefits for electronic record-keeping of records: easily exchangeable, appropriate for structured textual information, application of an explicit document model, self-describing to a large extent. For databases that contain Binary Large Objects (BLOBs – a large block of data, such as an image or sound file), XML is used as metadata format for archived records.

The migration process of records within databases to XML documents consists of several steps. The first step involves editing a document model for the records. Initially, DTDs were developed for this, but gradually the switch is being made to XML schemas. This document model is based mainly on the inherent structure of the documents. It can be identical to the internal database record structure, but this is not essential. In relational databases the record is often spread over several tables, so joins and queries usually precede the unload or export of the data. One guideline when putting together a document model is to consider the way in which the input, and especially the output, was presented to the user of the active information system.

The mapping of a relational data model to a hierarchical document model is not always obvious. Both data models have a number of fundamental differences. It is possible to put the internal logic of the documents in the archived documents by assuring good nesting and by attributing semantic tags. Therefore exporting data from the database often requires query and join actions. If necessary, stylesheets can be used to demonstrate more explicitly the way in which the records were shown to the users of the active system.

The process of unloading databases through text files deserves special attention because of the encoding of the characters. First, the characters are transferred, preferably to Unicode. The next step is to replace the preserved XML characters by entities and to filter out the invalid XML characters (for instance control characters). This is achieved with a tool specifically developed for the purpose. Finally, the last step is tagging the XML characters and adding the XML declarations. When choosing the tag names it is advisable to choose semantic tags, even though this can lead to redundancy when using large files

and when the same tags appear multiple times. Tagging individual fields offers the advantage that the data from different fields can be addressed individually and that database filters can be eliminated. The use of abbreviations and codes as tag names can offer a partial solution, but this does mean that documentation on the meaning of tag names needs to be kept. Because of this, XML documents lose part of their autonomy and self-reliance.

## METADATA

At the moment of archiving, all of the metadata of the records and of the information system are archived as well. By adding as many semantic XML tags to the database fields and records as possible, important metadata in the record itself is archived. Because of this, there is no need for external documentation in order to find out the function of characters in certain positions within the database record. We keep depending on external documentation to discern the meanings of such abbreviations as tag names or code tables. Essential metadata on the information system are present in the information system inventory. This metadata is exported to an XML document and further replenished with metadata at the moment when an archiving action is undertaken. This XML document is archived at the same time as the records.

The DAVID project highlights the importance of understanding the overall structure of an information system in order to design methods of record keeping which are appropriate to the data stored and which will maintain high quality of both metadata and records for the long term.

[BACK TO PAGE 1](#)

More information on this archiving system and on the DAVID project is available on the Website:  
<http://www.antwerpen.be/david> or by emailing: [david@stad.antwerpen.be](mailto:david@stad.antwerpen.be)

# NEWS FROM UK ARCHIVES

## NEW REPORT ON UK ARCHIVES

Resource, the UK Council for Museums, Archives and Libraries, (<http://www.resource.gov.uk/>) has published online a new report dealing with the effective long-term management of digital records and the current situation in the UK. "Archives in the Digital Age" can be downloaded from <http://www.resource.gov.uk/information/tenders/completed.asp>, project ID 435. The study considers current research and frameworks and the specific needs of archives in addition to obstacles to progress and the legal environment. For more information, e-mail: [emma.halsall@resource.gov.uk](mailto:emma.halsall@resource.gov.uk).

## NEW MOVING IMAGE ARCHIVE RESOURCE

Archival film and television material provides a unique and rich resource for academic study, teaching and research as well as for artistic projects. The UK has twelve public sector moving images archives at both national and regional levels who are dedicated to providing the widest possible public access to all preserved materials and to furthering the use of their collections in learning, teaching and research. Several of these collections are based in Higher Education institutes and the film archive community seeks to maintain its good links with the academic sector.

On this note, a new resource for UK film and video archives has been produced and is now available online. Moving History (<http://www.movinghistory.ac.uk>) is a Website aimed at academic communities in the arts and humanities sectors. It is hoped that Moving History will promote scholarly research into moving image material and archive film in par-

ticular and will increase awareness of the value of the UK's film and video archives in teaching and understanding film.

The Website provides information about all twelve of the UK's public sector moving image archives including access procedures, facilities and services, as well as study tools such as academic research case studies and guidance on understanding the collections. Moving History gives an overview of the content of each collection and features descriptive and contextual data on over 100 selected clips from the archives. It also acts as a gateway to the archives themselves and related organisations and collections.

As a result of research initiated by the Arts and Humanities Research Board Centre for British Film and Television Studies (<http://www.bftv.ac.uk>), a two-year project led by the South East Film and Video Archive (SEFVA: <http://www.brighton.ac.uk/sefva/>) led to Moving History being created. The project sought to address and rectify the issue that underfunding film collections results in a lack of academic research and the consequent under-valuing of these resources as powerful artistic, historical and sociological study tools. A full description of the project can be found at <http://www.bftv.ac.uk/projects/sefva.htm>.

### UK national collections:

- bfi National Film and Television Archive (<http://www.bfi.org.uk/collections/>)
- Imperial War Museum Film and Video Archive (<http://www.iwm.org.uk/collections/film.htm>)
- National Screen and Sound Archive of Wales (<http://screenandsound.llgc.org.uk/>)

- Scottish Screen Archive (<http://www.scottishscreen.com/>)

### English regional collections:

- East Anglian Film Archive (<http://www.uea.ac.uk/eafa/>)
- Media Archive for Central England (<http://www.nottingham.ac.uk/film/mace/>)
- North West Film Archive (<http://www.nwfa.mmu.ac.uk/>)
- Northern Region Film and Television Archive (<http://www.nrfta.org.uk/>)
- South East Film and Video Archive (<http://www.brighton.ac.uk/sefva/>)
- The South West Film and Television Archive (<http://www.tswfta.co.uk/>)
- Wessex Film and Sound Archive (<http://www.hants.gov.uk/record-office/film/>)
- Yorkshire Film Archive ([http://www.yorks.ac.uk/dialect/yorkshire\\_film\\_archive.htm](http://www.yorks.ac.uk/dialect/yorkshire_film_archive.htm))

## NEW RESOURCES FROM THE IAML UK AND IRELAND BRANCH

There has been a lot of recent activity at the UK and Ireland branch of the International Association of Music Libraries, Archives and Documentation Centres (IAML), beginning with a celebration of the branch's Golden Jubilee.

The IAML (<http://www.cilea.it/music/iaml/iamlhome.htm>) has around 2000 individual and institutional members in forty-five countries throughout the world, although most members are European or North American. Founded in 1951 to promote international co-operation and to support the interests of the profession, IAML is a respected member of the international library and music communities.



IAML: UK & Ireland (<http://www.iaml-uk-irl.org/>) was founded two years later in 1953 and has spent fifty years aiding understanding of the cultural, educational and economic importance of music libraries, archives and documentation centres, providing professional education and training, and representing this sector. To mark the anniversary, "Music librarianship in the United Kingdom: fifty years of the United Kingdom branch of the

*"[Music] is a language that is understood worldwide without the need for translation. It is a reflection of our cultural heritage and traditions and a significant component of the UK economy. Our music libraries, archives and specialist centres provide access to a rich variety of collections for exploration and discovery, staff in those libraries offer unparalleled expertise and support in that journey for users and at the centre of all this activity sits IAML."*

Susi Woodhouse, President, IAML (UK & Ireland), speaking at Golden Jubilee celebrations held at the British Library

International Association of Music Libraries, Archives and Documentation Centres" (Aldershot : Ashgate, 2003) was published in April this year.

A special reception was held at the British Library to mark both the fifty year anniversary and the launch of Cecilia, an online database of music resources and collections throughout the UK and Ireland, and the publication of the branch's ten-year development plan, "Access to Music. Music Libraries and Archives in the United Kingdom and Ireland: current themes and a realistic vision for the future".

Named after the patron Saint of Music, Cecilia (<http://www.cecilia-uk.org/>) aims to increase awareness of and access to the wide variety of music-related materials held in institutions. The project database was launched on 1 July 2003 and is powered by the Performing Arts Data Service (PADS: <http://www.pads.ahds.ac.uk/>). A simple interface allows users to perform quick or advanced searches based on keyword, title, location and format, returning results from over 1500 collections based in 500 distinct institutions. It is hoped that the number of repositories and collections added to the database will continue to grow in the future. To contact the

Cecilia project, e-mail the project manager at: [cecilia-uk@britishlibrary.net](mailto:cecilia-uk@britishlibrary.net).

The Access to Music 10-year strategy addresses relevant issues in the context of the changing environment of music libraries and archives. It emphasises the importance of music to all aspects of society, the central enabling role played by music repositories and their staff, and the continuing need to ensure access for users. The strategy was developed by Malcolm Lewis of Nottingham City Libraries and Pamela Thompson, Chief Librarian at the Royal College of Music, who said "Music in libraries can give immense added value economically, culturally and educationally and this report identifies and discusses issues which must be addressed in order to continue to deliver this, and sets out a strategy to kick-start the process which builds on our [previous] achievement..." The strategy is funded by the Research Support Libraries Programme (<http://www.rslp.ac.uk/>). "Access to Music. Music Libraries and Archives in the United Kingdom and Ireland: current themes and a realistic vision for the future" can be purchased for £15 by e-mailing the IAML(UK & Ireland) Publications Officer: [mroll@bucksc.gov.uk](mailto:mroll@bucksc.gov.uk).

BACK TO PAGE 1

## THE OPEN VIDEO PROJECT: BUILDING AN OPEN SOURCE DIGITAL VIDEO ARCHIVE

**MENG YANG**  
**GARY MARCHIONINI**  
**BARBARA M. WILDEMUTH**  
**AND XIANGMING MU**

Interaction Design Laboratory,  
 School of Information and  
 Library Science, University  
 of North Carolina

There is an increasing demand for digital videos in the educational and research communities. Instructors may want to find a video segment to illustrate the topics they teach in class; students might like to insert a short video clip in their class projects; and researchers in video retrieval fields could try to find different types of digital videos to test their new algorithms for face recognition, object identification or motion detection. Unlike

text documents, the challenges in storage, indexing, retrieval and copyright issues make digital videos more difficult to find, download and use from WWW sources. The Open Video Project (<http://www.open-video.org>) aims to provide an easy-to-use open source digital video archive to the research and educational communities while serving as a testbed for digital library research and development at the Interaction Design Laboratory.



The Open Video Project grew out of experience designing the Baltimore Learning Community Project [Marchionini et al., 1997], which intended to provide multimedia materials to middle school science and social studies teachers. Since 1999, the goals have evolved in three directions: to create a publicly accessible digital video repository for educators and researchers, to develop and evaluate user interfaces of digital video libraries, and to develop open source digital library tools. Currently, the collection contains about 2000 individual video segments, including MPEG-1, MPEG-2, MPEG-4 and QuickTime files. The videos were contributed from various sources such as government agencies, universities, and personal collections. In this article we first introduce the current collection content and the contribution sources, then describe our user interface research issues, and finally give a brief summary of the digital library tools we have developed.

## AN OPEN VIDEO ARCHIVE

The first goal of the Open Video Project is to create an open source digital video archive for the research and educational communities. To build this open archive, we have acquired specific video collections. In addition, we try to instantiate and evaluate the Sharium concept for digital libraries (DLs) [Marchionini, 1999], which means inviting contributions and gathering patron participation in DLs. Our videos were contributed from various sources: US government agencies such as the Library of Congress and NASA, research projects such as Carnegie Mellon's Informedia Project (<http://www.informedia.cs.cmu.edu/>), the University of Maryland's Human-Computer Interaction Lab (HCIL: <http://www.cs.umd.edu/projects/hcil/>) and the Digital Himalaya Project (see Table 1), and also from personal collections such as the Prelinger Archives (<http://www.prelinger.com/>). In addition, we collaborate

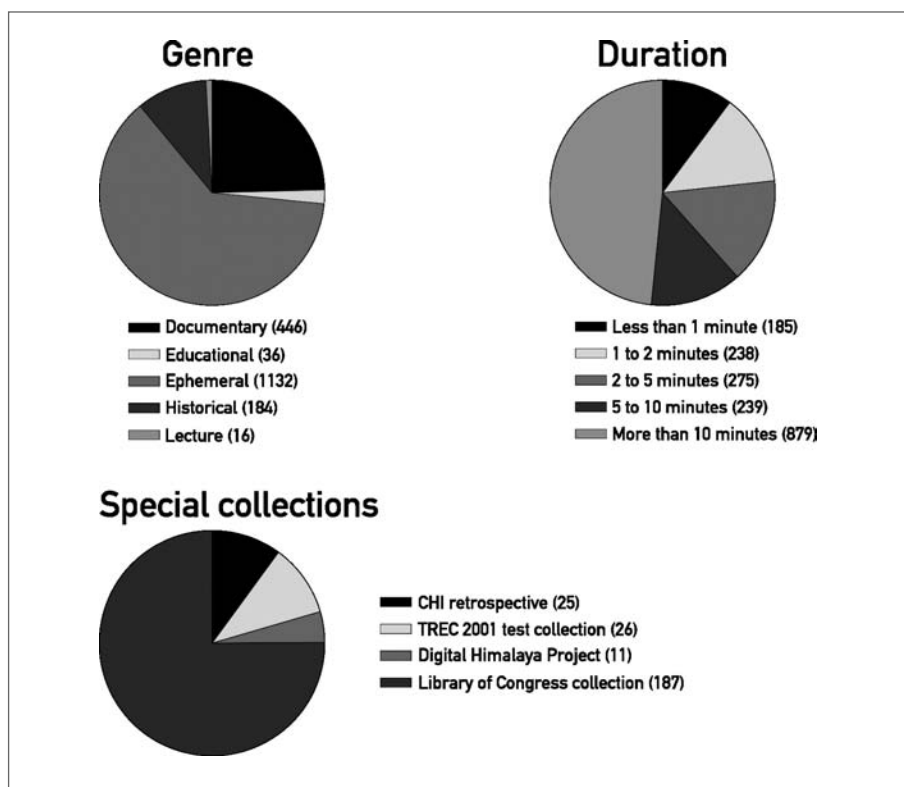
with other digital video collections such as Internet Archive Moving Image Collections (<http://webdev.archive.org/movies/movies.php>). We also encourage anyone who owns videos to make their collections available from the video archive. As of Spring 2003, the archive provides about 2000 video segments (more than 0.5 terabytes of content), representing 460 hours of video footage. Among them are documentaries, educational videos, and ephemeral videos (mostly from the Internet Archive, <http://www.archive.org/>).

Since downloading a whole video clip is time-consuming, video segmentation is a fundamental step for the efficient accessing, retrieving and browsing of large digital video libraries [Wang et al., 2000]. We also assume that most users want only short

segments of the videos, rather than the whole video. Therefore, we segmented the videos manually based on human judgement, for instance, by a story or event clip. The duration of the videos in our collection ranges from several seconds to about an hour (see Table 1 for details). In addition, we provide metadata at three levels: entire video, segment and frame level. The current metadata schema being used is Dublin Core compliant, and we are an Open Archives Initiative Data Provider.



Figure 1 - Example of a storyboard surrogate






## USER INTERFACE RESEARCH

Our second goal is to create and evaluate highly interactive user interfaces that allow people to select representations, and control them quickly and easily to achieve their information-seeking needs [Marchionini et al., 2002]. We are working towards this goal in two phases. In the first phase, we evaluated various kinds of video representations (known as surrogates) individually, trying to see how much people could understand from watching only these surrogates [see Wildemuth et al., 2002, and Wildemuth et al., 2003] and how they interacted with textual and visual information [Hughes et al., 2003]. In the second phase, we have integrated these surrogates together to create ‘Agile View’ interfaces for users to explore video content [see Marchionini et al., 2000; Geisler, 2002]. At present, we are creating and evaluating previews, overviews, reviews (or history views), peripheral views and shared views, and are initiating user studies focused on their combination.

Due to the complex characteristics of the video content (e.g. multiple channels, long download times), various kinds of video surrogates need to be created for users to search and browse before they decide to download and use the whole video. O’Connor [1985] first suggested that key frames – still images representative of scenes extracted from the video itself – could be used to construct video abstracts or ‘contour maps’. Several kinds of video surrogates (e.g. poster frame, storyboard, slideshow and fast forward) as well as the traditional textual surrogates (e.g. title, keywords and descriptions) have been applied in the design of Web interfaces of the Open Video archive. A poster frame [Christel et al., 1997] is a single ‘thumbnail’ image representing a video clip. A storyboard (see Figure 1) displays multiple key frames simultaneously and has been widely used as a video surrogate in video research projects [Christel et al.,

1997, Tse et al., 1998, Lee et al., 2002, etc.]. A slide show [Tse et al., 1998] is a dynamic video surrogate, which presents individual key frames consecutively. A fast forward is a new kind of surrogate we have been developing and investigating [Wildemuth et al., 2003]; it mimics the fast-forward function of a VCR. A fast forward plays the whole video content at a much faster speed than a VCR and, like the analogue fast forward, no audio is provided.

To view the fast-forward surrogate of this video clip, click here  (requires Quicktime) or to see the full video represented by this storyboard please click: [http://www.open-video.org/segment.php?seg\\_id=3937](http://www.open-video.org/segment.php?seg_id=3937)

Several user studies have been conducted to evaluate the effectiveness of video surrogates, based on user performance (i.e. object and action recognition, gist determination, and visual gist [Yang et al., 2003b, and Wildemuth et al. 2002]) compared five multi-modal video surrogates in relation to their usefulness and usability in accomplishing specific tasks. Those surrogates included storyboards with text or audio keywords, slide shows with text or audio keywords and fast-forward. The results showed that no surrogate was uni-versally judged ‘best’, but the fast-forward surrogate garnered strong support, particularly from experienced video users. Additionally, users suggested that the interaction between textual and visual information was quite important to understand the videos. In a subsequent study, we evaluated four speeds of fast-forward surrogate [Wildemuth et al., 2003] and, based upon users’ performance, we set the default speed for fast-forward to be 1:64 in the Open Video archive. These performance studies were augmented with an eye-tracking study to see “where, when, and how long users looked at text and image surrogates” [Hughes et al., 2003, p1]. The results showed that the participants looked at textual information first and fixed their

gaze on it much longer than on the images. Most of them used the text as an anchor from which to make judgements about the search results and the images as confirmatory evidence for their selections.

Our current work is more focused on the second phase of the user interface research: to create ‘Agile View’ interfaces. We aim to give people several classes of views: overviews of collections of video segments; previews of specific video objects; reviews or history views of past searches or community activities; peripheral views of related objects; and shared views instantiated by others active in the sharium environment (using recommendations based on user patterns captured in the server logs).

## OPEN VIDEO TOOLS

There are a lot of challenges in building a digital video library, such as creating video surrogates, cataloguing and indexing videos at different levels, and sharing videos among different users. As we gain more experience, the backend processes of managing a video digital library become more routinised. We have developed a variety of tools and scripts in this regard. Three tools have been developed or modified to help to generate video metadata. We modified the MERIT system [Kobla et al., 1998] to extract keyframes on our Linux platform. VAST [Mu et al., 2003b] is a semi-automatic tool to generate storyboards and fast-forward surrogates. VIVO [Yang et al., 2003a] is a tool to help digital video librarians to input, edit and manage video metadata elements on different levels. We have also developed some video sharing application tools. ISEE [Mu et al., 2003a] is a real-time collaborative multimedia distance learning tool. Brief descriptions of these three tools follow.

**V**AST (Video Annotation and Summarization Tool) [Mu et al., 2003b] is a novel video metadata authorisation system that integrates both semantic and visual metadata (see Figure 2). Balance



between accuracy and efficiency is achieved by adopting a semi-automatic authorisation approach. Semantic information such as video annotation is integrated with the visual information under an XML schema. It can act as a semi-automatic tool to generate storyboard and fast-forward surrogates.



Figure 2 - VAST interface

**VIVO** (Video Indexing and Visualization Organizer) [Yang et al., 2003a] is a tool we developed to help digital video librarians to input, edit and manage video metadata elements on different levels. It uses a hierarchical tree to organise the multiple-level (e.g. video, segment and frame) video structure (see Figure 3). Each video level has its own textual metadata and visual surrogate. Librarians can catalogue and index videos at each level without losing context.

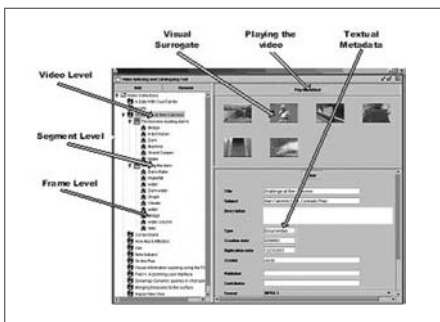


Figure 3 - VIVO interface

The **ISEE** (Interactive Shared Educational Environment) [Mu et al., 2003a] is a real-time collaborative multimedia distance learning tool suitable for heterogeneous network connections (see

Figure 4). ISEE supports real-time text chat, video playback, video navigation with a storyboard, video synchronisation across users, hypertext browsing, Web information access and Website URL sharing.

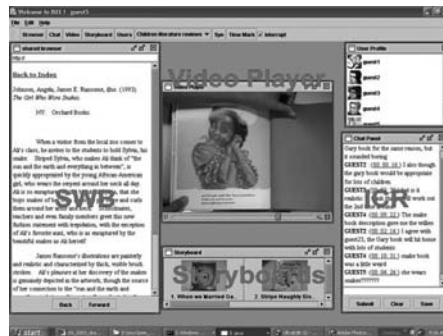


Figure 4 - ISEE interface

These tools will continue to be tested and modified over time. In addition, a variety of other tools as well as programs for managing files and surrogates and the MySQL database and PHP code to link the Open Video interface to the backend have been developed. Over the long term, we hope to create an integrated and open source toolkit that can be shared by other research groups.

## SUMMARY

In this article, we have given a brief overview of the current achievements in our NSF-funded Open Video Project. We have about 2000 individual video segments in our collection that have been donated by various contributors. We have tested and evaluated several different types of video surrogates individually and tried to integrate them together to create an 'Agile View' user interface. We have also developed a set of tools to help with cataloguing and indexing videos, generating video surrogates and supporting video sharing and collaboration. We will continue our work in these areas and hope to communicate and collaborate with other digital video archive groups.


This work is supported by National Science Foundation (NSF) Grant IIS 0099638.

## REFERENCES

- [Christel et al., 1997] Christel, M., Winkler, D. and Taylor, C. R. (1997). Improving access to a digital video library. Paper presented at the Human-Computer Interaction: INTERACT97, the 6th IFIP Conference on Human-Computer Interaction, Sydney, Australia, 14-18 July 1997.
- [Geisler, 2002] Geisler, G. (2002). AgileViews: A Framework for Creating More Effective Information Seeking Interfaces. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.
- [Hughes et al., 2003] Hughes, A., Wilkens, T., Wildemuth, B. and Marchionini, G. (2003, to be published). Text or Pictures? An Eyetracking Study of How People View Digital Video Surrogates. International Conference on Image and Video Retrieval, Beckman Institute, Urbana, IL, 24-25 July 2003.
- [Lee et al., 2002] Lee, H. and Smeaton, A.F.(2002). Designing the User-Interface for the Fischlár Digital Video Library. Journal of Digital Information, 2(4), Special Issue on Interactivity in Digital Libraries. Available at: <http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Lee/>
- [Kobla et al., 1998] Kobla, V., Doermann, D. and Faloutsos, C. (1998). Developing High-Level Representations of Video Clips using Video Trails. Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases VI, pp. 81-92.
- [Marchionini et al., 1997] Marchionini, G., Nolet, V., Williams, H., Ding, W., Beale, J., Rose, A., Gordon, A., Enomoto, E. and Harbinson, L. (1997). Content + Connectivity => Community: Digital Resources for a Learning Community. Proceedings of ACM Digital Libraries 97 (Philadelphia, PA: 23-26 July 1997), pp. 212-220.
- [Marchionini, 1999] Marchionini, G. (1999). Augmenting library services: Toward the sharium. Proceedings of International Symposium on Digital Libraries 1999 (Tsukuba, Japan, 28-29 September 1999), pp. 40-47.
- [Marchionini et al., 2000] Marchionini, G., Geisler, G. and Brunk, B. (2000). Agileviews: A Human-Centered Framework for Interfaces to Information Spaces. Proceedings of the Annual Meeting of the American Society for Information Science (Chicago, 12-16 November 2000). pp. 271-280.
- [Marchionini et al., 2002] Marchionini, G. and Geisler, G. (2002). The Open Video Digital Library. D-Lib Magazine, 8 (12), December. Available at: <http://www.dlib.org/dlib/december02/marchionini/12marchionini.html>
- [Mu et al., 2003a] Mu, X., Marchionini, G. and Pattee, A. (2003). The Interactive Shared Educational Environment:

Continued next page

BACK TO PAGE 1



“User interface, system architecture and field study”. Proceedings of the Third ACM+IEEE Joint Conference on Digital Libraries (Houston, Texas, 27-30 June 2003), pp. 291-300.

[**Mu et al., 2003b**] Mu, X. and Marchionini, G. (2003) (to be published). “Enriched video semantic metadata: authorization, integration and presentation”, The Annual Meeting of the American Society for Information Science (Long Beach, LA, 19-22 October 2003).

[**O’Connor, 1985**] O’Connor, B. (1985). Access to moving image documents: background concepts and proposals for surrogates for film and video works. *Journal of Documentation*, 41, pp. 209-220.

[**Slaughter et al., 2000**] Slaughter, L., Marchionini, G. and Geisler, G. (2000). Open Video: A Framework for a Test Collection. *Journal of Network and Computer Applications, Special Issue on Network-Based Storage Services*. pp. 219-245.

[**Tse et al., 1998**] Tse, T., Marchionini, G., Ding, W., Slaughter, L. and Komlodi, A. (1998). Dynamic key frame presentation techniques for augmenting video browsing. Proceedings of AVI’98: Advanced Visual Interfaces (L’Aquila, Italy, 25-27 May 1998), pp. 185-194.

[**Wang et al., 2000**] Wang, Y., Liu, Z. and Huang, J. (2000). Multimedia Content Analysis Using Both Audio and Visual Cues. *IEEE Signal Processing Magazine*, November 2000, pp. 12-36.

[**Wildemuth et al., 2002**] Wildemuth, B., Marchionini, G., Wilkens, T., Yang, M., Geisler, G., Fowler, B., Hughes, A. and Mu, X. (2002). Alternative Surrogates for Video Objects in a Digital Library: Users’ Perspectives on Their Relative Usability. Proceedings of the 6th European Conference ECDL 2002 on Research and Advanced Technology for Digital Libraries, Berlin: Springer. (Rome, 16-18 September 2002) pp. 493-507.

[**Wildemuth et al., 2003**] Wildemuth, B., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A. and Gruss, R. How fast is too fast? Evaluating fast forward surrogates for digital video. Proceedings of the Third ACM+IEEE Joint Conference on Digital Libraries, (Houston, Texas, 27-30 June 2003), pp. 221-230.

[**Yang et al., 2003a**] Meng, Y., Mu, X. and Marchionini, G. (2003) (poster), VIVO - A video indexing and visualization organizer, Proceedings of the Third ACM+IEEE Joint Conference on Digital Libraries, (Houston, Texas, 27-30 June 2003), p.403.

[**Yang et al., 2003b**] Yang, M., Wildemuth, B., Marchionini, G., Wilkens, T., Geisler, G., Hughes, A. and Gruss, R. (2003) (to be published). Measures of User Performance in Video Retrieval Research. The Annual Meeting of the American Society for Information Science (Long Beach, LA, 19-22 October 2003).

## SURVEY OF CULTURAL HERITAGE WEBSITES

24 Hour Museum (<http://www.24hourmuseum.org.uk/>), The National Virtual Museum for the UK, have published the results of a recently compiled survey of Web statistics across the cultural heritage sector. The survey presents user statistics for 2002 of almost twenty Websites including national museums and the National Archive. Results can be downloaded from: [http://www.24hourmuseum.org.uk/etc/press/TXT17762\\_gfx\\_en.html](http://www.24hourmuseum.org.uk/etc/press/TXT17762_gfx_en.html)

[BACK TO PAGE 1](#)

## INTELLECTUAL PROPERTY AND TECHNOLOGY FORUM

A group of students at Boston College Law School have published an online repository of information about current developments in technology law and intellectual property and how it relates to cyberspace, patent and trademark issues, and regulation on the Internet.

The Intellectual Property and Technology Forum ([http://www.bc.edu/bc\\_org/avp/law/st\\_org/iptf/index.html](http://www.bc.edu/bc_org/avp/law/st_org/iptf/index.html)) uses the electronic publishing format to present scholarly articles and breaking news about legal issues concerning IP and technology law alongside subjective discussions and commentaries. The site was designed to complement the content provided by traditional legal journals.

[BACK TO PAGE 1](#)

## WEBKIT – INTUITIVE PHYSICAL INTERFACES FOR INFORMATION RETRIEVAL AND MANAGEMENT

From an early age we can move, manipulate and assemble a variety of physical objects with very little cognitive effort. Yet current Human-Computer Interfaces (HCIs) use only a limited range of these abilities. Graphical User Interfaces (GUIs) have brought huge benefits; however, there is growing evidence that direct manipulation is far more important than pictorial metaphors (e.g. the ‘desktop’) for assisting end-users. Tangible User Interfaces (TUIs) give physical form to information, using objects both as representations of information and as controls to manipulate the underlying platform.

WebKit, a twenty-seventh-month collaborative project to be completed in May 2004 and which receives funding

from the European Commission under the Framework 5 Programme, utilises a Tangible User Interface (TUI) for navigating sources of information (e.g. the WWW) that places priority on physical manipulation by the users. Essentially users control the system and navigate, manipulate and utilise data and information by selecting and positioning physical objects, not just icons on a PC screen. Thus WebKit combines a ‘tangible’ element with digital representations (graphics, audio) that provide a dynamic element – showing how manipulations interact with and utilise the underlying data.

WebKit combines an examination of novel multimodal interfaces with innovative work in Semantic Web technology and information management.

WebKit is developing an innovative end-to-end system, expected to yield new exploitation opportunities, principally in the field of education. WebKit will be evaluated extensively both inside school and outside (e.g. in museums) by children in a variety of age groups.

[BACK TO PAGE 1](#)

#### WebKit and the trial ASTRAL project

will be covered in more detail in the next issue of DigiCULT.info. The WebKit project co ordinator, Dr John Holden of Central Research Laboratories, can be contacted at: [jholden@crl.co.uk](mailto:jholden@crl.co.uk)

WebKit Participants	
Central Research Laboratories Ltd	United Kingdom
University of Strathclyde	United Kingdom
Associazione Europea per l'Ambiente - Tecnopower - Ricerca Scientifica Informatica	Italy
ITEL, Informatics & Telematics Ltd.	Greece
The University of Cambridge	United Kingdom
Université de Genève	Switzerland
The University of Warwick	United Kingdom
Neurosoft Sp. Z. O.O.	Poland
Giunti Interactive Labs S.R.l.	Italy

## STEWARDSHIP RESOURCES ON THE WEB: RAISING AWARENESS OF THE ROLE STEWARDSHIP PLAYS IN ENSURING THE PRESERVATION OF CULTURAL MATERIAL

DARRYL MEAD AND  
JUDY MEAD.

M Squared Consulting

This research project was commissioned by Resource, the Council for Museums, Archives and Libraries (<http://www.resource.gov.uk/>). Resource was launched in April 2000 as the strategic body working with and for museums, archives and libraries, tapping the potential for collaboration between them. The new organisation replaced the Museums and Galleries Commission (MGC) and the Library and Information Commission (LIC), and now includes archives within its portfolio. Resource is a non-departmental public body sponsored by the Department for Culture, Media and Sport, part of the British government. Shortly after it was established, Resource commissioned consultants to produce a report from which it could develop its stewardship agenda. The outcome of this research and consultation was published in February 2002 as *Preserving the Past for the Future*. This

study noted that, too often, insufficient importance is attached to conservation and collections management. As a consequence, there is a risk that unconsidered neglect of collections could result in a loss to future generations of users.

In late 2002 Resource commissioned M Squared Consulting to conduct a new research project under the banner of Raising awareness of the role stewardship plays in ensuring the preservation of cultural material. The full report, *Advocacy of Stewardship on the Web*, is now available on Resource's Website at: <http://www.resource.gov.uk/documents/id693rep.pdf>.

### OVERVIEW

Caring for collections in museums, archives and libraries is a broad topic which in recent years has been gathered under the umbrella term of stewardship. In

this study, stewardship was defined as collections management, care and access and was only applied to portable cultural heritage.

Users of stewardship information are a diverse group ranging from individuals wishing to preserve an heirloom, to industry professionals, as well as libraries, museums and archives of every size and budget, even funding agencies, civil servants and politicians.

The focus of the study was on identifying English-language resources available on the World Wide Web that best provide information on the delivery of effective stewardship across all three domains. Creating this list provided potential users with a pre-screening service for their research. Web-based resource delivery was essential, now being available to virtually all potential users.

## RESEARCH

Research was carried out in February and March 2003. Over 100 sites were examined and graded on a five-point scale according to their likely usefulness to individuals or organisations engaged in any aspect of delivering stewardship (5 for the Best sites down to 1 for sites that provided no real information, and 0 if the Website could not be located). In the full study, detailed reviews were provided for the twenty best sites, with commentaries on their individual strengths. Direct links were provided by URLs to the top level of the area of each site most relevant to stewardship issues.

Care was taken to ensure a fairly even spread of sites across the museum, library and archive domains. Examples were explicitly sought from English-language sites around the world. Where there were similar organisations in a geographical area with relevant content, such as the national libraries in Australasia, only the most useful example was kept in the 'Best' group.

While the research team took every care to review sites rigorously and fairly, the review process was at least as much an art as a science. If navigation directions within a site did not lead an experienced researcher to stewardship advocacy material by a short and clear path, we felt that it was not appropriate to rank the site 'Best' as there were sufficient sites that combined easy navigation with effective information provision which received our highest rating.

Equipment and technology used to view the sites were only partially standardised. All reviewers employed Microsoft's Internet Explorer Version 5 as a browser. The hardware included computers running various Windows and Macintosh operating systems. Internet connections were made at speeds ranging from 56K

dial-up to 600K Broadband and 2MB Local Area Networks. None of the reviewed sites performed in an unreasonably slow way, though graphics-intensive examples benefited from a connection at a speed greater than 56K. This wide range of hardware, software and connection speeds provided a good test-bed for real-world usage. No particular effort was made to find sites that made use of audio content although clever and effective sounds were found on several of the more interactive sites.

## RESULTS

Site availability was very high; only one site out of over one hundred was unavailable during our tests, and then only for three days. The top-rated sites (scoring 5) were spread widely around the world, with 40% coming from North America and another quarter from Europe and Australia. This reflected both the international distribution of major trade bodies and the strong voice of well-funded American institutions. This list was also dominated by national organisations. The United Kingdom was clearly the dominant nation in the Very Useful category (scoring 4) with a dozen English sites and three from Scotland. Part of this strength stemmed from the desire of the research team to assess all the relevant bodies in the UK and part came from the team's superior knowledge of the UK scene in each domain.

A surprising and welcome revelation was the finding that a substantial portion of the stewardship sites were equally relevant for libraries, archives and museums, although, somewhat oddly, no sites specifically relevant to the museum and library combination were given a high rating.

Two broad categories of sites were evident. The first, those promoting stewardship, were mostly provided by industry bodies and national institutions, or were rooted in academia. The second were those

libraries, archives and museums demonstrating best practice in their approach to stewardship. They were selected to showcase what could be achieved at different operational scales and in diverse environments. While the main emphasis was directed to the world of real objects in real-world environments, the migration of many collections into the world of digital artifacts, virtual collections and metadata also presented stewardship challenges. Relevant sites supporting digitisation issues for libraries, archives and museums have been included as well.

*"A substantial portion of stewardship sites were equally relevant for libraries, archives and museums."*

The twenty best stewardship sites are identified on page 10 of the report *Advocacy of Stewardship on the Web*: <http://www.resource.gov.uk/documents/id693rep.pdf>

## SHORTCOMINGS IN THE WEBSITES

While the study noted that there were many positive aspects to stewardship resources on the Web, there were also shortcomings. Many heritage Web sites in the UK failed to do justice to the collections they hold. There appeared to be an assumption that commercial information was all the online viewer wanted or needed. It was noted that there was a failure to re-engineer booklets and handouts from the physical world for the Web in order to speed up download times. Only occasionally was any attempt made to fully utilise the functionality of the Web by treating the user as a vital partner in the enterprise, a source of deposits, useful information and donations. This contrasted strongly with the best sites in the United States, Australasia and Europe where the partnership between research and collections management and delivery was transparent and informs the commercial perspective, rather than being opposed to it.



One lesson that could be drawn from this exercise was that perhaps it is time for many stewardship professionals to re-evaluate how they approach the presentation of their work, particularly on the Web. It has already been noted that the audience for this work is now much more broad-based than it has been in the past and it is important that all potential users are catered for. Broadening, simplifying and clarifying sites will impress the funders of the Websites and their attendant programmes, as they look for ways to increase the potential and visibility of their investments.

### SOME CONCLUSIONS

It was confirmed that the Web is an excellent place to look for resources to support stewardship at all levels, from the interested amateur to the professional seeking any type of advice. When the top 42 sites found in this study were considered, the volume and quality of resources available on the Web were currently greater than can be found in all but a handful of the world's most lavishly equipped libraries and research institutes. There were enormous advantages in the fact that most of

the content on offer was free. Stewardship Websites often featured the most up-to-date sources, some of which were work-in-progress or will not be printed by traditional means for some time to come (if at all).

There was also a very positive move towards greater openness about what we do and how we do it, particularly in the public sector. This is a result of both changing political attitudes and the new opportunities that communications media such as the Web have provided. In many areas, stewardship being one of them, there is a much greater tendency to share practice and experience that would not, a generation ago, have resulted in writing a book or even a journal article.

Stewardship on the Web will remain a rapidly evolving subject. New sites will continue to appear and the Best selection of this study will inevitably lose some of its currency. However, against this background of change, sites recommended in this study are expected to have a long future life, as we assumed that, in most cases, the host organisation will continue to update their

material. In cases where the publishing organisation does run into financial difficulties or the limited-life programme comes to an end, it was considered highly likely that the service will migrate to another server and/or supporting organisation, a practice of which there are already examples.

### FOR THE FUTURE

In order to improve working practices, the principal UK funding bodies such as the Heritage Lottery Fund (<http://www.hlf.org.uk/>), Resource (<http://www.resource.gov.uk/>), the Arts and Humanities Research Council (<http://www.ahrc.ac.uk/>), and the Joint Information Systems Committee (<http://www.jisc.ac.uk/>), need to agree on a set of attributes which should form part of the Website of any organisation in receipt of funding. Why, if the HLF insists, as it does, on strategy documents and business plans, are these not available online? There needs to be a reasonable amount of prototyping, support and an appreciation of the unique facilities of the Web in order to create quality exemplars which can provide these valuable resources into the future.

[BACK TO PAGE 1](#)

## SEMANTIC WEB CHALLENGE

DigiCULT's recent publication Thematic Issue 3 "Towards a Semantic Web for Heritage Resources" considers issues relating to the basis, development and usage of Semantic Web technology by the cultural heritage sector. Prior to the 2nd International Semantic Web Conference in October, the Semantic Web Challenge (<http://challenge.semanticweb.org/>) aims to highlight new applications which use Semantic Web techniques to create applications which "integrate, combine and deduce information from various sources to assist users in

performing specific tasks." The challenge lasts for at least five years and each year, as the knowledge and development of semantic applications increases, a new goal is added to the challenge. In 2003, submissions must incorporate two external data sources which follow different schemata, although the data sets are not specified due to the potentially very wide applicability of the Semantic Web.

In 2003, the challenge invited submission of Semantic Web applications from industry and academia which adhere to a

set of minimum criteria and aim to satisfy an additional list of desired functionality. Minimum requirements were as follows: information sources should be geographically distributed, structurally and semantically diverse, have diverse ownership and contain real-world data. It was also required that the applications assume that information is never fully complete and that some formal description of the meaning of data was included.

Registration closed on 15 August and applicants have until 15 September



2003 to submit Web pages online. Entries are judged by the Semantic Web Challenge Advisory Board and the best applications and winners will be announced at the 2nd International Semantic Web Conference (ISWC 2003: <http://iswc2003.semanticweb.org/>).

The challenge will run next year, with slightly different specific goals.

[BACK TO PAGE 1](#)

### DigiCULT Thematic Issue 3

“Towards a Semantic Web for Heritage Resources” [http://www.digicult.info/downloads/ti3\\_high.pdf](http://www.digicult.info/downloads/ti3_high.pdf) presents a detailed analysis of the Semantic Web.



## DELOS NETWORK OF EXCELLENCE

MARIA SLIWINSKA

The International Centre for Information Management Systems and Services, ICIMSS, Torun, Poland

The DELOS Network of Excellence on Digital Libraries brought together sixty-eight academic and industrial organisations interested in research and applications in the domain of Digital Libraries. This enabled the institutions previously working separately to consolidate forces as part of the European Commission’s Fifth Framework Programme.

DELOS partners represented Austria, Czech Republic, Denmark, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Latvia, Norway, Poland, Portugal, Russia, Slovakia, Spain, Sweden, Switzerland, The Netherlands, and the United Kingdom. The European Research Consortium in Informatics and Mathematics (ERCIM, <http://www.ercim.org/>) played a Contractor role, and the scientific part of the project was handled by Consiglio Nazionale delle Ricerche – Istituto di Elaborazione della Informazione in Pisa (<http://www.cnr.it/>). A full list of partners can be found on the DELOS Website <http://delos-noe.iei.pi.cnr.it/>.

The objectives of the DELOS Network of Excellence (NoE) were defined as:

- Provide a context to the international research agenda for further activities in

- digital library domains.
- Contribute to the creation of the European digital library research community.
- Contribute towards improving the effectiveness of European research in the area of digital libraries.
- Provide a forum for researchers, practitioners and industries for exchange of ideas and experience.
- Co-operate with ongoing standardisation activities.
- Facilitate take-up of digital libraries technologies.
- Create a forum where results of the European Union funded digital libraries projects could be discussed.
- Promote co-operation between European and national digital library initiatives.
- Improve international co-operation in the digital library research area.

The DELOS project started its activities in 1999 with the following five forums organised around specific issues, each with its own aims and objectives and organisational structure.

### DIGITAL LIBRARY RESEARCH FORUM

The Digital Library Research Forum presented a series of thematic work-

shops on such issues as: Users and User Interfaces, Evaluation in Context, Metrics and Testbeds, Evaluation of Digital Library Services and Scalability. This Forum organised European Conferences on Digital Libraries (ECDL, [http://delos-noe.iei.pi.cnr.it/activities/researchforum/ECDL/ecdl.html?content=ecdl\\_c.html](http://delos-noe.iei.pi.cnr.it/activities/researchforum/ECDL/ecdl.html?content=ecdl_c.html)). It also developed an important collaboration with the National Science Foundation (NSF <http://www.nsf.gov/>) in the USA, agreeing to co-operate in eight working groups: Spoken-Word Digital Audio Collections, Information Extraction from Digital Libraries, Personalisation and Recommended Systems in Digital Libraries, Emerging Language Technologies and the Rediscovery of the Past, Digital Imagery for Significant Cultural and Historical Materials, Preservation and Archiving, and Agents in Digital Libraries.

### DIGITAL LIBRARY EVALUATION FORUM

In the Digital Library Evaluation Forum, working groups were established for the purpose of a Digital Library Test Suite and Metalibrary and Digital Library schema. A joint working group with the NSF was established to structure the field of Digital Library evaluation. An important part of this Forum was an Initiative for the Evaluation of XML Retrieval (INEX), and



the Evaluation of Cross-Language Spoken Document Retrieval Systems (CLEF). As a result of INEX a new Digital Library test-bed was created with a range of 12,000 articles from the field of computer science published during the years 1995-2001. The CLEF activities became so rich that they have been converted into a separate new project (<http://clef.iei.pi.cnr.it:2002/>).

### DIGITAL LIBRARY STANDARDISATION FORUM

The Forum for Standardisation included work on Registries, Ontology Harmonisation, and Project Metadata. Working groups included researchers not only from the DELOS project but from the USA, Japan and Australia. The groups produced a number of White Papers devoted to these topics, including: Building Core Ontologies, Principles of Metadata Registries, Digital Libraries Future Directions For A European Research Programme, Describing Projects.

### DIGITAL LIBRARY TRAINING AND TECHNOLOGY FORUM

This forum created two International Summer Schools in Pisa to foster research and understanding in the fundamental technologies underlying the Digital Libraries field, in addition to national workshops. Materials from these Summer Schools are available in printed or digital form from: <http://delos-noe.iei.pi.cnr.it/activities/trainingforum/SummerSchool/summerschool.html>

### DIGITAL LIBRARY INTERNATIONAL CO-OPERATION FORUM

The forum for International Co-operation encouraged co-operation between EU and US researchers, largely facilitated by an All Projects meeting where fifty-two presentations were given in eighteen parallel sessions to an audience of over 130 participants. In this Forum, co-operation with Newly Accessed States (NAS) has begun.

Information on DELOS activities has already been published in various papers from ERCIM News ([http://www.ercim.org/publication/Ercim\\_News/](http://www.ercim.org/publication/Ercim_News/)) and D-Lib Magazine (<http://www.dlib.org/>), therefore this article concentrates on some lesser known activities of NAS countries in Central and Eastern Europe.

Inclusion of Central and East European cultural heritage institutions in the DELOS project was in line with the opportunities offered by the European Commission within the Fifth Framework Programme and Key Action VIII.1.6 objectives. Activities undertaken by Central and East European (CEE) partners can be described as dissemination of the DELOS project achievements through the planned organisation of a number of customised scientific events (thematic workshops, training and tutorials) according to the existing real needs of the cultural institutions from CEE, in co-operation with colleagues from the European Union.

The established DELOS – CEE consortium included the following members:

- European Research Consortium in Informatics and Mathematics (ERCIM), Sophia-Antipolis, France – Contractor
- The Head Office for State Archives (NDAP, <http://www.archiwa.gov.pl/>), Warsaw, Poland – Co-ordinator
- The International Centre for Information Management, Systems, and Services (ICIMSS, <http://www.icimss.edu>), Toruń, Poland
- Comenius University (<http://www.uniba.sk/webuk/>), Bratislava, Slovak Republic
- Latvian University (<http://www.lu.lv/>), Riga, Latvia
- Hungarian Library Institute (<http://www.ki.oszk.hu>), Budapest, Hungary

In addition, CNR (Italy) played an essential role as a scientific co-ordinator.

Participation in the DELOS NoE was a valuable experience for Central and East European institutions, which had a long way to go during a very short time. At the time of the conception of DELOS, archives, libraries and museums often created their own home pages to publicise their institutions and attract users, and these users did not have to physically visit cultural heritage institutions to check opening hours, obtain information about exhibitions, or check the latest acquisitions. However, it was still extremely unlikely to find the actual content of the resources online. In such an environment, the DELOS project was extremely welcome. It should be emphasised that, in the last ten years, the most active heritage institutions have made a long metaphorical journey from the 19th to the 21st century – today, many archives, libraries and museums have electronic catalogues and there is a real likelihood of providing digitised materials, allowing users to consult documents without actually visiting the institution.

DELOS-CEE activities focused on the dissemination of achievements of the DELOS members by training, translation and publication of selected materials as well as by technology transfer. DELOS members were offered the option of taking part in events organised by the DELOS consortium and other relevant events. Taking this opportunity, DELOS-CEE members participated in an All Projects' meeting held in Rome in March 2002 where European and American projects concerning the problems of digitisation were presented. In addition, the ECDL conference took place in Rome in September 2002, where DELOS-CEE members ran a session, and evaluated and selected a number of papers. DELOS-CEE members and colleagues from other Central and East



European countries also took part in two Summer Schools – training offered by CNR.

Among others, the following training activities were chosen: ERPANET training devoted to the preservation of Internet resources, MINERVA workshop on co-ordination of digitisation activities in Europe, and Electronic Publication conference held in Guimaraes (Portugal). Moreover, DELOS-CEE members organised a number of events in their own countries including workshops, meetings with national librarians and archivists, and national events. The list of DELOS-CEE events, detailed programmes and PowerPoint presentations can be found on the DELOS-CEE homepage: <http://www.icimss.edu/delos/>. Selected materials have been published in hard copy.

As digitisation became an increasingly important issue in Central and Eastern Europe, some funds in the DELOS project were allocated to facilitate the

translation of recommended materials in order to reach a broader audience. The translated documents were:

- a) Lund Report on Digitisation:
  - Digitisation Europe (translated into Polish and Hungarian)
- b) White Papers and reports prepared by DELOS working groups (translated into Polish and Latvian)
  - o Building Core Ontologies
  - o Principles of Metadata Registries
  - o Digital Libraries Future Directions For A European Research Programme
  - o Describing Projects
- c) Some other publications translated into Polish include *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship* (<http://www.bl.uk/services/preservation/occpaper.pdf>) and *Cyberculture, Cultural Asset Management, and Ethnohistory – Preserving the Process and Understanding the Past* ([http://www.deflink.dk/upload/doc\\_filer/doc\\_alle/740\\_sross\\_cyberculture\\_rev2.doc](http://www.deflink.dk/upload/doc_filer/doc_alle/740_sross_cyberculture_rev2.doc)), both by Seamus Ross, and *From Gutenberg to*

Global Information Infrastructure by Christine Borgman (available from MIT Press).

Another benefit was technology transfer: copying a digital library from CNR and mounting it on the ICIMSS server. It will be possible to make good use of this technology even after the end of the project.

Participation in the DELOS project was an excellent opportunity to become familiar with the latest developments in the area of digitisation and a valuable learning tool. Most importantly of all, DELOS-CEE is included in the group of centres of excellence, an achievement of which we are very proud. In the next phase of the DELOS consortium, it would be interesting to see representatives from the library and archival sector working together with academics and researchers, using the knowledge and expertise of practised users to fully understand the potential of digital library solutions. [BACK TO PAGE 1](#)

## JODI ISSUE ON ECONOMIC FACTORS FOR DIGITAL LIBRARIES

The Journal of Digital Information (<http://jodi.ecs.soton.ac.uk/>) was one of the first electronic-only journals to be published and aims to present serious research and links to other Web resources in an easily accessible way. It is available online, free, due to the support of the British Computer Society and Oxford University Press. The theme for the most recent issue (Vol. 4 Issue 2, June 2003) is “Economic Factors for Digital Libraries”. It contains several articles on the economic costs and impact of digital preservation and storage within the library and publishing fields.

There are several types of cost involved in implementing a digital library, from the immediate start-up costs of acquiring the digital objects (creating or purchasing) and the further costs of establishing the library, to the long-term expenses of managing, maintaining and preserving the resource and access to it. Whether the values and benefits of the resource can be effectively measured and offset against its cost is the subject of a wide range of discussions in this issue. You can view “Economic Factors for Digital Libraries” at: <http://jodi.ecs.soton.ac.uk/?vol=4&iss=2>. [BACK TO PAGE 1](#)

**DigiCULT Technology Watch Report 2** due to be published early in 2004, will devote a chapter to issues related to digital rights management and economic factors in cultural and heritage organisations.

## UNESCO CONVENTION FOR THE SAFEGUARDING OF THE INTANGIBLE HERITAGE OF THE WORLD

“Preliminary Draft International Convention for the Safeguarding of the Intangible Cultural Heritage and report by the Director-General on the situation calling for standard-setting and on the possible scope of such standard-setting” has been distributed by UNESCO in preparation for its biennial General Conference in autumn 2003.

This issue will affect intangible heritage (e.g. performances, crafts, oral traditions) and the way in which it is presented in museums, historic sites, libraries and archives at a local, regional and national level. The draft is available for downloading from: <http://unesdoc.unesco.org/images/0013/001307/130784e.pdf>

[BACK TO PAGE 1](#)

## E-CULTURE NET THEMATIC NETWORK (2002-2003)

**KIM H. VELTMAN** Scientific Director of the Maastricht Mc Luhan Institute and co-ordinator of a new EU Network of Centres of Excellence in Digital Cultural Heritage

In the past year, the E-Culture Net Thematic network has produced a number of reports ranging from surveys of content in Germany and Russia to discussions of Distributed Autonomous Cultural Objects (DACOs), a



### E-CULTURE NET WEBSITE

(<http://www.eculturenet.org>)

Our vision is to provide access to Europe's cultural heritage through co-ordination of networks and to develop a Distributed European Electronic Resource (DEER) through a grid for culture; to use the DEER to develop critical meth-

ods and thought via European Masters and Doctorates; and to keep it up-to-date through research matrices to understand the digital knowledge production life-cycle and generate new research topics. Interest in joining the Network of Excellence came from institutions across the EU, from the EU's prospective members (Newly Accessed States) and internationally.

E-Culture Net has taken up afresh the idea of a Network of Networks, introduced in 1989 by UNESCO through Culture Link (Network of Networks for Research and Co-operation in Cultural Development, <http://www.culturelink.org/>) and CIRCLE (Cultural Information and Research Centres Liaison in Europe,

<http://www.circle-network.org/>). Whereas the earlier effort was limited to a loose connection through a Website with addresses, E-Culture Net foresees a more systematic co-operation by involving existing specialised networks in specific elements of the DEER vision. By joining these smaller networks in a common task, E-Culture Net achieves a new critical mass; at the same time it brings into new light (valorising, in the language of Brussels) various excellent achievements such as the Inventaire, which were previously often unknown beyond their country of origin. Among the networks included to date are ERPANET (<http://www.erpanet.org>), C2RMF (<http://www.c2rmf.fr/>), NEHRN (<http://nehrn.hum.sdu.dk/>), SCRAN (<http://www.scran.ac.uk/>), ORION (<http://www.orion-net.org/>) and SIMILAR (<http://www.tele.ucl.ac.be/SIMILAR/>). Other specialised networks such as the Réseau d'Ethnologie and GLEN (Greek and Latin Epigraphy Network) are in the process of joining.

A series of links with UNESCO and with other cultural networks have been initiated and strengthened. These networks include: Canadian Heritage Information Network (CHIN, <http://www.chin.gc.ca/>), Chinese Cultural Heritage Network (CCHN), Asian Network of Excellence on Digital Silk Roads (ANoEDSR), Russian Cultural Heritage Network (RCHN, <http://www.rchn.org.ru/>), Digital Silk Roads Cultural Grid (NSF DSRCG), and the UN Educational, Scientific and Cultural Organisation (UNESCO, <http://www.unesco.org/>).

### DISTRIBUTED EUROPEAN ELECTRONIC RESOURCE (DEER)

How does one translate a long-term vision into a practical reality? A first step when considering how to create the DEER was to identify a series of 12 mod-

### INTRODUCTION

E-Culture Net (IST-2001-37491) began on 1 July 2002 as a one-year thematic network and resulted in twelve outcomes including: a Website with definition of vision, structure and expansion; taking the first steps towards Distributed European Electronic Resource (DEER); the development and demonstration of a DACO protocol for access to local databases; the development of new research areas, topics, European Masters and Doctorates, and the development of national networks.



ules and find existing groups (e.g. consortia) to develop each of these (Table 1). Next, a series of papers were produced resulting in a Foundation for the DEER with three essential features: a Distributed Repository, a Virtual Reference Room and a Virtual Agora – a forum for collaborative research and creation. The report outlined premises, issues and challenges and concluded that the DEER is both feasible and necessary.

**A**DACO protocol was developed that builds on the Open Archive Initiative (OAI) and offers access at much greater levels of granularity, namely individual images, pages or even paragraphs, without requiring a re-organisation of existing databases. This provides a European solution to the challenge of accessing national, regional and local diversity while maintaining a “unity of diversities” (Ruffolo). The DACO protocol was applied to five distributed databases to provide access to 1.5 million items (largely pages from books and images). A next phase is planned to involve 500 databases.

**A** concrete broadband demonstration used images of a wall size fresco of a Map of Bologna in the Papal Apartments of the Vatican, using the national super-computers at CINECA (<http://www.cineca.it/>) and the Universidad Complutense. Specifications of how subsets of this map could be used over the Internet were the subject of a *tesi di laurea* at CINECA. These examples are also concrete steps in the direction of a vision of Virtual Heritage Centres (VHCs).

**N**ext, a paradigmatical model for understanding new media developments was developed which identified three integrating themes: Intelligent Multimedia Museums, Hypermedia Online and Hypermedia E-Learning. These integrating themes were used to define twenty-three projects to serve as near-term

<ol style="list-style-type: none"> <li>1) Security</li> <li>2) Legal Arrangements</li> <li>3) Storage</li> <li>4) Multilingual Mapping</li> <li>5) Semantic Mapping</li> <li>6) Digital Autonomous Cultural Objects</li> <li>7) Interface Institutes</li> <li>8) Multimodal Interfaces</li> <li>9) Spatial</li> <li>10) Temporal-Spatial</li> <li>11) Virtual Environments</li> <li>12) Virtual Heritage Centres</li> </ol>	<p>UC Louvain Valencia Team UNESCO AMP Consortium SEMKOS Consortium University of Cologne/DFG Intermedia/Interactive SIMILAR NoE ORION TN Portsmouth EoI CINECA CNR-ITABC/VHN</p>
---	---

Table 1. Twelve modules for a first step towards a DEER.

inputs for prototypes for the DEER. A concrete example of work in progress includes a reconstruction of the Monastery of Santa Cruz (now the Universidad SEK), which uses the NUME project (<http://www.storiaeinformatica.it/nume/italiano/ntitolo.html>) as a model and forms part of the Segovia in 4-D project. Meanwhile, the work on Hypermedia E-Learning Online led to collaboration with the BELE Consortium and agreements to link learning and training networks, especially in Spain and France, with existing satellite networks in South America and Russia comprising content, learning tools and new interfaces in an NoE.

**A**nother aim is to develop critical methods and thought using the DEER. Criteria and focal points for emerging European Masters degrees as well as an initial draft for course work have been identified, which can be grouped as three thematic cycles of courses: Textual and Visual Databases, Databases Processing for Conservation Intervention on Monumental Complexes, and Projecting (Planning) of Museum Communication.

**I**n addition, a team has been formed to develop further European Masters courses in the realm of digital culture. These include representatives from the Politecnico di Milano, Complutense (Madrid), Sorbonne (Paris), UNESCO’s

World Heritage Studies Programme, and the Northern European Historical Research Network (NEHRN). In addition, four representatives will reflect Masters Courses in Russia. A next step will be to develop a similar small group to represent the Mediterranean especially with respect to Greek, Hebrew and Arabic. The European Masters and Doctorates will use resources from prototypes of the DEER.

*“The so-called revolution in new media is much more fundamental than hype about more powerful computers linked by faster connections. It is changing the methods for creating, preserving, accessing, and learning (presenting) knowledge. It is changing the boundaries between disciplines and in many ways transforming the nature of knowledge itself.”*

**T**raditionally the knowledge production life-cycle was different for each medium, i.e. the production of a manuscript in a scriptorium was very different from the production of a book in a printing press and publishing house. In analogue mode, the person who produced the content in one medium was typically separate from those who developed its context and communication. Traditionally senses were also treated separately: those concerned with visual media were separate from those who worked on tactile media. In digital mode, all media and senses are potentially



interlinked, hence the rise of terms such as multi-media, inter-media, trans-media and cross-media. There are new links between content, context and communication, which leads to new interplay among cultural organisations, industry, research institutions and government. New overviews are needed to fully understand these developments, therefore E-Culture Net began to develop a prototype for such research matrices. This began by identifying basic categories for a macro- and a micro-research matrix (details available on <http://www.eculturenet.org/>). Next, a working model was developed to which members can add content. Netzspannung.org (<http://www.netzspannung.org/>) has been working on knowledge discovery tools with semantic maps and other features, which will in future enable users to walk through knowledge landscapes. A next stage will be to combine the research matrices with Netzspannung's approach to achieve new understanding of the digital knowledge production life-cycle. In the longer term the research matrices will become one of the dimensions of the DEER.

In the short term, by way of preparation for the DEER, two research topics were identified: access to (and preservation of) existing cultural knowledge, and production and preservation of new cultural knowledge. These two areas generated four integrating themes for further research which could become Integrated Projects (IPs)<sup>5</sup>: multilingual, semantic access and knowledge organisation; spatio-temporal access with historical-cultural dimensions; collaborative creation with multimodal interfaces; personal and collaborative e-learning.

Two areas for long-term research were also considered: dynamic knowledge, and new models of culture. Five further implications were identified: massive new content; new methods of scholarship; multilingual methods to reflect diversity at

national, regional and local levels; new fields of study; and changing boundaries of scholarship. These new areas for future research confirm that the so-called revolution in new media is much more fundamental than hype about more powerful computers linked by faster connections. It is changing the methods for creating, preserving, accessing and learning (presenting) knowledge. It is changing the boundaries between disciplines and in many ways transforming the nature of knowledge itself.

### NATIONAL NETWORKS

A model of national networks ensures that E-Culture Net can expand to achieve a critical mass that is representative of Europe's enormous diversity. In the course of the first year Spain established a solid national network complete with its own Website at: <http://www.ucm.es/info/eculture/index.htm>. Content pilots served as a motivation to integrate these efforts. In most European countries a permanent or temporary representative was found to lead the national network. Within the thematic network, membership expanded to sixteen European countries, and nine NAS countries.

National networks, with close links to the policy strands of national governments (e.g. via MINERVA) can address the challenges of subsidiarity and assure that local, regional and national interests become more visible and are fostered through a European E-Culture Net. They can raise awareness of Europe's remarkable unity of diversities as emphasised by Giorgio Ruffolo (of the European Parliament) and thus contribute to a new, emerging consciousness of what it means to be a European.

Written agreements with UNESCO, a number of organisations and existing projects ensure that members will

enter into a firm framework of co-operation. A permanent home for E-Culture Net has been found at the new European University of Culture in Strasbourg. This co-location with the premises of the Council of Europe (CoE) ensures a fitting and enduring home for the network.

E-Culture Net has twenty-four main participants, and an attached 132 organisations with a total of 537 researchers. In effect, the NoE is a consortium of a number of smaller, specialised NoEs which, in isolation, lack the critical mass to achieve what is now possible. The activities of E-Culture Net included two conferences for members (11-14 December 2002, 16-17 June 2003), which included guests from the Artnouveau and Netzspannung networks, resulting in agreements for further co-operation.

Next steps for E-Culture Net include the development of a Network of Excellence to integrate our goals through a prototype DEER, with joint research and concrete pilots for e-learning. The DEED (Distributed European Electronic Dynamic) resource is a prototype for the long-term DEER and will combine an integration of projects and activities and the development of networks in each country. The DEED will be updated using the defined Research Matrices. There will follow concertation meetings to link the NoE's and the IP's and a proposed accompanying measure that explores how the long-term framework for the DEER can be further operationalised.

More detailed information on E-Culture Net, including deliverables, resources and the DEER, is available from: <http://www.eculturenet.org>.

<sup>5</sup> Four such possible IPs were identified:

- i) DILIGEANT (Digital Libraries in a GEANT framework)
- ii) ACE (Augmented Cities and Environments)
- iii) CO-CREATE (Collaborative Creation)
- iv) PACE (Personal and Collaborative E-Learning).

# OPEN SOURCE AND FREE SOFTWARE SOLUTIONS: RELIABLE, STANDARDISED, FREE

ANDREW MCHUGH

of the Humanities Advanced Technology & Information Institute at the University of Glasgow examines some alternatives to commercial applications and how they can benefit the cultural heritage sector.

and leaves the user at the mercy of the developers and any strategic decisions that they may take.

The Open Source Initiative (OSI, <http://www.opensource.org/>) and the Free Software Foundation (FSF, <http://www.fsf.org/>) eschew such closed practices in favour of greater transparency, collaboration and mutual progress.

Several attempts have been made to offer a definition of free or open source software, with a variety of success. The clearest issue is that free is used in the sense of 'unrestrictive', like free speech, and is not taken to mean gratis, or without charge, a common misconception.

Frequently, distributors of these products will make them available for little or no monetary cost, but this is not implicit in the terms of their licences. Prominent criteria of the Open Source Definition include free distribution, availability of source code and provisions to ensure that modifications and derived works are permitted and should be distributable under the same terms as the licence of the original software. Further clarification is offered by the FSF definition, granting to users four central freedoms: to run the program for any purpose, to study how the program works and adapt it (implying source code availability), to redistribute copies to help your neighbour, and to improve the pro-

*"With Open Source software, the obvious winner is the consumer, as overall functionality is prioritised ahead of corporate practices and intellectual property issues."*

Many brochures, Websites, catalogues and magazines are quick to highlight that there is a proprietary commercial software solution for every IT problem faced by today's users, businesses and institutions. The user is left in little doubt as to the revolutionary potential of commercial tools, of the wrongs they can right, and of the time that will be saved in doing so. Whilst many commercial applications are certainly very effective, the inevitable maelstrom of marketing spin and rhetoric may convince an unwary user that these kinds of heavy-weight proprietary applications offer everything that one could possibly ever need and are the only choices that can meet the productivity requirements of the modern company or institution.

However, for as long as the computer world has been developing, accompanying the many advancements and innovations is a decline in terms of user choice. Commercial applications, often with proprietary file formats and restrictive licences, can entrap users in a course of operation that cannot deviate from the strategic path determined by one particular software vendor. Bugs must be tolerated, functionality cannot be customised, and if software should become obsolete or producers cease trading then picking up the pieces can be an expensive process.



Andrew McHugh

It was such issues that led to the evolution of a movement that would press for seismic changes in the culture of software production and distribution. Since the late 1970s the Free Software movement has campaigned and lobbied for a re-evaluation of the models upon which software is created and sold, and of the legal system governing its licensing, redistribution and

reuse. On 3 February 1998 in Palo Alto, California, the Open Source label was conceived, to characterise and push to the commercial fore this philosophy that had begun some twenty years before.

The state of proprietary software development and distribution has evolved into a fairly widely accepted model. A centralised and compact team of developers will design and implement a program to provide a solution, usually in keeping with the overarching commercial goal of the umbrella company. The strictly concealed raw source code that forms the comprehensible essence of a program is compiled down to a single binary executable program, which can then be distributed to users. Typically such products are accompanied by strict non-disclosure licensing agreements that prohibit their sharing, copying, reproduction or customisation. This relationship puts the software producer in a powerful position. It becomes possible to exploit customers with the introduction of costly 'essential' upgrades,



gram and release the amended version to the public so that the whole community benefits. In addition, FSF explicitly incorporates a concept that has become known as ‘copyleft’. In stark contrast with copy-right laws this is intended to preserve the wide outreach potential of free software, and prevents those redistributing free software from adding restrictions to deny other people of any of the central freedoms. Many open source licences stop short of requiring the copyleft requirement but it offers a number of advantages, ensuring that with each reworking of a particular software tool it is not necessary to ‘reinvent the wheel’. The obvious winner under such a system is the consumer, as overall functionality is prioritised ahead of corporate practices and intellectual property issues.

The most popular qualifying licence for both open source and free software is the GNU General Public Licence (GPL). The most obvious example of a software product that is subject to the GPL is the GNU/Linux Operating System.<sup>6</sup> The GPL is just one of several different licences that meet the criteria of the OSI, which each correspond to and satisfy the requirements of open source.

A useful consideration at this point is of the variety of open source software solutions that offer commensurate or improved functionality over proprietary, non OSS/FS products. Consider a fictional small museum whose IT requirements are all met using closed-source proprietary software. For example, administrative tasks such as word-processing, accounting information and maintaining a mailing list and database of exhibits are facilitated with the use of Microsoft’s Office suite of software. The museum has a simple Website, which is hosted internally, served on a Microsoft Windows machine with the built-in proprietary Internet Information Server (IIS) software. The museum’s newsletter is published using



*Screenshot of Open Office software*

Adobe’s Indesign software. Images for publication on the Website and in the newsletter are formatted, cropped and resized using Jasc’s Paint Shop Pro.

This software configuration offers a great deal in terms of convenience; each application offers powerful functionality and is no doubt familiar. However, each application listed above could be replaced by an equivalent Open Source alternative. As well as offering comparable levels of functionality, OSS/FS (open source or free) software removes the restrictions that often accompany the use of proprietary formats. The technical specifications of these formats are invariably closely guarded secrets and are frequently specific to particular applications and vendors. In addition, any decisions regarding the future development and direction of these formats and applications are completely at the manufacturer’s discretion. Some users may find themselves bound to regular expensive upgrading whenever the formats in use require it.

An OSS/FS alternative to commercial ‘office’ packages is Open Office (<http://www.openoffice.org>), a comprehensive and functionally complete suite of tools for a variety of office productivity tasks. Because its source code is freely available, there is no danger of users becoming prisoners of their chosen file

format, since anyone else is free to create their own programs for reading and editing the encoded files that Open Office creates. Concerns may still exist about reading existing documents that have been created using proprietary closed-source software, or attachments from colleagues that are in such a format. Open Office has been designed with this in mind too, and as far as possible seeks to ensure compatibility with all existing Microsoft formats.

While Open Office does not currently come equipped with an application commensurate to Microsoft’s Access, the database requirements of this fictional institution can be just as easily met using another highly regarded OSS/FS application that is becoming increasingly widely used in a variety of fields. MySQL (available for free download from: <http://www.mysql.com>) is a fully featured relational database package that can integrate within Open Office to offer a convenient free alternative to commercial database applications. Around 4 million installations worldwide and a user base including Motorola, NASA and Yahoo! emphasise its pedigree.

The museum’s Website can also be served very successfully using OSS/FS tools. The Linux operating system, often regarded as the most important development in the history of OSS/FS software, is particularly well suited to serving digital materials, with its basis in the Unix operating system. It is becoming increasingly highly regarded as a platform for desktop productivity, and it is generally regarded that, in terms of quality and reliability, Linux has already won the ‘server war’ and proved itself the best choice for such a role. It is also widely acknowledged that the best Web server available is Apache (<http://www.apache.org>), another open source technology.

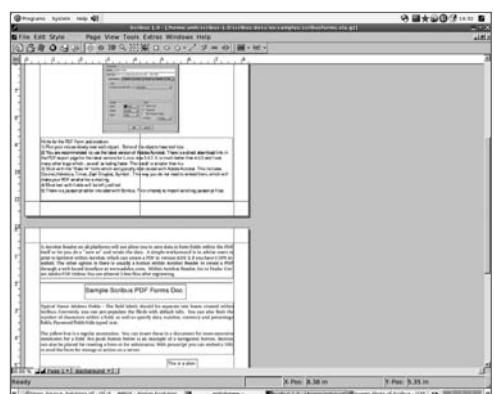
<sup>6</sup> The GNU Project was launched in 1984 to develop a complete Unix-like operating system which is free software: the GNU system (GNU is a recursive acronym for ‘GNUs Not Unix’). Variants of the GNU operating system, which use the kernel Linux, are now widely used; though these systems are often referred to as ‘Linux’, they are more accurately called GNU/Linux systems.



Finally, the two highly visual tasks that the museum currently completes could both be accomplished using OSS/FS applications. Scribus (<http://www.atlantechsolutions.com/scribusdocs/projects.html>) is an open source DTP package that, while still in its relative infancy, has garnered a great deal of praise and enthusiasm from a wide variety of users. With a selection of industry award nominations to its credit including Linux User UK's 'Best Linux Software', Scribus aims to strike a balance between ease of use and powerful functionality in the creation of a variety of published materials. The OSS/FS imaging tool of choice is The GNU Image Manipulation Program (The GIMP, <http://www.gimp.org/>) which offers a wide variety of functionality for the editing and creation of a comprehensive variety of digital picture formats. While The GIMP stops short of the exhaustive and specialised functionality included in software such as Adobe's Photoshop, it incorporates more than enough to far exceed the imaging requirements of most institutions.

*"The open source development model is geared towards mass co-operation and collaboration; many institutions and individuals working together to create something more than could otherwise be attained."*

It is clear that OSS/FS software offers sufficient functionality to facilitate a wide variety of IT and IT-related tasks. Quality and depth of functionality is facilitated by the open source development model, which is geared towards mass co-operation and collaboration: many institutions and individuals working together to create something more than could otherwise be attained. Programmers are motivated by their own interest in the software they create, as well as their own needs to have the tools that they devise unencumbered by restrictive licences. Bugs are iden-



Screenshot of Scribus software

tified and fixed with almost ruthless efficiency and worldwide, keen-eyed scrutiny ensures that security issues are quickly resolved. Similarly, open source users find themselves less likely to be the target of viruses and invasive software, which can be associated with or motivated by resentment to large corporate distributors.

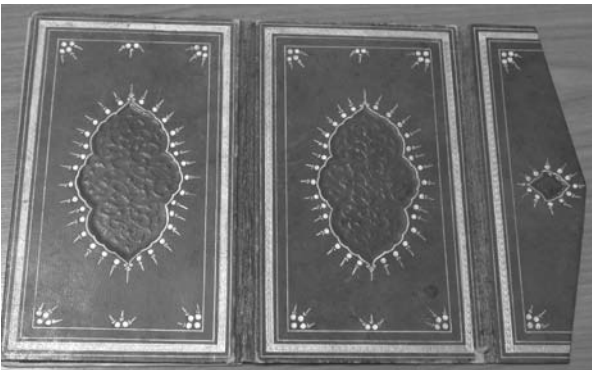
Choosing to embrace open source solutions need not be a particularly dramatic or revolutionary process but the positive outcomes in terms of increased productivity and lower overheads may be vast. By removing the risks commonly associated with software use, such as vendor collapse, unfinished bugs, potential legal exposure from restrictive licences and becoming tied to proprietary file formats, institutions are free to concentrate fully on achieving their computing goals. Any choice that offers maximum functionality with none of the oft-associated restrictions is worth at least a great deal of consideration.

[BACK TO PAGE 1](#)

# INTRODUCING DIGITISATION INTO TURKISH LIBRARIES: CURRENT ATTITUDES AND THE WAY FORWARD

SELENAY AYTAÇ

Işık University, Istanbul, Turkey



*Kaside' by Yusuf Inb-i Muhammed-177 x 110 mm-1895 (golden 11 folio)*

This Poster Paper Session was presented at the 69th IFLA Annual Conference, Berlin, Germany, on 5-6 August 2003.

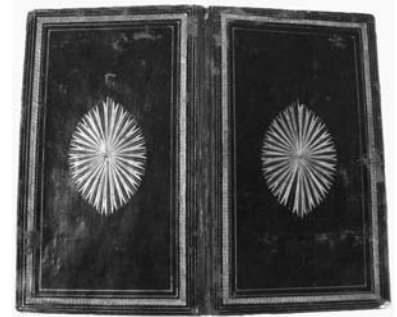
Despite the wealth of historical materials held in Turkish libraries, such as the 120,000 Ottoman manuscripts held by the Süleymaniye Library in Istanbul, only a few attempts have been made to preserve such cultural heritage by using digitisation technologies due to a lack of educated staff, technological equipment, know-how, and financial freedom [1]. In order to introduce Turkish librarians to the fundamentals of digitisation techniques and digitisation practice and to investigate the current level of knowledge of these preservation issues among library professionals, a tutorial on the digital preservation of written cultural heritage was held at the Academic Computing Conference (2 February 2003, Adana, Turkey [2]).

A questionnaire was handed out at the end of the tutorial to evaluate the

views of the attendees on the use of digitisation technology in Turkish libraries. This study involved thirty-seven participants, around a third of whom were female, representing a wide age range. Thirty-three participants responded to the survey. The institutions represented were university libraries and LIS departments, research libraries, the National Library, and the Library of Higher Education Council of Turkey. Although the majority of respondents were from university libraries, a range of positions were represented (see Chart 1) including Library and Information Science (LIS) faculty members, librarians, other academics, a multimedia specialist, and an architect.

The questionnaire was designed in two categories: open-ended questions (to gain suggestions) and closed questions (questions designed to examine how strongly the subject was satisfied with statements on a five-point scale and yes or no choices). This survey described the various attitudes of the library professionals in Turkey. Results indicate that participants were satisfied with the tutorial in general.

The initial findings from the survey revealed only a limited degree of knowledge of digitisation technology among library professionals, despite their



*Güfte Mecmuası' by Selim III, Ottoman Sultan-290 x 170 mm-18th century (golden 27 folio)*

acknowledgement of its importance (100%) in the preservation of Turkey's written cultural heritage. Some of the attendees (24%) had not heard about the issue before the tutorial or were not familiar with any international digitisation projects (60%). Many had never attended any seminar or training course on digitisation technology (82%).

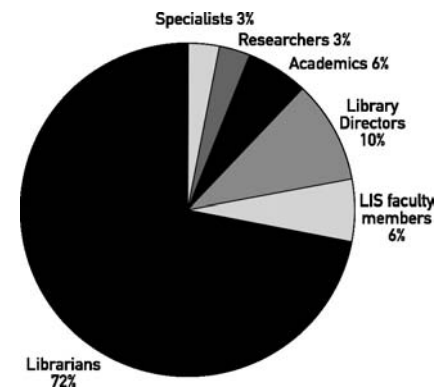
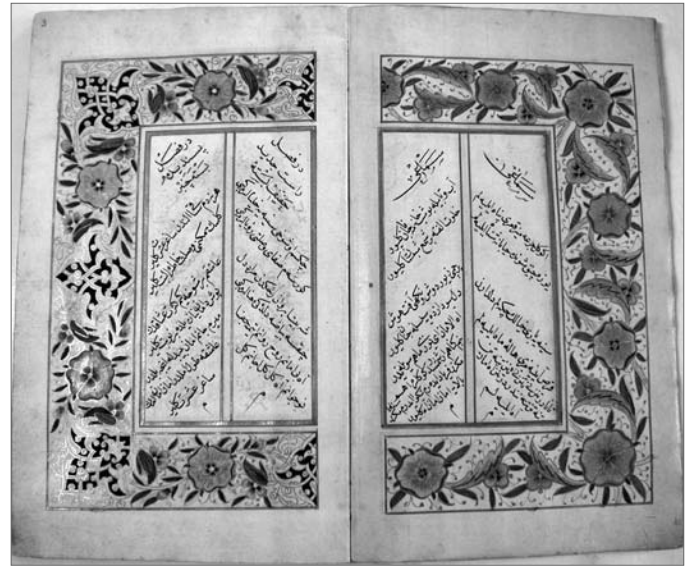
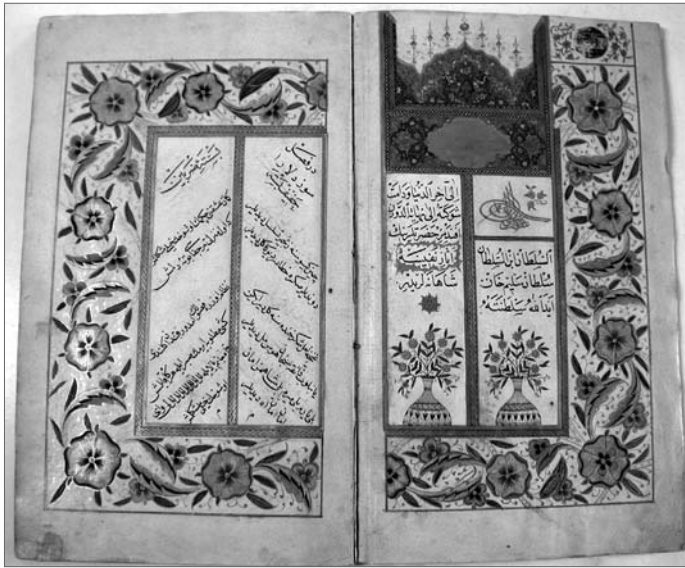


Chart 1: Positions represented

Recommendations were made to address these areas and to improve the level of knowledge of Turkish librarians with the continuous education facilities provided by LIS faculty, library associations in Turkey and by enhanced co-operation between technology-savvy associations such as Türkiye Bili\_m Derne\_i (Informatics Association of Turkey, <http://www.tbd.org.tr/>) and Türk Kütüphaneciler Derne\_i (Turkish Librarians' Association, <http://www.kutuphaneci.org.tr/>). LIS departments in Turkey



Güfte Mecmuasi' by Selim III, Ottoman Sultan-290 x 170 mm-18th century (golden 27 folio) (picture left)  
Güfte Mecmuasi' by Selim III, Ottoman Sultan-290 x 170 mm-18th century (golden 27 folio) (picture right)



Kaside' by Yusuf Inb-i Muhammed-177 x 110 mm-1895 (golden 11 folio)

should provide continuing education in new information technologies for their graduates to attain continuous improvements in current and future developments in Information Science.

In conclusion, some possible ways to improve the current state of digital preservation in Turkey include:

- **continuous education for librarians:** life-long learning by the three LIS departments which is currently available in Istanbul [3] (one department) and Ankara [4, 5] (two departments).
- **a single gateway to Turkish resources:** making images of all these manuscripts available from a single Web-based database, established with

the collaboration of all Turkish institutions, would raise their profile and encourage institutions to contribute their resources.

- **universal standards:** common universal standards for metadata and file formats should be applied. This would allow the sharing of resources with many other Web-based manuscript archives in the international area.
- **user-centred interfaces:** multilingual interfaces to these images should be set up to allow the full richness of the materials to be exploited [6].
- **quality control:** consistent quality control procedures should be applied to all components of the archive.

## REFERENCES

1. Aytaç, S., Türkiye Yazmalarının Sayısallaştırılması (Digitisation of Turkey Manuscripts) Turk Kutuphaneciligi, Turkish Librarian Association, Ankara, 16 (1), pp. 3-19, 2002.
2. Aytaç, S., Yazılı Kültürel Mirasın Sayısal Korunması, Sayısal Koleksiyonlar ve Yönetimi (Digital Preservation of Cultural Heritage, Digital Collections and Management), tutorial presented in. Academical Computing Conference, 3-5 February 2003, Çukurova University, Adana, Turkey.
3. Istanbul University, Department of Information Management, Istanbul URL: <http://www.istanbul.edu.tr/edebiyat/edebiyat/p3a8111.htm>
4. Ankara University, Department of Information Management, Ankara URL: <http://www.ankara.edu.tr/letters/bilgibelge/index.htm>
5. Hacettepe University, Department of Information Management, Ankara URL: <http://www.kut.hacettepe.edu.tr/>
6. Aytaç, S., Development of a User-Centered Digital Library for Ottoman Manuscripts, IATUL (International Association of Technological University Libraries) 24th Annual Conference, Ankara, Turkey, 2-5 June 2003, URL: [ftp://ftp.lib.metu.edu.tr/pub/doc/iatul/paper\\_presenters/AYTAC,%20Selenay-IATUL03-SelenayAytacpresentation.ppt](ftp://ftp.lib.metu.edu.tr/pub/doc/iatul/paper_presenters/AYTAC,%20Selenay-IATUL03-SelenayAytacpresentation.ppt)

# DIGITAL ACCESS TO AERIAL AND ASTRONOMICAL ARCHIVES

## Project Co-ordinator and Project

### Scientist Jean-Pierre

**De Cuyper**, of the Royal Observatory in Belgium, spoke to DigiCULT about the digitisation of astronomical and aerial photographic plates and the unique technical challenges posed by this new archiving project.



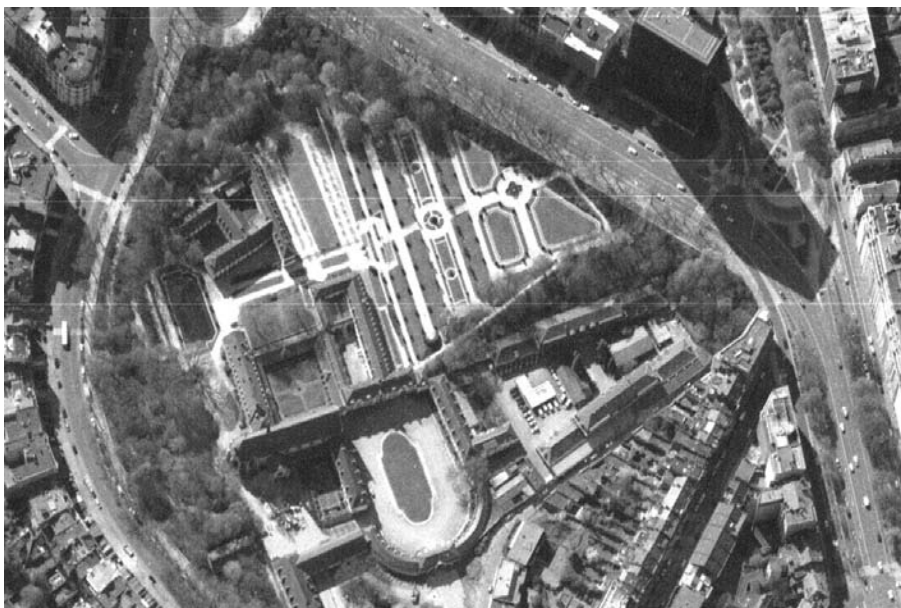
<http://africamuseum.be>) and in the astrophotographic plate archive of the Royal Observatory of Belgium (KSB, <http://www.astro.oma.be/>). In collaboration with the astronomical institutes of the Vrije Universiteit Brussel (VUB) and the Universiteit Antwerpen (UA) and AGFA-Gevaert, a world-leader in photographic matters, the goal is to acquire the necessary know-

*“Old photographic plates contain valuable information about the position and the brightness of heavenly objects that cannot be reproduced by modern instruments. Digitising these plates in an optimal way is crucial to the accuracy of the data.”*

The aim of this pilot project, financed by the Belgian Federal Public Planning Service Science Policy (<http://www.belspo.be/>), is to preserve the historic-scientific information contained in the aerial photographic archives of the National Geographical Institute (NGI, <http://www.ngi.be>) and the Royal Museum of Central Africa (KMMA,

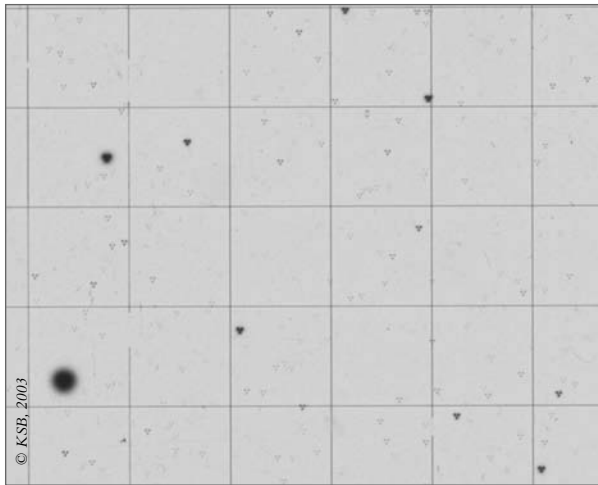
how, hardware and software to digitise the information contained in the photographic plates, as well as the associated metadata. The project set out to offer the results to the public and to make them directly usable for scientific research through the modern techniques of the information society.

The collection of the aerial photographs of the NGI contains about 80,000 greyscale images of Belgium, starting after the Second World War. These images were used for the photogrammetric extraction of all necessary information needed by cartographers to make detailed maps. As topographic reconstruction of the altitude of the surface requires a stereographic view of the landscape, these aerial photographs were taken in strips with at least 60% overlap in the flight direction and at least 15% overlap between adjacent strips. These historical aerial images are useful for private as well as professional purposes ranging from visual to photogrammetric usage. The oldest aerial photographs are on glass plates of 18 cm x 18 cm; later, polyester film rolls with images of 23 cm x 23 cm were used but cut into single photographic sheets for further use. The KMMA's collection contains 300,000 aerial photographs on glass plates (partly only in the form of paper contact copies) taken by the NGI in the Congo, Rwanda and Burundi after World War II until the end of the colonial occupation in 1960. These images are often the only source of visual information available for ecological studies concerning erosion, human occupation, and vegetation. The first aerial photographs were taken on glass plates. A special machine was designed that made it possible to exchange the glass plates in the aerial camera automatically during the flight. Later, 24-cm film rolls were used.



National Geographical Institute at the old Ter Kameren abbey in Brussels

Used with kind permission of N.G.I. Belgium



*A plate from the Carte du Ciel collection, enlarged to show triple exposure.*

More recently, colour film has been used; however, for photogrammetric purposes greyscale film is still being used by the NGI.

The KSB archives contain 30,000 photographic images of the sky mostly on glass plates, some on polyester sheets, with dimensions ranging between 16 cm x 16 cm and 30 cm x 30 cm, medium to high resolution spectra of individual stars and low resolution field spectra used for star classification. The oldest collection is formed by the plates taken around 1900 for the Carte du Ciel international project, which aimed to conserve an image of the sky at that time. Two series of plates were taken. Long exposures, containing three exposures of about twenty minutes on each plate, were used to make an atlas of the heavens using newly developed heliographic technology. These plates used fine-grained emulsions in order to obtain the best possible geometric (astrometric) precision. A second series of short-time multiple exposure plates were taken to produce the first astrographic catalogue. To photograph fainter stars and to take shorter exposures, coarser grained emulsions were used. Photographic plates were also used to take spectra of individual stars or to make low resolution field spectra (i.e. a sky image

with the light of the individual stars split up in wavelength). Later on, emulsions on polyester sheets were used. The largest collection consists of the images taken for the discovery of comets and small planets. Usage of photographic techniques for astronomical purposes gradually stopped after the introduction of the CCD detector in the 1980s.

Plates consist of a distribution of silver grains embedded in a gelatine layer fixed on a glass plate or polyester sheet. As such, they are very sensitive to changes in temperature, relative humidity (RH), and chemicals, and are at great risk of degradation, such as chemical reactions from fingerprints, humidity causing destructive fungi and so on. Most photographic collections were/are stored in conditions that are far from being ideal. In order to improve the lifetime of its astrophotographic plates, the KSB is constructing an acclimatised plate archive that will be kept at 18°C and 50% RH. As the plates have been stored in more humid conditions up till now, a gradual change of the relative humidity will have to be applied. The photochemical re-bathing of old plates, in order to remove remains of historic fixing agents (which cause a gradual fading of the image and deposit a purple shade), dust and fungi, is currently being studied in order to determine possible distortions and to find out what effect this bathing has on the lifetime of the plates.

These old photographic plates contain valuable information about the position and the brightness of heavenly objects that cannot be reproduced. By comparing stellar positions obtained from these historical photographic plates with positions

obtained recently, very accurate stellar proper motions can be calculated for stars as faint as magnitude 15. This goes deeper than the ones obtained from the HIPPARCOS astrometric satellite. Digitising these plates in an optimal way (using a very precise XY table, the appropriate optics, and a stable environment) is crucial to the accuracy of the data. Furthermore, the data reduction software used for calculating the stellar positions should use an appropriate fitting profile.

Digitising a photographic image can be done 'on the fly' (using a digital detector moving with constant speed in one direction) or 'on the step' (using a digital detector at rest). The way traditional scanners work means some of the finer details of the image can be smeared out over neighbouring pixels, creating a soft-looking image. However, certain photogrammetric scanners create a 'dead' time between integration intervals in order to produce a hard(er) image. The level of detail in the photographic plates requires a very high optical resolution to produce high geometric and radiometric accuracy in the digital copy, and precludes the use of sharpening filters in the scanning process. The D4A digitiser that will be used for this project is able to digitise photographic images and spectra on glass plates and polyester sheets as well as on polyester film rolls to an extremely high level of precision.

This project will develop a two-dimensional plate digitiser that will operate on the step in order to create a precise digital optical copy of the original image. For most astronomical applications, overlapping digital sub-images can be used. Bright stars in the overlaps are used to tie up the whole image and to transform measured X and Y positions on the image into celestial 'alpha' and 'delta' co-ordinates. The accuracy of the photographs depends on the type of emulsion used, the type of supporting layer (glass plate or



© KSB, 2003

*The Schmidt telescope building at the KSB in Ukkel (Brussels) that will house in its basements the D4A digitiser in a climate controlled clean room with adjacent plate archive.*

Carte du Ciel plates the digitised monochrome heliographic atlas images can be used for this purpose.

**A**erial photographs need to be digitised as raster images, requiring an accurate stepping in both the X and Y directions. The digital image can then be stored as one large image, or in the form of a tiled-TIFF file containing the adjacent non-overlapping images of the individual footprints.

**T**he D4A pilot project will develop the necessary hardware and software to digitise the plates by the end of 2005. The real production phase should be realised through a future follow-up project. When the D4A high-resolution scanner is available, this project will hold an international workshop based on the methods and challenges of digitising collections. A digital catalogue is under construction that will be integrated into a user-friendly database accessible from the Internet to allow public access to this important scientific and historic information.

[BACK TO PAGE 1](#)

polyester sheet), the optical quality of the telescope used and the exposure time. Photographic plates containing stellar images can have a density range of 5 (i.e. a grey scale ranging from 0 to 100,000) and a sub-micron stellar positional accuracy. The photographic image is made up of an irregular distribution of developed grains of varying sizes whereas a digital image consists of equally spaced and sized square or rectangular pixels. In order to capture the level of accuracy of the analogue photographic images as closely as possible, a digital detector is needed with at least a 10-bit ADU (Analogue-to-Digital converter Unit) read-out and a pixel size of about 5 micrometres. We will mount the digital camera above the plate, perpendicular to its surface, and use an air-bearing open frame XY table to allow us to position the plate with a geometric accuracy of some ten nanometres. We will use a two-sided telecentric lens; this is to ensure that, if the original image is not perfectly flat, the introduced error will only slightly enlarge the projected image of a point source, while keeping it isotropic and without displacing it. The part of the footprint of the

telecentric objective used will be limited to its central part where the distortion is less than a pre-defined maximum. In this way an 'optical' contact copy of the original image onto the digital detector will be achieved. In order to be able to reach and maintain a high geometric and radiometric accuracy, the digitiser will be placed in a clean room, at a temperature of  $18^{\circ}\text{C} \pm 0.1^{\circ}\text{C}$  (1 sigma) and a relative humidity of  $50\% \text{ RH} \pm 1\% \text{ RH}$  (1 sigma). These conditions will be obtained by circulating air over a chilled water radiator in order to produce a stable laminar flow. To obtain a sharp image, the air between the photographic emulsion and the objective should, however, be kept at rest. The exact technical details are still under study.

**I**n order to save the historic notes on the glass side of the hand-measured astrophotographic plates, a graphical scanner will be used to produce low-resolution 'pre-scan' digital images. These will be coupled to the database and used to re-determine the position of the plate centre and the magnitude range of stars, etc. measurable/visible on the plate. For the