

DigiCULT .Info

Issue 8 A Newsletter on Digital Culture

ISSN 1609-3941

August 2004



INTRODUCTION

For the first time, DigiCULT has devoted a section of this publication to a detailed focus on the cultural and scientific heritage work being carried out in one particular country. *DigiCULT.Info's* Content Editor, Daisy Abbott, spent a week interviewing sector professionals in France and their thoughts and experiences are presented in this special section, alongside articles and reports from French projects and events.

Continued next page



© Daisy Abbott, HATII, University of Glasgow, 2004

Technology explained and applied ...

- 21 European Conference on Encoding Research Aids
- 21 Making Handwritten Archives Documents accessible to the Public with a Document Image Analysis System
- 26 Interfaces for digital libraries at CNAM (1993–2003)
- 40 CWIS – Collection Workflow Integration System
- 40 Steganography Explained
- 48 Using Spatial Knowledge to Classify Metadata

France in Focus

- 5 DigiCULT Focus on France
- 5 Jean-Pierre Dalbera, Chef de la Mission de la recherche et de la technologie, Ministère de la culture et de la communication
- 7 Alain Maulny, Conservateur du patrimoine, Chargé de mission à la mission de la recherche et de la technologie
- 9 Jean-Pierre Teil, Head of the Constance Programme, Archives Nationales, Centre des Archives Contemporaines
- 12 Catherine Dhérent, Chef de la mission pour la gestion de la production documentaire et des archives, Bibliothèque Nationale de France
- 15 Catherine Lupovici, Directeur, Département de la Bibliothèque Numérique, Bibliothèque Nationale de France
- 16 Daniel Teruggi, Directeur recherche et experimentation, Institut national de l'audiovisuel (INA)
- 19 Abdelaziz Abid, Division de la Société de l'information, UNESCO

News, events ...

- 44 Event Report: PrestoSpace and Audiovisual Preservation
- 49 News From Around the Globe
- 53 ERPANET/CODATA Seminar final report
- 53 International Internet Preservation Consortium
- 53 New Curation Heritage Reports
- 53 Digital Curation Center

Challenges, strategic issues, new initiatives ...

- 35 Recognising Advances in Digital Preservation
- 37 Edition Production Technology (EPT) and the ARCHway Project
- 41 Memory of The Netherlands (Het Geheugen van Nederland): a programmatic approach to the digitisation of cultural-historical resources in The Netherlands
- 51 Changing Technologies Changing Methods: An Interview with Paola Moscati
- 51 Training of the Future Archivists and Records Managers of Europe: the situation in the UK and Ireland

News from DigiCULT's Regional Correspondents ...

- 30 Greece
- 30 Italy
- 31 Lithuania
- 33 The Netherlands
- 34 Serbia and Montenegro



From the stately elegance of the castle at Fontainebleau to the innovative design of the recently built Bibliothèque Nationale de France, from the fine art on display at the Louvre to the new visualisation techniques being used in Web sites, from the Ministry of Culture's emphasis on public access to heritage to the 'invisible' research being carried out at the Institut national de l'audiovisuel,¹ making the most of culture in France demands approaches as diverse as the culture itself.

While the articles towards the end of the section describe the details of smaller projects, Jean-Pierre Dalbera and Alain Maulny from the Ministère de la culture et de la communication demonstrate the effects of a coherent, nationwide policy from the top down. Jean-Pierre Teil of the Archives Nationales agrees that a pool of resources at the national level is crucial to best organise regional and local activities, and both he and Catherine Dhérent highlight the evolution of the field of preservation of records in recent years.

Along with Daniel Teruggi from INA, Teil and Dhérent are emphatic that archivists cannot wait for the 'perfect format' and that, to ensure information is preserved, action must be taken immediately, but always with an eye on the future, ensuring that an appropriate migration strategy is in place to safeguard the longevity of the data. Specific work on archival access is presented in the article 'Making Handwritten Archives Documents accessible to the Public with a Document Image Analysis System' by a team from IRISA (<http://www.irisa.fr/>) in collaboration with several different archives from across France. As shown in the DigiCULT Report: Technological Landscapes for Tomorrow's Cultural Economy, pp. 162-174, 'Turning Archival Databases into Goldmines: The Genealogy Case', genealogists are driving the demand for fuller access to public records and other archival

material. However without effective annotation of hand written records users may find themselves overwhelmed by the need to leaf through a sizeable number of digital images. The research team at the IRISA/INRIA laboratory in Rennes, France presents a family of tools and approaches developed for a document image analysis system which is based on automatic annotation of image records. Images not fit for automatic methods are supported by manual annotation which is managed on a collective annotation platform for the user community.

Catherine Lupovici, Director of the BnF Digital Library, describes her work in digitising and providing access to the materials held at the BnF, while the article from the team at the Conservatoire National des Arts et Métiers presents the technical aspects of digital library interfaces. In the final interview, Abdelaziz Abid from the Division de la Société de l'information at UNESCO describes the Memory of the World project, its benefits, and place within the work of UNESCO, casting an international perspective on work taking place within France.

To encourage the role of digital preservation in the production and management of born-digital resources, the Digital Preservation Coalition has established the Pilgrim Trust Conservation Awards to recognise innovation and achievement in the area. Adam Rusbridge takes us through the shortlist which includes specific tools, programmes and large-scale initiatives that all seek to advance the successful preservation of our digital cultural society. And the winner is...

An area of research that has received particular attention of late is the preservation of audiovisual resources. The nature of audiovisual material poses immediate challenges that the European Commission PrestoSpace project, an integrated project within the IST sixth frame-

work programme, has firmly placed on its radar. The User Requirements meeting held in Amsterdam 18-19 March 2004, formulated the PrestoSpace Plan in response to the 'real needs' assessment of the users in the field. In closing the delegates all agreed that the problems were 'larger than any one sector, viewpoint or institution' - which needs to be recognised with a structure that ensures comprehensive European co-operation.

The 'Memory of the Netherlands', is an initiative to build, manage and make available the cultural richness of the Netherlands to a wider audience, based on a digitisation programme. It first went 'live' online in May 2003 - and is a work in progress. We therefore feel extremely fortunate for Paul Doorenbosch's frank assessment of the first phase. Of particular interest are the lessons learnt, which mirror many of the concerns from other large-scale digitisation initiatives in the cultural heritage community. A notable challenge is the development and deployment of educational applications of digital heritage. Closer collaboration between the cultural heritage sector and education sector reinforces the social value and to a lesser extent the commercial value of the former. In the 'Memory of the Netherlands', executing the vision involved abandoning the initial idea to create a separate education interface in favour of varied applications designed along specific user requirements, with the focus on the presentation of the content.

In the ARCHway project from the University of Kentucky the focus is on scholarly collaboration. The computing and humanities communities were brought together to build and integrate electronic editing tools in the building of image-based electronic editions from cultural material. The toolkit being

¹ The importance of INA's research is highlighted in a report on Prestospace, also in this issue. INA is the co-ordinator of the Prestospace project.

developed is called Edition Production Technology (EPT) and is build on top of Eclipse (<http://eclipse.org>) a programming environment selected for its collaborative teaching and learning features. It is hoped that the tools will provide practical solutions in the preservation and use of hand-written manuscripts.

With all the technology advancement whizzing by on our screens, Paola Moscati asks whether we are aware of the effect that such progress is having on the research methods used to study the cultural heritage material. Her journal *Archeologia e Calcolatori* first pub-

lished in 1990 is dedicated to this investigation. Her fifteen years of experience leaves us with the cautionary message that: ‘Methods and results are the most important thing, not the technology used to achieve them, and often the “wide” solutions offered by the application of some technologies overlook the scientific part of discovery’.

DigiCULT continues to extend its network of regional correspondents. Their work is increasingly central to our activities as not only do they contribute quarterly reviews of activities in their countries, but they point the DigiCULT team to topics and authors for feature

articles. Readers will find much in these reports of value as they produce a picture of the extent to which ICT continues to transform management, study, interpretation, presentation, and use and enjoyment of the cultural and scientific heritage across Europe. There are a few European countries that do not yet have a national correspondent and those interested should contact the *DigiCULT.Info* Content Editor, Daisy Abbott at D.Abbott@hatii.arts.gla.ac.uk. To our current correspondents the Editors wish to extend our warm thanks for all your efforts that have helped to make *DigiCULT.Info* so rich in content.

Seamus, John & Daisy.

DIGICULT PUBLICATIONS

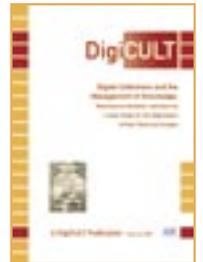
DigiCULT Thematic Issues



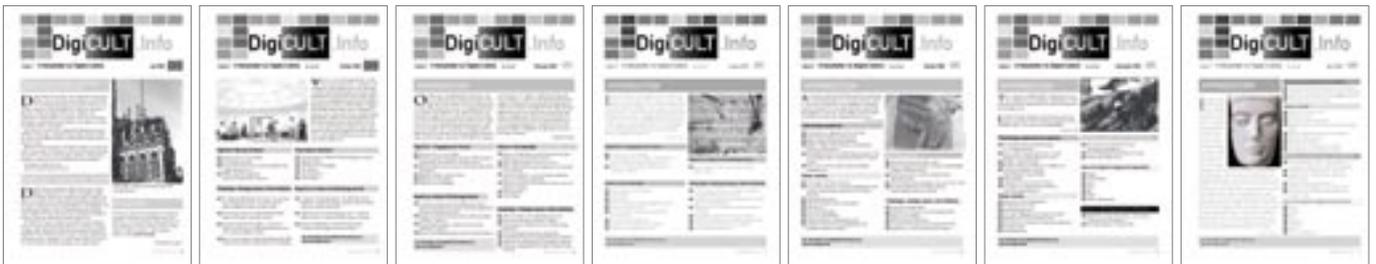
DigiCULT Technology Reports



DigiCULT Special Publication: Emblemsbook



DigiCULT.Info



CONTACT THE PEOPLE BEHIND DigiCULT FORUM

Salzburg Research

Forschungsgesellschaft m.b.H

salzburg|research

DigiCULT Project Manager

John Pereira, john.pereira@salzburgresearch.at

DigiCULT Thematic Issue Scientific Editor

Guntram Geser, guntram.geser@salzburgresearch.at

DigiCULT Network Co-ordinator

Birgit Retsch, birgit.retsch@salzburgresearch.at

Copyright Notice:

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

HATII – Humanities Advanced Technology and Information Institute



Director, HATII & ERPANET

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

DigiCULT Forum Technology Assessor

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

DigiCULT.Info Content & Managing Editor

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

DigiCULT Web Developer

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

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

DIGICULT FOCUS ON FRANCE

Jean-Pierre Dalbera, Chef de la Mission de la recherche et de la technologie, Ministère de la culture et de la communication

As MINERVA published the report 'Co-ordinating Digitisation in Europe' a progress report of the National Representatives Group for 2003,² the Chef de la Mission de la recherche et de la technologie, Jean-Pierre Dalbera, introduced DigiCULT to some of the best French examples of digitised cultural heritage now available on the Web.

It is very important to realise that designing an effective and valuable learning resource for online delivery relies on the skills and expertise of more than one group of people. Researchers, experimenters, educators, designers and technicians are all involved in creating detailed and accessible cultural heritage Web sites and the understanding of all of these processes is important to the presentation of authoritative and



© Daisy Abbott, HATII, University of Glasgow, 2004

Jean-Pierre Dalbera

²The 2002 report is available from Minerva's Web site: <http://www.minervaeurope.org/publications/globalreport.htm>.



relevant information. This approach requires extensive co-ordination both within and across projects to maximise efficient use of resources. By discussing five different online products produced by French teams, Jean-Pierre Dalbera demonstrated the Ministry of Culture's cohesive approach to the digitisation and presentation of educational materials for the Web.

PARIS, VILLE ANTIQUE ³

(<http://www.paris.culture.fr/>)

One of the collections of archaeological sites, which are listed, along with other Ministry of Culture Web sites, on <http://www.culture.gouv.fr/culture/arnat/fr/>, is *Paris, ville antique*. This site is truly multimedia, with different kinds of information meaningfully layered to allow users to explore interactively various aspects of historic Paris, in relation to contemporary views of the city. The site will soon be available in English as well as French and can be purchased as a printed product.

Dalbera: 'This site is one of the most recent projects we have been working on, and I think it is one of the best. It represents a collaboration between archaeologists and multimedia designers, combining tools such as virtual reality with images of floor plans and maps. The mixture of 3D reconstructions and animation and the interaction that is possible between text and images make it a very useful and memorable learning tool. For example, difficult words appear in an online glossary, and the meanings can be ascertained simply by hovering over the word with your mouse. In addition to the architectural reconstructions, the site presents digitised objects from museums, which are explored in a similar fashion. What is particularly interesting about this site is that, alongside the digital objects themselves, it presents the research methods in archaeology. Archaeology is one of the services of the Ministry of Culture, and making the utilities themselves available for exploration (in conjunction with the digital materials) will enrich an understanding of the whole subject area.'

DE SAQQARA AU MUSÉE DU LOUVRE

(<http://www.culture.gouv.fr/culture/arnat/saqqara/fr/index.html>)

This Web site presents the *chapelle*, a small part of a *mastaba* (on display in the Louvre in Paris), in its original context in Egypt. The site opens with an introduction in Flash, followed by a slideshow with audio commentary and text labels on the history of the purchase of the *mastaba* by France.

Dalbera: 'The Ministry of Culture funds Egyptological study within the Louvre, so this Web site explains the research conducted in Egypt. Like *Paris, ville antique*, the means of discovery and digitisation of the tomb and found objects is presented alongside the digital materials, with equal importance. For example, there is a step-by-step translation of the hieroglyphs displayed on the walls, whilst each panel is shown in its original position within the *chapelle*. The techniques used in the presentation of material and the production of the Web site are as interesting and relevant as the subject of study! Digitisation is a technical process; however, our policy recognises that not only technical work is required to realise a resource such as this in a published form. A product such as this Web site combines technical expertise with design and usability, user interaction and, in this instance, Egyptology. We must remember that cultural heritage products like this are always scientific as well, combining the skills of experts in the subject with those of technicians and designers.'

Both *Paris, ville antique* and *Saqqara* can be accessed from the *Grands sites archéologiques* portal at <http://www.culture.gouv.fr/culture/arnat/fr/> which contains links to many other archaeological projects from across France.

MARC-ANTOINE CHARPENTIER – MUSICIEN DU BAROQUE

(<http://www.charpentier.culture.fr>)

This Web site opens with a Flash intro-

duction to the music and work of Marc-Antoine Charpentier, but the content of the site itself has a much wider scope; it presents all of the research in musicology of that period.

Dalbera: 'This site combines many different media – audio clips of the music of course, with images of manuscripts and digital lyrics, and prints of some of the places in which Charpentier's works were first performed. It is an excellent learning resource as it offers the opportunity to access materials that usually only music researchers would know about. It also presents the scientific side of musicology with the same reverence as the art of this great composer.'

Other Web sites celebrating the work of French men and women are available from the *Célébrations nationales* portal at <http://www.celebrations.culture.fr/>.

PORTALS TO ONLINE RESOURCES

'Although many valuable links can be followed from the Web site <http://www.culture.fr/>, including contemporary sites for scientific and cultural heritage professionals and virtual communities, there are many other Web sites which I think are especially worthy of mention – firstly, *Science et patrimoine culturel* (<http://www.culture.gouv.fr/culture/conservation/fr/>), then *Le Laboratoire de Recherche des Monuments Historiques* (<http://www.lrmh.fr/>) which provides links to projects and information on some of the scientific works funded by the Ministry of Research, the *Centre de recherche en conservation des documents graphiques* (<http://www.crcdg.culture.fr>) and the *Centre de recherche et de restauration des musées de France* (<http://www.c2rmf.fr/>). Last but not least, the *Catalogue des fonds numérisés*, which can be found at http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_02.htm.'

³ Many of the Web sites produced by the Ministry of Culture and available through the portal <http://www.culture.gouv.fr/culture/arnat/fr/> are winners of the Prix Möbius des Multimédias, a prize for Web site design (for more information, see <http://www.prix-mobius.net>).

3D MONUMENTS

<http://www.map.archi.fr/3d-monuments/>

‘Last year I launched a big project with Jean-Marie Besnier and Michel Florenzano, called *3D Monuments*. It presents three-dimensional reconstructions of monuments throughout the world using the 3D Virtools plugin for display. The site also includes the methodology of the construction process including descriptions and images of the different techniques used – photomodelling and 3D scanning to produce point clouds. The tools used within sites such as this offer a new means of visiting world heritage. For example, a section of the site is devoted to le théâtre romain d’Arles, which it is possible to view as augmented reality through a VR helmet. This is currently limited to the accessibility of the hardware, but will have many uses in the future when the technology has become more widespread. Of course, the digital models must be created first! It is important to this project that we present only the reality of the research into each site, with no speculation.

☛ The change in use of 3D technologies is very important to note. These models are created by architects using a “digital maquette”, not engineers with a tool such as AutoCAD. It can be seen that a digital model can be created from primary sources, and the database behind the model contains information on all parts of the building as described by an *expert*: an architect or an archaeologist. This is an excellent method for preserving the scientific basis of these presentations. Not enough 3D reconstructions are exact, and the facility to place an object or portion of the whole back into its original context, and the simultaneous presentation of the science behind these discoveries is one of the most important issues which will shape the future of this subject. Through projects such as *3D Monuments* we can present expert knowledge to the public in a highly accessible way. 3D models aid investigative understanding, but they must have an editorial context. We publish these

data and visualisations as a product allowing offline access to this high-quality knowledge and design. As some of the 3D visualisation tools are still not ergonomic and smooth enough, offline access is an advantage. Web sites and products such as this, while still having a little room for improvement in terms of usability, retain the full integrity of the data with the interpretations and explanations of experts in the subject. Information is presented in order to show the multiple layers of scientific and cultural research that is being undertaken. At the Ministry of Culture, we are very proud of our role in supporting this work and ensuring it reaches a wider audience.’

The integration of cultural heritage with the science of its discovery and investigation is a theme throughout the Web sites being produced by projects under the Ministry of Culture. The five Web sites described show the attention being paid to accuracy, detail, and the interactive layering of information to provide users with a product that is as relevant and personalised as possible. These sites are not simply authoritative information sources, but comprehensive learning tools. Science, technology and culture converge to provide innovative access to fascinating heritage information. The Ministère de la culture has every right to be proud of its achievements.

Alain Maulny, Conservateur du patrimoine, Chargé de mission à la mission de la recherche et de la technologie

From his office in the heart of Paris, Alain Maulny spoke to DigiCULT about the Ministry of Culture’s steps to encourage digitisation of cultural objects within France and some of the progress made over the last ten years.²

☛ In 1996, the Ministry of Culture in France launched a digitisation plan, and



Alain Maulny

since that date we have digitised over 5 million documents. The basic principle behind the plan is simple – credits are given to museums, libraries, archives and other heritage institutions to assist them in digitising their collections. However, to do it properly and to build up a coherent national policy, we must liaise with collections of associations and foundations. Although we do not give funds to large institutions like the Louvre or BnF as they have their own budgets, we provide financial assistance to smaller public institutions and foundations and we encourage the building of relationships between all sizes of institution to maintain consistent and standardised plans, indexes, and so on. Although the documentation is not particularly centralised, the central policy stresses that we exist to *encourage* high-quality standardised digitisation across France, not to take over the process entirely. The credits are distributed over the whole of France, are representative for each region, and the funding we offer is often matched or exceeded at a regional or local level.

All projects can be accessed through a single portal: *Numérisation du patrimoine culturel* (<http://www.numerique.culture.fr/>). This site provides links to the 750 separate resources from 376 different institutions, making as much of our digital resources as possible available to the public. The portal can be browsed by lists and

² DigiCULT would like to extend warm thanks to the translator who made this interview possible.



gives a description of each resource and a link to the project's Web site, so that digital objects free from copyright can be visited (unfortunately, there is necessarily very little information on twentieth-century resources due to rights restrictions). It also presents all projects by particular institutions, and the lists show how active our various archives are in the field of digitisation. Archivists often produce a product such as a CD-ROM, which is very useful to provide as an alternative to the fragile original objects. We consider this catalogue to be extremely successful and valuable. It is the first example of this kind of inventory on this scale and is a model for further work at a European level. It has already inspired some work within MINERVA on the new MICHAEL portal (Multilingual Inventory of Cultural Heritage), which comprises our catalogue as well as others. The reason it works so well is that it provides a general view for people to see what cultural heritage has been digitised over our entire country. We are very proud of it.³

The organisations digitise all kinds of different objects of cultural significance: examples include photographs, blueprints, books, glass plates and precious objects. The data relating to each object are added to the database of the service that owns the object. Although the databases are separate, each must adhere to the markup rules and guidelines we set out. Each different direction has a group of people who together designed an appropriate common system for all objects. We also encourage the digitisation of audio and moving image materials.

To aid in the selection of materials, a catalogue was created, using XML, which classified what had already been digitised through projects – initially only by the Ministry of Culture, but now across a wider group of projects in France. Similar description of collections will be used in Belgium and in Italy. The most precious items have priority, although our selection

process is also affected by the staff members working within each institution – they have to show an interest in digitisation, otherwise it is unfeasible to pursue it. One example is the digitisation of manuscript illuminations; we have 14,000 images at the moment and by the end of 2004 we will have included 92,000 illuminations held in public libraries. These images are available online from <http://www.enluminures.culture.fr/>. We also select materials for digitisation based on a geographical strategy; for example, we are currently working on projects to digitise marks on jewellery and a collection of precious porcelain. These collections should be understood as complete, therefore, we will always try to complete the precious collections in one village, before moving on to an institution somewhere else in France. Our selection strategy does take into account materials that are under threat, to digitise them for preservation. However, we tend to work by completing whole collections, including modern objects, rather than digitising individual items, therefore an object under threat may not necessarily be our priority. Our selection process is also, necessarily, driven by the interests of the public, who are the ones who will eventually be accessing these digital resources. One example is the digitisation of old public registers, where the originals are too fragile for public research, driven by the great interest in genealogy. A good representation of this plan is the civil register of Paris, which will take four or five years to digitise completely. Another way to understand our selection process comes from the project within the Central Union of Decorative Arts (<http://www.ucad.fr/>). They have over 5000 works, each with hundreds of pages; therefore, by collaboration with the experts on these works, the best 400 were chosen for digitisation. This makes the task more manageable, while still providing an excellent representative sample of the cultural objects available at this institution.

When this programme was first planned, back in 1994, our aim was to feed the roots of information, i.e. to make information about these objects widely available, and not initially for preservation of the objects themselves. However, taking steps towards preservation by digital means is expensive and so the two activities have been combined: we now store a high-resolution master for the future. Nevertheless, our main objective is still to provide access.

The past ten years have seen a lot of progress in the Ministry of Culture's digitisation programme. Before the national policy was put in place, all candidates used to digitise separately. This was both expensive and incoherent in terms of interoperability and strategy. So, one company, Jouve (<http://www.jouve.org>), was put in charge to standardise the methods used, and went on to run training workshops for digitisers across France. The quality of digital documents produced by some companies often did not meet the very precise standards we required. The Ministry of Culture deal with projects individually, which can be a complicated process, but our relationships are all informed by the central direction of the programme and by the growing pool of expertise which is being developed by experience and training, and the standards of digitisation are improving greatly.

Another problem commonly experienced by digitisers is the movement of precious items. There is a significant security risk and the logistics of transport are often difficult. For example, institutions often have to take out special temporary insurance for the objects to protect against damage or theft that occurs outside the usual place of storage, or need to ensure that adequate security is provided.

However, the single largest problem that faces us is what medium to use

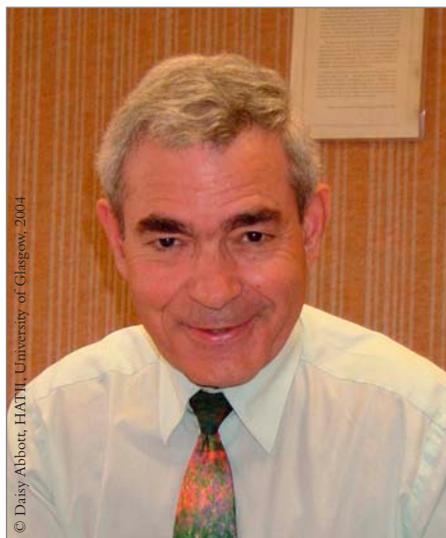
³ Search the catalogue at http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_02.htm.

for storage in the long term. We currently use CDs (no DVDs yet) and we have already experienced some degradation of data. We check the TIFF data both before and after the burning process and require that the producers retain a copy of the content for a short time (three to six months only), but this is no guarantee against loss of valuable information in the long term. Additionally, the data sources are not saved by the producers. One of the problems with CDs is that there is no standardisation between different producers – each company makes the CD differently and, therefore, reducing risk on one type may not protect against data loss on CDs from a different supplier. This problem has led to the development of a scientific community within France to study the fabrication of CDs and perhaps one day to solve what I see as the industry's biggest problem.

I am very proud of our achievements to date. We have created over 5 million digital documents and 4000 hours of audio. We aim to publish many more of these documents online (in JPEG format) for public access and to begin putting audio files on the Web as well, as none of the audio materials are currently online. This task is very time-consuming and digital rights become a significant issue. In 2003, IRCAM (<http://www.ircam.fr>) was asked to create a Web tool to allow access to audio materials. The tool is still in development, but initial work looks very promising. We think that the work carried out by the Ministry of Culture in the last ten years has had a profound effect on digital access to our heritage and we hope it continues to do so well into the future.'

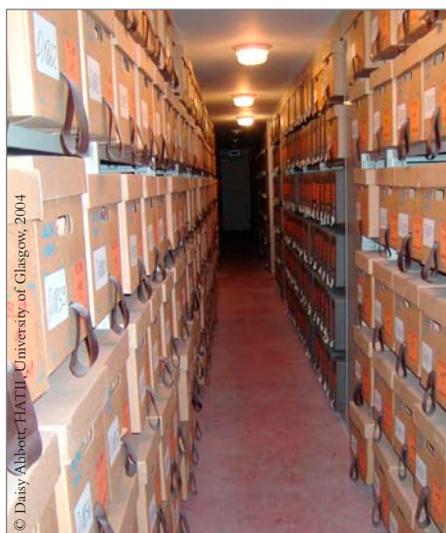
Jean-Pierre Teil, Head of the Constance Programme, Archives Nationales, Centre des Archives Contemporaines

6 I started working with the National Archives in 1977, before which I worked



Jean-Pierre Teil

in computing for the research centre of computing at the National School of Mines here in Fontainebleau. The new centre for archives that was set up at that time required practical help and experience in computerising the new repository. I began the task of using computers to manage the 80 km of shelves of records and developed the initial software that was used to manage our acquisitions: records totalling five or six kilometres of shelf space every year. Our archival holdings have grown ever since, and now nearly fill the storage space with records occupying more than 200 km of shelves.



View between shelves on one of the five underground levels of the Centre des Archives Contemporaines. Each level has 40km of storage space. The stable environment of the underground levels make temperature control much easier.

The setting up of a new specialised department to manage the electronic archives began in 1983. It was a dedicated computing centre, with a mainframe computer, and archivists working alongside information technology staff to handle the huge amounts of data coming in, for example, six thousand magnetic tapes which arrived and were added to the collections. Three years after my work at the

'Archives have many "oysters", but few "pearls" – the problem is: how to select the best pearls.'

Archives Nationales began, the head of the archives in Paris launched the Constance programme (CONSeRvation et Traitements des Archives Nouvelles Constituées par l'Electronique) to deal with e-archives. As this was only 1981, it was a very early foray into digital techniques.⁴ In 1984, I moved from my previous position to take charge of the electronic archives at the new centre in Fontainebleau. In retrospect, we can identify mistakes that were made in these early planning stages due to our collective inexperience in dealing with and managing electronic archives. Electronic archives are comprised of many oysters, but very few pearls – with hindsight, we should have perhaps concentrated on how to select the best 'pearls' from all of these magnetic tapes.

Ten years after the work on the new centre for archives began, we ran our first experiment: to demonstrate to archivists how to deal with technical metadata requirements. By 1989 the centre employed twelve members of staff, five of whom worked full-time for the Constance Programme. This archive service was for all ministries of the government, but changes to the administrative structure led to staff numbers being reduced in 1995. The changes within governmental ministries

⁴ For a 2003 ERPA.net workshop presentation on the work of Constance over the last twenty years, please see: <http://www.erpanet.org/www/products/bern/presentations/WS%20Bern%20Presentation%20Teil.pdf>.

may have been at least peripherally related to the 'big bang' in administration, government and society due to the rise of the Internet and the Information Society. Leaving mainframe computers behind in favour of PCs and with more and more people using networks and the Internet, the old way of working with computers has changed radically. Nowadays, the Internet is much more pervasive; almost everyone who uses a computer becomes a producer of electronic documents as well as paper ones. This represents a huge challenge – it is in effect a new type of science that must be learned, a revolution with new questions and problems to be solved and (as yet) very few experts and little practice.

There is often a feeling of isolation among archivists. I think that not enough young people are entering the profession, and often changes to our ways of working and the acceptance of these new challenges are met with resistance. The development of the field is now being considered at both national and European level, as confirmed by an international meeting several years ago in Brussels, where archivists discussed approaches to electronic archives in different countries.⁵ As a result of the meeting, France set up a national programme to respond to the needs of this new area in archiving. One agency was created: ADAE (Agence pour le Développement de l'Administration électronique: <http://www.adae.gouv.fr>), the official programme for electronic administration; and the ADELE Project (http://www.adae.gouv.fr/rubrique.php3?id_rubrique=3), a three-year programme (2004–2007) that comprises a strategic report and an action report.⁶ The

It is extremely important to have a pool of resources at a national level that can support management of electronic documents down at regional and local levels.



Centre des Archives Contemporaines

basic aim of the programme was that the Archives Nationales would handle and care for dematerialised documents from all the French governmental and local institutions.

The Contemporary Archives deal with central rather than local administration: documents that must be stored for historical value, although, of course, a lot of documents have the potential to become historically significant. The Constance project represents twenty years of experience in this area – an extreme rarity! One problem is that there simply isn't enough 'critical mass' in smaller, local archives to set up a centre for digital archives, therefore it is extremely important to have a pool of resources at a national level which can support management of electronic documents down at regional and local levels. France is very centralised in terms of organisation and administration, which, although it has advantages, is not ideal for the local archives. The symbolism of having a national centre outside Paris is not to be underestimated! It is interesting that the Ministry of Culture recently announced that a new centre will be created in St Denis by 2009 or 2010 which will have some overlap with the work of the Archives Nationales and the Centre des Archives Contemporaines. As yet it has not been confirmed exactly what materials the new centre will take, so the role of this new centre is still unclear.



The Centre des Archives Contemporaines is close to the grounds of Fontainebleau Castle

Part of the problem in France is the sheer scale of the task. In a way, it would be easier for a smaller country, as the quantity of records that need to be preserved is less.

⁵ For more information, see <http://europa.eu.int/ISPO/dlm/dlm96/proceed-en1.pdf>.

⁶ The strategic report can be downloaded in several formats from http://www.adae.gouv.fr/article.php3?id_article=315.



© Daisy Abbott, HATIL, University of Glasgow, 2004

Fountain within the grounds of Fontainebleau Castle

The structure of the French state means that administration can be slow, but it tends to be very well organised. Now correct structures and procedures are in place, and the government is determined to speed up electronic administration. Creation, use and acquisition of e-documents is spreading quickly in local and central agencies; in fact, it has gained a momentum that we can scarcely keep up with! However, we must preserve these documents, as not retaining contemporary records and only concentrating on the historical materials, as some other countries do, is too narrow a view.

I will now describe briefly our selection process of electronic documents. We meet with the material's producers and require them to make the transformation of their data into a 'flat' ASCII format, as this is the best standard for the long term in order to maintain accessibility and ensure that migrations are possible in the future. We don't take on software or an operating system, only the actual data. As 95% of the data we receive are from databases, they must be converted by the producers into ASCII, and metadata provided (technical and descriptive metadata as well as information on the volume and context) which are then checked. After delivery to the Contemporary Archives the con-

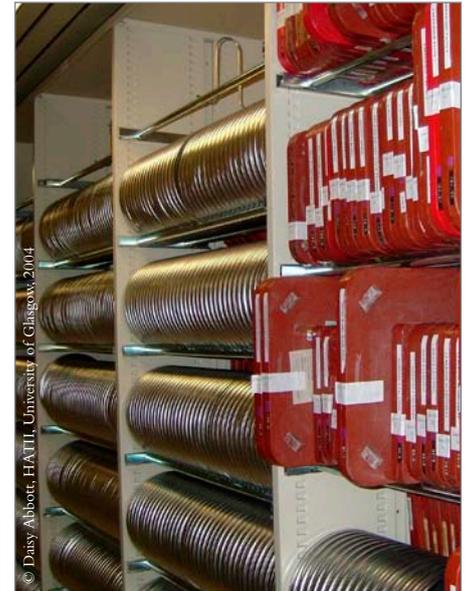
tent is copied and checked for compatibility and to ensure that the metadata matches the actual data. These procedures identify a growing task for archivists: to manage a 'quality chain' from the design of the concept all the way to the acquisition and storage of data. Each record we keep is copied every five years on to a new medium as the magnetic tapes we use to store the data can degrade.



© Daisy Abbott, HATIL, University of Glasgow, 2004

Jean-Pierre Teil holds one of the many magnetic reels stored in the collection.

It is important to remember that our duty of care does not stop when the data are in storage and that stock management must be very rigorous; the integrity of data must be checked at all stages, to maintain the authenticity of the records. As mentioned, we migrate the records on to new media every five years; however format migration is much more tricky. In 1995 we had to migrate 7000 files from EBCDIC format to ASCII and we had to take great care to ensure that steak was not converted to mincemeat during this process! It took eighteen months of manual migration (for example, converting formats such as scientific binary to standardised ASCII). Each record was initially checked to determine whether transformation was necessary. Some datasets were left because the metadata was not sufficient to tell whether a hand-transformation was needed. This example shows how quality control must be very strict. One day perhaps, all our records might be converted to UNICODE – but it is likely that this would be automated.



© Daisy Abbott, HATIL, University of Glasgow, 2004

Shelves storing magnetic reels.

Electronic archives are very different from traditional record keeping in terms of storage. A lot of space is not needed to keep digital data but the data are more complicated to access than simply looking in a box, so instead there is a greater emphasis on good tools and a good strategy! Human resources are the most valuable asset to a good e-archive. Expertise is the most difficult (and expensive) resource to obtain, and an efficient team must be built up over several years in order to best understand good strategies and tools. It is also incredibly important that effective training is provided for the archivists (and archives) of the future. The development of expertise can also be achieved through international meetings and the sharing of knowledge and experience. Projects such as ERPAnet (<http://www.erpanet.org>) are of crucial importance in this regard. Archivists must master the problems presented by electronic records before we can move on to produce effective tools for managing them – it is still too early for automatic tools to be fully utilised here. For example, electronic files can be subdivided into different types, e.g. image, sound and text. Each type requires a different tool and different procedures to ensure that the record is stored properly and can be accessed at a later date.



Jean-Pierre Teil demonstrates the formats used for storage in the Centre des Archives Contemporaines

One subset of e-documents is digital images. It is possible that the current formats (for example JPEG) may become obsolete in the future, giving us a finite timeslot to migrate the data into a new format, before the old one becomes unreadable. The principle is the same for text in, for example, Word format, or for audio formats. These different subsets of records may need to be changed at different times. For these reasons it is important to watch technical trends in the market to identify as soon as possible when migration might be necessary and to plan and budget for this time. There is some flexibility, of course: if you miss one rung of a ladder, you'll still be able to climb it, but if you miss three rungs... it is too late, it will be technically and financially unfeasible to rescue the data from its old format. Digital archives are in more danger than paper archives in this sense; it is very easy to lose them to progress. However, they are often so much more supple and useful than their analogue equivalents – these advantages must be balanced with the constraints.

All French archives are free (naturally, with a small charge for services such as photocopying). It costs a lot of money to provide appropriate assistance to the public on a one-to-one basis. The different modes of access across time and space allowed by networked delivery of e-archives will change usage so considerably over the



Fontainebleau Castle

next twenty years that it will again alter the role of the archivist. The physically smaller e-archive repositories will be accessed in a widely dispersed number of locations and times. The next generation of archivists won't ask for money for new shelves or a bigger reading room – they'll ask for more powerful computers and more technically qualified staff. Archivists must become partners in the new electronic administration, assisting the state in addressing the legal frameworks of the developing electronic society. Our records need to be authentic and be seen to be authentic, and therefore a trusted third party is required to handle, for example, digital signatures ensuring security for the materials. These issues are related to, but not the same as, the traditional record-keeping role of the archivist, changing and developing our roles into the future. Of course, for a professional body to reflect these changes, it may take even longer than another twenty years!

Catherine Dhérent, Chef de la mission pour la gestion de la production documentaire et des archives, Bibliothèque Nationale de France

From her office in the East tower of the famous Bibliothèque Nationale building in Paris, Catherine Dhérent explained her work preserving the administrative records produced by the library.

☛ The Bibliothèque Nationale has 2800 workers, who are dispersed over seven

different sites, three of which are outside Paris. As a public institution, we have a special legal status, making the need for good administrative record keeping even more important. The President of the library was aware of the need for better management of archives than we previously had in place, and wanted to improve our knowledge of what and how documents are used, by putting tracking procedures in place. Until the end of 2003, no one knew who was responsible for keeping each file or document, or even where they were held. Archiving was not consistent, with some departments setting up their own specif-

'It is important to remember that this library is not only a repository for knowledge, but that it also has its own history to manage.'



Catherine Dhérent



ic archive for departmental administration while others relied on the central library to manage the documents. For example, the manuscript department had an archive with holdings from the seventeenth century until the end of the nineteenth. In the twentieth century, a general secretary was set up within the library to centralise records between the end of World War II and 1984, but it was extremely difficult to manage records during the period that the new library was being built (1980–94). There is a lack of information about the administration of the library during this time, records were not well managed and, as a result, much of our history has been lost.

In 2002, the President decided to create a new mission for the library's documents and archives, run by a specialist archivist, and I joined the BnF at the beginning of 2004. My background is twenty-five years of experience in the Archives of France for the Ministry of Culture, which provided the specialist knowledge that the President required to run this new venture. The two major goals were to put in place procedures for dealing with administrative documents (e.g. mail registration) and to track documents to find out who was using them and where they are stored. My job is to manage the almost 6 km of administrative archives which have been collected since the seventeenth century. Our aims are linked to one of the higher-level aims of the BnF: to manage and record our own personal and financial affairs, which is currently not at a high level within the structure of the library.

Although our mission is very modest (only three people!), we have already made some progress in tracking documents and discovering who their readers are. For example, we have rich resources on the tools used by the library – for instance, how the use of typewriters was introduced to the BnF. This type of information is very useful to historians or universities who are interested in the history of an institution.

The histories hidden in archives like these can be fascinating; here, the period of the Second World War is of particular interest as one of the library's administrators, a famous librarian, Julien Cain, was captured and deported to German camps in 1940 and took over the general administration of the library again after the war, instead of another administrator who was linked with the Vichy government. New perspectives on history, within both an institutional and a national context, can be discovered through these archives, and in some cases these records could be the only evidence of historical facts.

There is, therefore, a necessity both to show and explain these documents, and

One difficulty that was immediately apparent is that the administrative documents are widely dispersed. It is perhaps not the best solution to centralise documents as, although this makes standardisation of procedures easier, it separates items from their creators or producers, and in our case some individual departments have been managing their own archives for decades. It is more important simply to know *where* each file or document is and to manage them wherever they are physically located. Each item can be physically located (be it in somebody's office or stored in a stack) by its unique code. This approach creates an easier relationship between the library's departments and the archivist – it is often not easy being an archivist in a



Wooded area within the four towers of the Bibliothèque Nationale de France

to preserve contemporary documents for the future. Managing the archives at the BnF has two parts. The procedures of handling contemporary documents have a lot of similarity with records management, whereas the historical documents require a more traditional archiving approach. My first task was to prepare a report defining what was necessary in terms of staff, budget and new tools in order to achieve the President's aims of allowing the team to proceed in planning and implementing our archiving strategy.

world of librarians! We have to ensure that our work is complementary and that staff members know that the purpose of good archiving is not to lock the documents away. Indeed, some items have a very long period of usefulness where they are often consulted and are required to be near the people who need them. As, in our specially designed system, each office also has a unique code, the management of where items are stored becomes much easier. We are still in the midst of setting up our



retention schedule, however, so there is still a long way to go.

To implement the changes as efficiently as possible, I meet each week with three or four people who create documents and analyse both what they are producing and how their methods may vary. Then, step-by-step, the retention schedule is added to. This task will take a long time to complete – there are as many as 1500 staff at the BnF who are document creators and it takes one year to properly analyse around 150 creators, as each requires a report, a change to the retention schedule and work on finding aids.

At the same time, my team is working on the procedures of communication of documents for both BnF staff and the public. It is pleasing that, in the first three months, we do have a demand for material and that we can be accessible enough to help users of the archives through our new procedural channels.

The tools we use are very helpful in dealing with such a large institution and number of items. For instance, as well as using comparative tools, a great deal of work can be done using our intranet – for example, analysing and tracking the phases of treatment for each document. We are currently working on information services and hope to have implemented an automatic checking system for acquisition by the end of May, and we are also going to be perfecting automatic caching for inclusion of documents in 2005.

Of course, archiving has changed a lot in recent years with the advent of electronic records. E-records create new challenges – they require different management styles and techniques, and I hope that records management continues to evolve to reflect this. One important concept is the distinction between records as documents of *proof* and records as *useful* documents. Many e-documents are extremely useful

and often consulted, but in many institutions are not recorded at all. Here in the BnF, we are able to record e-mails, but it is a continuing challenge to make the best use of the technology we have available. For example, there is often confusion among users about the difference between an electronic document being stored on their local hard disk or on the intranet. It will involve a lot of work to create appropriate procedures and tools for adding metadata to these documents; people have vastly different methods of working, which are totally unstandardised, so storing these documents for maximum quality in the long term is a very difficult task. We will use the existing system (LotusNotes) to implement the new teleprocedures; however, it is important to remember that the emphasis should not be on the tools but on enforcing the procedures and rules themselves, which are validated by the aims of the institution. We are working together with staff from the Department of Preservation to best design the metadata for e-records – we must have enough information to be able to use the document again in the future, but not so much that it is unrealistic to expect people to fill it in! Of course, in the future we will be working on ways to make metadata collection automatic and ways of allowing querying over our intranet (with rights access), with the useful part of the document being stored online.

There has been a big change in the concept of archives and archiving systems and theories, the traditional standpoints almost destroyed by the characteristics of electronic records. E-documents blur the boundaries between the traditional ‘ages’ of paper records: historical, intermediary and current. It is incorrect to apply these stages so simply to e-documents; they have a longer current age and we are losing the intermediary stage. Instead, the current and historical stages overlap, making the three-stage theory no longer relevant and further identifying the need for a different approach. Research nowadays often con-

centrates on online materials; with many researchers not checking the paper record at all, use is changing and so must archiving standards.

There has been a big change in the concept of archives and archiving systems and theories, the traditional standpoints almost destroyed by the characteristics of electronic records.

This field is changing very quickly and will continue to do so into the future. One day, if metadata is created at the point of the document’s creation and is embedded, documents may even be able to live without us! All of the current archivist’s duties will be incorporated into a persistently identified record. Archivists could perfect these systems, so our work will necessarily evolve – to concentrate more on providing help for users, for example, or quality control, ensuring that the laws regarding rights are respected. We will, of course, have to continue to manage the documents still on paper; all issues cannot be resolved by automatic systems! There are always new records being discovered from abundant sources and on a large scale, and our work in this area will continue.

At the moment, we are just at the beginning in the BnF, following the example of other European institutions that have already begun this process and provide examples of good practice. It is a challenge to reach the standards of some other countries’ best institutions, as Latin countries have traditionally not had the same management practices. But it is one of the major aims of the President of the BnF to ‘make up for lost time’ and we’re making good progress – I hope that in the future people will look to our work here as another example of good practice!

**Catherine Lupovici, Directeur,
Département de la Bibliothèque
Numérique, Bibliothèque
Nationale de France**



Catherine Lupovici

The Director of the Digital Library department at the Bibliothèque Nationale de France, Catherine Lupovici, describes digitisation at the library over the relatively short period of the last sixteen years. Having previously worked in a private digitisation company, Lupovici's significant technical knowledge informs her discussion of the work of her department within the BnF.

‘The BnF began to consider electronic resources at the same time as constructing the new building in 1988. The aim was to create an encyclopaedic version of the documents held by the library. Progress towards creating these digital resources began in 1989 and the digitisation of our collections started in 1992. We began with printed documents and the cultural heritage of the library: our unique material and special collections. After 10 years, we had digitised over 30 million printed pages.’

‘At the moment, over 20 million of these are in the public domain and many are available online from GALLICA (<http://gallica.bnf.fr/>), the online section of the digital library of the BnF. The remaining pages have current copyright restrictions and are instead accessible via our intranet,

with the rights to view and print and so forth carefully managed’, Lupovici explains. Documents for the digital library are primarily processed in image mode for technical reasons and are selected for digitisation based on both their rarity and their copyright situation. In addition to textual materials, la Bibliothèque Numérique holds 100,000 digitised images drawn from collections outside the BnF and from patrimonial collections among its own departments.

‘In addition to printed materials and images, the audiovisual department in the library is also digitising sound. During planning for the new building, we bought 10,000 audio CDs that were made available to visitors through robots – after searching the catalogue and selecting a record, the robot would then play it for the user. We also make moving image material available; for example, analogue VHS tapes are digitised on the fly and stored locally before being delivered to users, but we don't store the digital copy long term.’ GALLICA also now offers access to multimedia documents, and, as its Director explains, the Département de la Bibliothèque Numérique is now beginning to look towards longer-term solutions: ‘Of course, our systems were designed ten years ago, so now the audiovisual department is changing its strategy and beginning to digitise for the long term. We now understand that digitisation is critical for conservation of these materials, as well as facilitating access. The 80,000 titles represent 400 terabytes of storage – and some of the originals are under threat, so quality digitisation is imperative.’

‘GALLICA was designed to deliver materials online; however, the masters were also carefully preserved, most being stored on glass disks. Although we used to photograph materials in both analogue and digital formats, the reprographic world has changed, and digital photography is developing so quickly that we now use the initial high-quality digital reproduction as the master rather than the analogue pho-



Initial work on the digitisation of resources began at the same time as the building of the new Bibliothèque Nationale

tograph for our colour materials. Slowly, the analogue cameras are being phased out – we currently have mixed cameras that take both an analogue and a digital photograph at the same time, but this equipment is expensive and so is only being replaced in some departments such as Special Collections, while other departments are moving to digital only. Of course, we store reproductions in different formats for different purposes so that they can be more easily supplied on demand for our users.’

‘It is important to remember that we are not acting as a publisher but as a repository. Therefore we are not aiming to deliver finished digital products, but to retain the same methods and mission as we did when recording materials in analogue format. Tools do exist to, for example, digitally ‘clean up’ old films or to remove noise in audio tracks, but the BnF's purpose is to offer our users a trusted source for materials, not to attempt to restore them.’

‘We don't have any choice but to digitise now, or important materials will be lost.’

When asked about the progress of the work of the Département de la Bibliothèque Numérique across the library, Lupovici pauses to consider. ‘Digitisation within an institution such as this is much more complex than merely the technical aspects. Some departments simply aren't ready for digitisation – in terms of accepting that it has become a necessity for the preservation of their materials. XX

This isn't a private digitisation company, it is a library, so it can be difficult for staff to fully endorse the importance of what we do. It is such a fast-moving field, but I think that time will help more and more people to understand the necessity of the digital library. Of course, the technology is changing all the time, but we can't wait for the "perfect" digitisation technology – we have to act now, or important materials will be lost. There isn't really any choice.'

¶ The real challenges for the future will be in managing the rights of digital objects. The situation in France is relatively restricted as regards access rights. For example, unlike most European countries, there is no waiving of rights restrictions for research or teaching purposes. We are working on adapting the European directive to rights management within France. There are, unfortunately, no ways in which to bypass the technical demands and issues relating to access rights at the moment. Digitisation for preservation and digitisation for access have different requirements and complications – it is our job to address these differences as quickly as possible in order to best serve the aims of the BnF and the needs of its users.'

Daniel Teruggi, *Directeur recherche et experimentation, Institut national de l'audiovisuel (INA)*

INA (<http://www.ina.fr>) was set up in 1975 to preserve French national audiovisual heritage, make it more widely accessible, and keep abreast of changes in the audiovisual sector. All public broadcast programmes have been archived since INA's inception, resulting in over 1.5 million hours of audiovisual materials. Since 1992 INA has been in charge of Radio and Television Legal Deposit, keeping a recording of all broadcast programmes for research purposes.

Daniel Teruggi started working in the musical research group (GRM), founded in 1948, which was incorporated into INA from the outset, and became Director of Research and Experimentation in 2001.



Daniel Teruggi

¶ The Research and Experimentation Department at INA contains different focus groups for research. GRM (*recherches musicales*: <http://www.ina.fr/grm/>) comprises around fifteen people and works on building tools for sound and the production of programmes and CDs, while the

GRAMM, audiovisual and multimedia research group (<http://www.ina.fr/recherche/index.fr.html>), employs about thirty staff and ten PhD students working on problems related to archives, mainly image and sound restoration, content identification and content description.

We now provide online access to around 200,000 hours of material that is browsable by registered customers. Before a high-resolution version of a programme or clip is delivered, the access rights and permissions are cleared, but browsing versions are available. Our selection criteria for placing materials online can be considered in three main areas:

1. Technical considerations: for example, is the material under threat?
2. Content: for example, is the material very likely to be used? Is it a unique or particularly excellent example?
3. Commercial factors: for example, we digitise 'on demand' for our customers.

Although at the moment INA offers these materials for professional use rather than to the public, in the future we want to open up public access to the digitised materials.



INA offices at Radio House

Our broadcast quality programmes are rented out, with an appropriate fee being paid for each commercial use. Negotiating rights for audiovisual materials is complicated at the best of times, and, as copies are delivered to the users, it is therefore possible for resources to be used beyond the licence agreement. Since legal deposit began, we have developed a new project to track and monitor usage of our materials, ensuring that the owner of, say, a news clipping, is known, and to help us identify unauthorised and/or illegal usage. Our news archives are the most used materials, and this is one example of an anticipatory project: high-level research for fast identification of images, to survey what is being broadcast and identify anything which belongs to us.

Our research methodology combines an academic approach to research with a developmental model. We have around ten PhD students who are semi-funded by the government and (currently) 25 senior researchers who work both on projects and initiatives specific to INA and on generic solutions to modern problems in the audiovisual sector. This combination of academic and industrial methodologies is very useful as it allows us to view INA from different perspectives. Without this distance, it would be easy to fall into the trap of becoming simply a software developer.

The value of your archive depends on the knowledge you have of it.

Ten years ago, INA began a new mission: as well as archiving all public broadcast programmes, we would also become the legal deposit repository for all broadcast programmes within France, like a public library but for film and television. INA now stores all of these programmes and makes them available for research. Legal deposit materials are stored in compressed format, as opposed to the public



Audio record showing signs of decay. Shellac disks suffer from a slow chemical deterioration whilst the 'vinegar syndrome' of acetate media is well known in the archival world.

broadcasts, which are retained at a broadcast quality. Producers send their materials to us in a digital format where it is compressed, checked, documented and several copies are made (for storage, access and backup). The traditional challenges of conserving and storing media still apply, but we are also faced with many more modern problems: digital storage, description, documentation and compression. The whole span of archival activities is reflected in the work of the research department, from the chemistry of magnetic tapes to tools for semantic interpretation. One major problem for audiovisual archives is that of long-term preservation, as digitising every item is a very ambitious goal.

Our audio materials are generally stored in shellac, magnetic tape and DAT tape formats, and we also have 16 mm film and video (of which there are six different magnetic formats). There is a crucial problem in that it is not only the storage formats themselves that are at risk, but the

machines that can read them are also dying out. If these audiovisual materials can be compared to books, we must ensure that we also keep the 'eyes' that can read them. To identify items that are under threat is a difficult and time-consuming task, as you have to look through the entire collection! It is a fight against time to check each disk for damage, remove the 'sick' disks (as the vinegar syndrome suffered by acetate formats is contagious and will spread to other materials unless it is quarantined) and take steps to preserve the threatened data. Once an item is identified as under threat, we store it separately in a cold area and digitise it as soon as possible. It is very important that we digitise as much of our material as possible before the analogue original becomes useless.

There are two main factors that affect our ability to digitise: money and research. Digitisation on INA's scale requires a very large investment, and the investment must be over the long term; at

'It is useless to wait for the 'perfect' storage format – we will never have it!'

least the next ten years. As with many other institutions, political changes can affect the amount of investment that we can rely upon as the priorities of funders can change. In terms of research, we have launched a project this year to develop technology specifically to accelerate the digitisation process and make it cheaper. This work has a tremendous value for institutions at all levels within a country; for example, an organisation like the BBC can get money to digitise its materials, but smaller institutions have less reliable funds, even though some of their materials are equally important. Who will protect these collections? Ideally, our research will provide a solution to the problem of the great expense of undertaking digitisation activities.

These steps are INA's approach to potential obsolescence of materials in the future. In my opinion it is useless for us to wait for the 'perfect container' for audiovisual materials, as we will never have a perfect storage format. In fact, the best one is the first one: film! If we waited for the perfect medium, data would be lost; instead it is evident that action must be taken now to prevent these cultural materials from becoming inaccessible. The only solution is to accept the fact that any medium is not eternal (although you should still choose the best format available) and to *build a good migration plan*. It is imperative to realise that the technologies will also change. This may well be the last opportunity that we have to digitise this particular item – so it must be performed at the best feasible quality with no modifications. One must also understand the value of the material and its uses in the future to justify decisions on the quality of digitisation. For example, the data needed by digital video are 50 Mb per second, whereas our master copies are only 8 Mb/s. Experts agree that at least 200 Mb/s is needed to replicate the quality of cine



Main entrance to Maison de Radio France, Paris

film – although 50 Mb/s is still an excellent digital quality. Balancing the quality requirements of the materials with the time and expense of creating digital versions is a difficult area. Once an unmodified master has been produced at the highest possible quality, then temporary versions can be created from it, for example, applying digital clean-up algorithms on an old television programme. Yet another factor to consider is that users' demands will change as the years go by – what is considered as a natural requirement now may not always be the case. For example, we supplied some moving image material to users after performing image restoration, but there were complaints about the quality of the sound attached to the images. Conversely, users were generally much happier if the sound had been 'cleaned' but the images were unmodified. This unexpected result shows that user needs are not always what we now consider them to be!

Our most recent major programme began around two years ago – our aim is to collect materials based around a specific theme, for instance, one person. We then attempt to digitise a comprehensive

(or at least representative) selection of these programmes to provide a detailed and varied view of the subject in various formats and styles. INA customers tend to be very demanding, each requiring material that is not only high quality but also original, i.e. has not been used already by a competitor. We hope that, by focusing on specific themes, we can provide material that is suitable for the purposes of a large number of our customers and will increase the choice they have.

As a participant of the information society, I have many tools to support my information needs. I can access millions of examples of written material in the modern 'global library'; however, there is comparatively little free access to audiovisual material. Digitisation will provide this access in the future and make film, television and radio programmes as ubiquitously available as the written word. My aim is to encourage work and research on archival materials and to catalogue the future of this sector. INA is involved in many forward-looking projects, from tracking user issues and the ways with which user interfaces are interacted, to the most important issue for

archive access in the future – meaningful querying and content-based retrieval.⁷ We hope that our current and future work will extend some of the concerns and solutions specific to INA across many different types of archive, and improve access to all holdings for our users.⁸

Abdelaziz Abid, Division de la Société de l'information, UNESCO

From an office with an elegant view of the Eiffel Tower, Aziz Abid spoke to DigiCULT about his work with UNESCO and the Memory of the World Project (<http://www.unesco.org/webworld/mdm>).

☞ To introduce myself briefly, before working at UNESCO, I was in charge of the Library of Alexandria, where my work included fundraising for an international architecture competition. I felt that the time was right to move on to something different and noticed that UNESCO (<http://www.unesco.org>) was seeking an officer for a specific project in Egypt. I could not go myself, but this led my work in a new direction, and in 1991 I began to work on the development of the Memory of the World programme, a different project with UNESCO.

At the general conference of UNESCO, one of the major elements was a discussion stemming from several countries from Eastern Europe about the preservation of their heritage, the emphasis being on the fact that co-operation would be needed throughout Europe to prevent the loss of some of this heritage. Since the collapse of the Berlin Wall, there has been very little co-operation across the full breadth of Europe, and the countries asked UNESCO to put together a scheme to assist in the preservation and digitisation of this rich but highly dispersed heritage.

As a result of this meeting, I wrote a one-page proposal, outlining the basics of the new programme, eventually named Memory of the World. The project was launched in 1992, funded by UNESCO with an initially small budget. In these early days, a paper was commissioned to define the needs of the sector and it concluded that a programme was neces-



Aziz Abid demonstrates a publicity board for the Memory of the World Programme

sary to co-ordinate activities and that digital technologies should be used to provide democratised access. In order to supervise and advise the programme, a body of fourteen people was created. The members are from all over the world and represent many different disciplines within culture, heritage and technology, for example, publishing, information science, legal, and history.

The first meeting of the Memory of the World Programme took place in September 1993 in Pultusk, Poland, a country identified as having suffered a great deal of damage to its heritage as a result of war. This meeting represented the birth of the project, laying the foundations for the future and setting up three major objectives.

The first was to act for the preservation of world significance and under the threat of disappearance. The most relevant materials needed to be identified and therefore we designed and planned surveys to meet these needs.

Secondly, it must be remembered that Memory of the World emphasises preservation not primarily for conservation but for access purposes. Part of this ultimate goal was to encourage existing holders of culturally valuable heritage to open up their collections to the public, despite a certain reluctance from some! At that time, the benefits of digitisation were not as obvious as they are today. Clearly, the original materials must be protected, not a process that is naturally linked to access; however, early suggestions of using digitisation technologies were met with some scepticism and resistance. For example, I remember a quite heated discussion about the relative benefits of microfiche as opposed to digitisation!

We began with a few pilot projects (for example, the Sana'a manuscripts at <http://www.unesco.org/webworld/mdm/visite/sanaa/en/present1.html> and the work at the National Library in Prague: <http://www.unesco.org/webworld/mdm/visite/prague/en/present1.html>, which I believe is one of the most successful projects based on relatively little initial funding). The programme grew, extending to many institutions all over the world and utilising the alliances forged by these pilot projects. More recently, our approach has moved more towards the systematic digitisation of collections. This is a very clever approach – scholars are offered a digital copy of an item upon demand as an alternative to viewing the original and have the option of buying a product containing the digital copy, such as a CD-ROM. This not only helps to preserve the original and offers researchers added functionality, but the digital resource is itself enriched by further papers, knowledge and expertise contributed by the knowledgeable users (such a system is used at the National Library of the Czech Republic: <http://www.nkp.cz/>).

⁷ Content-based Information Retrieval will be discussed in a chapter of the forthcoming DigiCULT *Technology Watch Report 3*.

⁸ One example of a project that developed from the work of INA is PrestoSpace. For more detail, see the article dedicated to PrestoSpace, also in this issue.



© UNESCO 2004



Prague Manuscripts

A digitisation programme has been launched by the National Library in Prague in co-operation with a private firm, Albertina Ltd. and resulted in the publication of a number of CD-ROMs entitled *Memoriae Mundi Series Bohemica*. Digitising the most beautiful manuscripts and old prints of the National Library has facilitated access to these treasures without exposing the originals to heavy use, thus contributing to their preservation.

Catalogue of Arabic Manuscripts. Al-Qur'an.
1211 A.H. (1796 C.H.)
Memoriae Mundi Series Bohemica
Photo: National Library of the Czech Republic



Sana'a manuscripts. Qur'anic fragments.
6th century
Photo: Great Mosque of Sana'a, Yemen

The Sana'a Manuscripts

In 1972, after heavy rain, a section of the wall of the Great Mosque of Sana'a collapsed. The work on the roof brought to light manuscripts which had been concealed in the ceiling for centuries. The parchment and paper fragments contain mostly extracts from the Qur'an, and they are of considerable interest for the linguistic, religious and palaeographic study of the literature of the early centuries of the Hegira.

Extracts from the UNESCO publication „Safeguarding the documentary heritage of humanity“. Used with permission.

The third objective formulated in the initial meeting was to promote awareness of the materials and the programme through both events and products. This resulted directly in the creation of the Memory of the World Register (<http://www.unesco.org/webworld/mdm/register/>), a list of documentary heritage that has been identified by the International Advisory Committee as being of world significance. Another UNESCO resource, the World Heritage List (<http://whc.unesco.org>), is an international list that brings together man-made cultural monuments and sites with naturally occurring heritage sites. Countries nominate individual sites and the World Heritage Committee votes to select those to be included on the list and ensures that there is balanced geographical representation. There are currently 754 properties from 129 countries inscribed on the World Heritage List (582 cultural, 149 natural and 23 mixed). The Memory of the World Register is like a young sister of the World Heritage list, focusing on documentary heritage of all kinds (e.g. film, sound, manuscripts and archives) rather than monuments. It was created after our 1997 meeting in Tashkent, Uzbekistan, and is therefore relatively recent in terms of the whole project. There are currently 91 items on the register from 45 different countries, including some real masterpieces such as Fritz Lang's film *Metropolis* (which we are showing to heighten awareness of the programme) and the Gutenberg Bible.



A poster for the film *Metropolis* outside the Bibliothèque nationale de France.

The idea of the Memory of the World Register, as well as our other activities, is to draw attention to our work, not (like the World Heritage List) to provide a financial or legal commitment to the projects and resources featured. UNESCO does fund some activities in poorer countries but is under no obligation, by including a property on the register, to provide support for it.

After ten years and six project meetings, the Memory of the World programme celebrated its anniversary by returning to Poland to meet at Gdansk in August 2003 at the invitation of Lech Walesa, former President of the Republic of Poland and Nobel Prize winner.⁹

National committees have now been formed in 59 countries across the world to identify digitisation priorities and address issues at a more local level. Committees at a regional level also consider the best way to preserve and provide access to multinational documentary heritage. These regional committees have their own Web sites and meetings.

The Memory of the World Programme also creates and distributes products derived from the programme such as CD-ROMs (of which we now have more than twenty) and technical documents, guiding future work in preservation or access. These documents are published in multiple languages and are free. We are also beginning training programmes all over the world to share experiences among different people and places.

Memory of the World is an example of a wide-reaching international project that uses a hierarchical organisation to encourage access to the most valuable documentary assets of the world. Although relatively recent in terms of UNESCO, the sea change in attitudes towards digitisation have been observed first-hand through Memory of the World's activities, and the incorporation of expert opinions emphasises its theme of sharing knowledge of these artefacts with as many people as possible.⁴

More information about this project and UNESCO can be found from the Web site at <http://www.unesco.org>.

⁹ Walesa was revolutionary in demanding workers' rights, leading the movement from a Gdansk shipyard where he worked as an electrician. Becoming chairman of the Solidarity labour union, he was involved in the formation of elections and the establishment of a non-communist government. More about Walesa's Twenty-One Demands can be found by searching Memory of the World at <http://www.unesco.org/webworld/mdm>.

MAKING HANDWRITTEN ARCHIVES DOCUMENTS ACCESSIBLE TO THE PUBLIC WITH A DOCUMENT IMAGE ANALYSIS SYSTEM

BERTRAND COÛASNON (IRISA/INRIA),
JEAN CAMILLERAPP (IRISA/INSA),
IVAN LEPLUMEY (IRISA/INSA)

EUROPEAN CONFERENCE ON ENCODING RESEARCH AIDS

Following the seminar on the application of the EAD DTD for cultural heritage purposes in archives and libraries, held two years ago in the National Library of France, the Directorate of the Archives of France is organising a European Conference on 7 and 8 October 2004 on data-processing software for encoding research aids for archives and authority data into XML. The last European Conference on the subject took place at the Public Record Office in Kew in October 1999. EAD, whose first public version (1998) was revised in September 2002, is a document type definition founded on the same principle of multilevel description as the General International Standard Archival Description, ISAD(G). EAD contains elements that are equivalent to MARC fields (an EAD inventory aid may be linked with a MARC note and, conversely, an EAD notice may be converted into MARC).

EAD is now well established in North America and in the rest of the world, and it has attained technical stability while continuing to keep abreast of technological advances. Training courses have been organised and handbooks and application guidelines have been produced. Several institutions using the EAD in France, the UK, Germany, The Netherlands and Spain have developed innovative solutions for encoding, editing and displaying their searching tools. It is now time to assess progress and to facilitate the exchange of experience. Since 2001 another DTD,

EAC (Encoded Archival Context), has been under development. Perfectly compatible with the International Standard Archival Authority Record for Corporate Bodies, Persons, and Families – ISAAR (CPF) – and complementary to the UNIMARC/Authorities format, the EAC combines bibliographic authority records and archival authority records, which give information about both the creator and the context of creation of archival material.

Version 1.0 of the EAC Tag Library should shortly become available. Tests have been carried out on the conversion of authority data into EAC/XML format, in particular within the LEAF project (Linking and Exploring Authority Files): local authority data will be downloaded from the local servers of the participating organisations to a central system which links automatically the authorities belonging to the same entity. The participants will present their research results and ideas on different issues:

- what is required for EAD implementation: thoughts on standardisation, training;
- EAD implementation: tools for creating EAD documents, interoperability, compatibility with other standards;
- publishing EAD/XML documents: editing tools, documentary management applications with documents stored on a Web server in native XML;
- the EAC DTD and the first examples of authority records encoded into XML.

INTRODUCTION

French archives, like other archives services around the world, own millions of pages of documents containing handwritten information which are difficult for members of the public to access. Even if archivists always built indexes and research tools, the quantities are so huge (the basic counting unit is a kilometre of shelving) that there are still a lot of documents which are both difficult and time-consuming to find, both from the shelves and within the document itself, if the reader is looking for a specific page. At the same time, a growing number of people, such as genealogists, are interested in these documents. How can archives offer public access to millions of pages of documents with handwritten information, when no research tools exist? Archives started to scan documents; however, even with digitised versions it is still often necessary to leaf (virtually) through a considerable number of images. Even if the page required has been digitised, the time needed to find it can still be enough to discourage the user to the point of giving up.

To solve this problem, systems must be defined to allow document retrieval by content. For this to be achieved it is necessary to associate annotations to the images of documents. Annotations for archival documents can be geometric (a position in the image represented by a rectangle or a polygon) or textual (e.g. a date, a place, a name, or a keyword). Of course, textual annotations can be linked to geometric ones, making it possible to select the appropriate images automatically.

We propose two complementary ways of producing these annotations: automatically with document image analysis and collectively, on the Internet, with the help of the readers during their use of the image. We present a platform developed to manage collective annotations built on automatic annotations. We show application examples on various documents: civil status registers, military forms of the nineteenth century and naturalisation decrees. For each document we present the automatic annotations that we are able to produce with DMOS (Description and MODification of Segmentation), a generic document recognition method we developed.¹² We also present the collective annotations that can be added by users with the help of the automatic annotation. This platform offers a uniform interface for accessing archival documents by content.

PRODUCING ANNOTATIONS

Automatic Annotations

Automatic annotations are produced by optical document recognition (OCR). On printed and recent documents, existing OCR systems are able to recognise almost all of the text and can then be used to build textual annotations for document retrieval. On archived and old documents, it is more difficult because documents can be damaged (tears, blots, tape repairs, smudges), ink on the reverse side of the paper may bleed through to the front of the page, stamps can be affixed and many other features also create problems for OCR. To automatically produce annotations on poor quality documents with handwritten text, with many different writers, on large vocabulary without dictionaries, it is necessary to first locate *where* the information for retrieval is within the image. This allows us to detect what part of the image contains handwriting and to begin work on handwriting recognition.

Geometric Annotations

To be able to detect the position of specific handwritten text, it is important that the document is structured, for example documents like forms or tables. We can work on less structured documents if handwritten text is graphically structured, e.g. within margins. Literature shows that structured document recognition systems are difficult to develop and systems are usually specific to one kind of document. A new kind of document often means a complete development of a new recognition system. This is why we proposed DMOS, a generic recognition method for structured documents. This method involves the new grammatical Enhanced Position formalism (EPF), which can be seen as a description language for structured documents. EPF makes possible graphical, syntactical or even semantic description of a document. DMOS contains the associated parser which is able to change the parsed structure during the parsing. This allows the system to try other segmentations with the help of context to improve recognition; it is an automatic generator of structured document recognition systems. Adaptation to a new kind of document is achieved by simply defining a description of the document with EPF grammar. This grammar is then compiled to produce a new structured document recognition system. With this generator, we have been able to produce various document recognition systems: for musical scores, mathematical formulae, and recursive table structures. We could even use it to recognise tennis courts in videos!

Later we present the application of DMOS to various documents to automatically produce geometric annotations on the document structure.

Annotations on Handwritten Text

Once handwritten text is located within the document with DMOS, it is possible to analyse the handwriting. As we are generally interested in last names, it is impos-

sible to use dictionaries, as they cannot be exhaustive. Moreover, these names are written by many different writers. It is then impossible to use handwriting recognition methods on large vocabularies,¹³ or even word spotting.¹⁴ We propose to extract (by image analysis) a grapheme representation of handwritten names (see Figure 1). This representation is stored as textual annotations.



Figure 1: Automatic extraction of grapheme. Dots represent the beginning of each grapheme.

This produces automatic indexing of handwritten names. A user will be able to make a textual request which is translated into a grapheme representation. Using edition distance, this representation is compared with all indexes to select the names according to the request. Later we present the first results of automatic access of handwritten names in military forms from the nineteenth century.

¹² See also: B. Coïasnon, 'DMOS: A generic document recognition method, application to an automatic generator of musical scores, mathematical formulae and table structures recognition systems' in *ICDAR, International Conference on Document Analysis and Recognition*, pages 215–220, Seattle, USA, September 2001, and B. Coïasnon, 'Dealing with noise in DMOS, a generic method for structured document recognition' in *LAPR International Workshop on Graphics Recognition*, Barcelona, Spain, July 2003.

¹³ Such as is described in A. Vinciarelli, S. Bengio and H. Bunke, 'Offline recognition of large vocabulary cursive handwritten text' in *Proceedings of the 7th International Conference on Document Analysis and Recognition*, volume 1, pages 1101–1105, Edinburgh, Scotland, August 2003.

¹⁴ See R. Manmatha and W. B. Croft, 'Word spotting: Indexing handwritten archives' in M. Maybury, editor, *Intelligent Multi-media Information Retrieval Collection*. AAAI/MIT Press, 1997, and T. M. Rath and R. Manmatha, 'Word image matching using dynamic time warping' in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, volume 2, pages 521–527, Madison, USA, June 2003.



Collective Annotations

Some handwritten archives documents can be too difficult to recognise fully with automatic methods. Therefore we propose to supplement automatic annotations with manual annotations which are made by readers during their reading. All the annotations are then put together, making them available for other readers to improve their access by content to documents. As the number of readers is significant, the number of annotations can grow very fast if a tool to manage them exists and if the process is initiated with automatic annotations.

A PLATFORM FOR IMAGE DOCUMENT ANNOTATIONS

We propose to build a platform for archive document retrieval that could deal with textual and geometric annotations at the same level. We will develop the platform with a classical architecture: a Web server (Apache) with a servlet container (TomCat); the Java servlet access to a relational database (PostGreSQL) to store annotations and send them to the client: a Java applet running in a Web browser (see Figure 2).

Using this platform through a Web browser, a user can leaf through images of documents, with all associated annotations presented on the interface. As shown in Figure 2, geometric annotations are drawn on the image; textual annotations are presented in tabs for the structural nodes and in field boxes (e.g. name or date). The reader can consult annotations, add or modify an annotation (if he has the right) but is limited by the allowed annotation structure set by the configuration file, according to the kind of document. The system can also store various interpretations if readers do not agree. Structured search or full text search is possible on all the annotations whether they have been produced automatically or manually.

A pen-based interface has been defined on this platform. Using specific gestures and online handwriting recognition, it has been possible to design a new way to interact with digital documents and paper documents (see Figure 3).



© IMADOC, 2004

Figure 3: A pen-based interface on the platform for annotations

EXAMPLES OF ARCHIVES DOCUMENTS

REGISTER OF BIRTHS, MARRIAGES AND DEATHS

Automatic Annotations

Documents like these are difficult to annotate automatically owing to the weak structure and the poor quality of the handwritten text. The documents are scanned in double pages and we defined a grammar in EPF describing the page. With the DMOS method we have been able to produce a recognition system that detects the position of each page and produces automatic annotations. A test was performed on over 4000 double pages: 99.3% were correctly detected. With these page annotations, a reader can leaf through a register page by page with an automatic zoom function for ease of use.

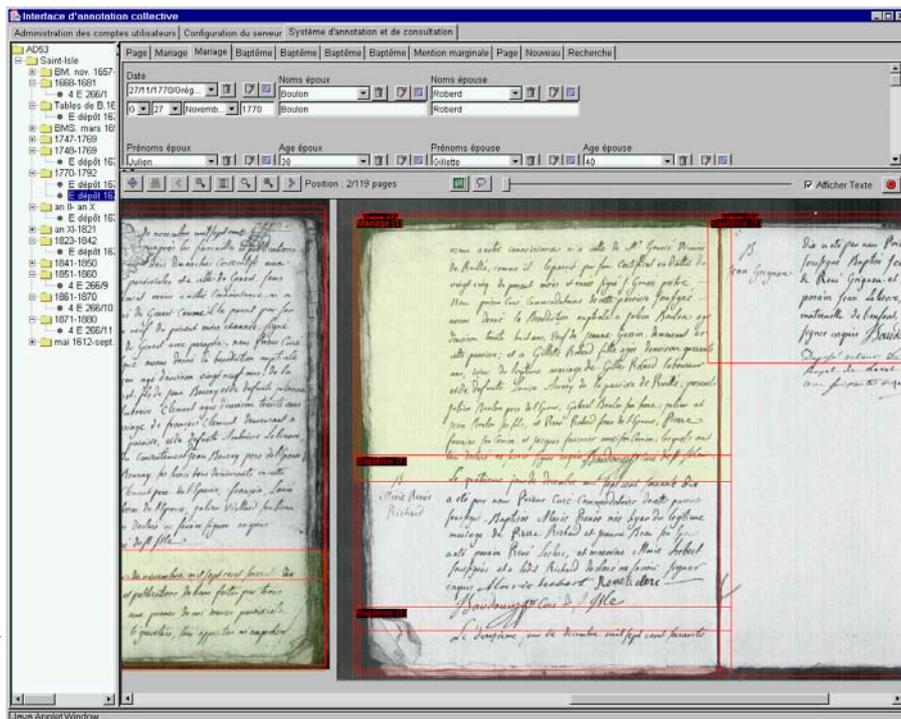
Collective Annotations

Collective annotations can be added, for example, the type of certificate (birth, marriage, death), and details such as the name and surname of the child, place of birth, and the name of the mother. The position of the certificate can be defined by the reader or the automatic annotation of page. There is no obligation for a reader to fill in all the fields.

Applications

A demonstration of the platform on civil status registers is available online at <http://imadoc-ar.irisa.fr/index.html>. It shows materials from two districts with automatic annotations on the position of pages and

Figure 2: One interface to consult, annotate, retrieve Archives documents.



© IMADOC, 2004



the collective annotations on 4000 images (8000 pages). This demo is also available in the reading rooms of the *Archives départementales de la Mayenne* (<http://www.cg53.fr/Fr/Archives/>) and the *Archives départementales d'Ille-et-Vilaine* (http://www.culture.gouv.fr/culture/nllefce/fr/rep_ress/ad_35700.htm).

REGISTER OF MILITARY FORMS

Automatic Geometric Annotations

We also worked on damaged military enrolment forms from the nineteenth century. Various problems make them difficult to recognise: the size of the cells changes from year to year, there are a lot of pasted sheets of paper which hide the form structure, stamps are stuck on, and ink has bled through the paper. Therefore we defined a specific EPF grammar which takes into account these difficulties. From this specific (and small) EPF grammar, we automatically produced a new recognition system to detect the form structure (see Figure 4). We successfully tested this recognition system on 88,725 forms from the *Archives départementales de la Mayenne* and the *Archives départementales des Yvelines* (<http://www.cg78.fr/archives/>): 98.83% (87,692) of forms showed correct detection of cell positions, with no error. Each cell produces an automatic annotation: a geometric annotation (the polygon of the cell) and a textual annotation (the name of the cell).



Figure 4: Some examples of last names on which automatic annotations are made.

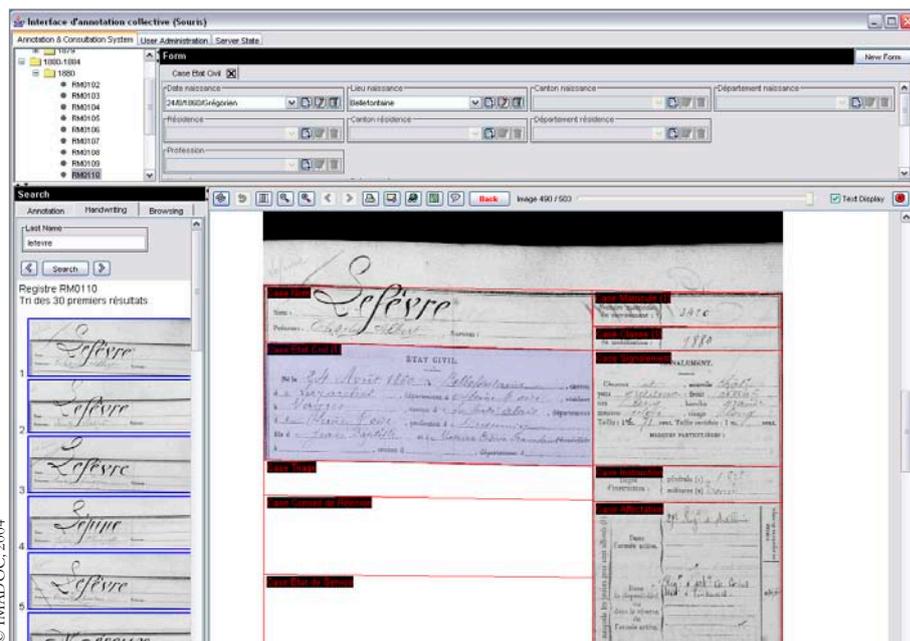


Figure 5: Access by handwritten names in military forms. The request is 'lefevre' from 503 pages. Answers are presented on the left (four 'lefevre' are automatically selected). The reader can annotate cells if he wishes.

Automatic Annotations on Handwritten Text

We worked on the automatic indexing of last names in those military forms with the help of cell locations (some examples of names are given in Figure 5). Dictionaries were not used as they cannot be exhaustive. Using the method previously stated, a user can make a textual request and the system selects the closest match. On 350 different last names, around 200 names are returned in first position when they are

used as request and 80% (280 names) occur in the first ten results. This offers an automatic document retrieval by handwritten last names (see Figure 5). Less than 1 second is needed to search for a name from 5000 images.

Collective Annotations

By changing the configuration file on the platform, it is possible to specify the allowed annotations on these military forms. For example, the cell containing birth information or the cell containing a physical description of the person could be collectively annotated. The user must select the cell (an automatic annotation) to zoom on it and to associate some textual annotations to it. All these annotations could then be used for a future query by another reader.

Applications

The 60,000 images (automatically cropped to remove protected information by using the geometric annotations produced automatically) are publicly available on the *Archives départementales de la Mayenne* Web site (<http://www.cg53.fr/Fr/Archives/>); follow *Archives en ligne* then *Conscrits de la*



Figure 6: Fast leaf-through using automatic annotations on naturalisation decrees. Names are presented in a table at the left. The reader can then access the complete page and annotate it if desired.

Mayenne). Our platform for searching handwritten last names on these military forms will be available in the reading room at the Archives départementales des Yvelines as well as on the Internet in 2004 (when ready, a link will be placed on <http://imadoc-ar.irisa.fr>). Initially 28,000 pages will be open to the public, which will be increased to 250,000 pages during 2004.

NATURALISATION DECREES

Automatic Annotations

These documents are from the end of the nineteenth and the beginning of the twentieth centuries. Documents can be fully hand- or type-written and are organised in two columns with a paragraph concerning one person, whose name is usually the first name in the paragraph. Retrieving a decree for one person is very tedious: the reader needs to leaf through all pages of all decrees.

Compared with the military forms, the structure is very weak – merely paragraphs of handwritten text. Owing to the generic nature of the DMOS method, we have been able to define an EPF grammar describing the organisation of decrees in handwritten text, in paragraphs and columns, using only the connected components detected in the image. From this

description, by compilation, a recognition system has been produced which is able to detect the position of the name and the file number. We tested 1126 images and 99.43% of positions were correctly detected. These positions are the automatic annotations which are added in the platform.

Collective Annotations

When a reader finds the name he was looking for, the platform presents the original page with all the existing annotations. The reader can then leaf through the original pages and, if he wants, add further annotations.

Applications

This work on naturalisation decrees has been carried out in co-operation with the French National Archives: *Centre Historique des Archives Nationales* (<http://www.archives-nationales.culture.gouv.fr/chan/>). A fast leaf-through capability using automatic annotations has also been included in the platform. A reader can then select the name he is looking for and has direct access to the full document (see Figure 6).

CONCLUSION

We presented a platform to improve access to archival documents containing handwritten text. For fast access,

annotations are needed. We consider that annotations for archived documents can be geometric or textual. The platform we propose to manage annotations deals with annotations produced in two complementary ways: automatically with document recognition and collectively with the help of the readers during their reading. The different documents (civil status registers, military forms and naturalisation decrees) on which we present the annotation platform show the importance of a generic system for document recognition. With the DMOS method we have been able to produce new recognition systems with a minimum of development work. The DMOS method has been tested on 95,000 pages and can be applied to highly structured and also less structured documents.

Both the platform and the methodology are important to facilitate access to the content of handwritten documents even if they are difficult to read. The automatic annotations become more or less important depending on how accurately the document can be recognised. The platform and the automatic production of annotations by document image analysis will be validated at a very large scale for the Archives départementales des Yvelines. In 2004, 28,000 pages of military forms will be automatically accessible on the Internet. 1,200,000 images of civil status registers will also be available on the Internet for collective annotation, and 250,000 pages of military forms will be introduced.

ACKNOWLEDGEMENTS

This work has been carried out in co-operation with the Archives départementales de la Mayenne, the Archives départementales d'Ille-et-Vilaine, the Archives départementales des Yvelines and the Centre Historique des Archives Nationales in France, with the support of the French Ministry of Culture, the regions of Brittany and Pays de la Loire, and the Conseil Général des Yvelines.

INTERFACES FOR DIGITAL LIBRARIES AT CNAM (1993-2003)

PIERRE CUBAUD, JÉRÔME DUPIRE AND ALEXANDRE TOPOL,
CEDRIC, CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS (CNAM, [HTTP://WWW.CNAM.FR](http://www.cnam.fr))

INTRODUCTION

Digital libraries (DL) technologies have benefited during the past decade from the impressive increase in digital data capture, storage, and transmission capabilities along with the consequent fall of their cost. The widespread use of the World Wide Web also enables digital libraries to be explored by a very large, international population. For example, the *Association des Bibliophiles Universels* (ABU, <http://abu.cnam.fr>), begun in October 1993, today has thousands of books downloaded daily and has become one of the most active French-speaking cultural Web sites. Another digital library, the *Conservatoire Numérique* (CNum, <http://cnum.cnam.fr>), was put online in January 2000 after being created in January 1998 from a partnership between the library of the CNAM, the research centre for the history of technologies, and the computer science research laboratory. Three hundred reproductions of old scientific and technical books are accessible on this Web site. These digital libraries are accessible online with a WWW-based interface and a standard architecture.

According to e-mails sent to the Web masters, these straightforward WWW services seem to meet users' needs. For ABU, texts are easily and quickly accessible (100 Kb is sufficient to encode a whole book without any compression). Accessing the CNum does not require high levels of computational power, and the GIF file format means that information is delivered in a suitable manner for average dial-up Internet connectivity. However, by looking at ten years of recorded activity (logs), one

can see that the Web interfaces for DLs do not offer a comfort of use sufficient enough to overcome the restricted role of facsimiles for remote printing. Fortunately, low-cost 3D graphic hardware and high bandwidth infrastructures for the Internet (e.g. cable and DSL) are becoming more commonly available and are now widely used by our users. New visualisation and interaction methods for online digital libraries can therefore be investigated to solve some of the common problems inherent in 2D interfaces.

We believe that 3D interaction can offer a better understanding of the three main functions of an online DL user interface:

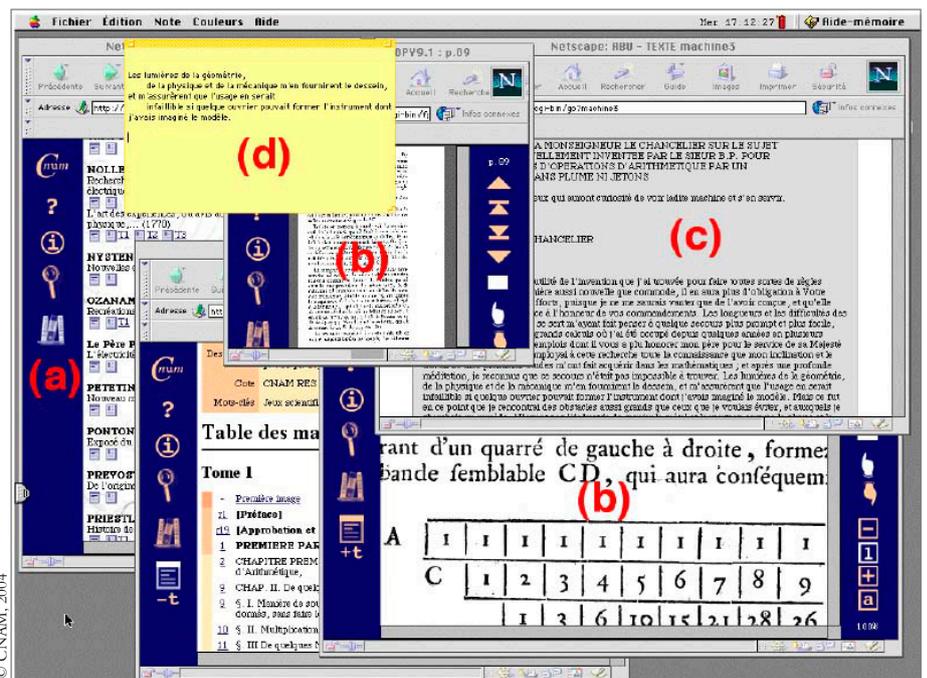
- Catalogue browsing and searching;
- Navigation within the selected docu-

ments, and

- Annotations and bookmark archiving.

Let us first describe how these three steps are handled within the CNum and more generally in many standard digital libraries using common Web technologies (HTML files and scripts). Figure 1 is a screen-shot of a probable user session. Browsing through textual lists is the only method offered to the user for evaluating the DL corpus. A book's content (e.g. chapters and plates) is sequentially described in another window. Cross-reading between books of the CNum and others from related Web services (such as the ABU) is made possible by opening multiple overlapping windows. No authoring tool is provided and CNum users rely on their local bookmarking, archiving, and word processing facilities.

Fig 1: Screenshot of a working session with CNUM



DESIGN CHOICES

Studies have shown that 3D interfaces can be more efficient and more powerful than classical 2D interfaces at managing overlapping windows.¹⁵ This efficiency can be reached when objects and navigation/interaction metaphors are designed depending on the particular context in which they should take place. For 3D digital libraries, there is no reason to reproduce faithfully a real library building and the various steps needed to select and read a book in real life.¹⁶ The only significant positive point in this 'real' to 'virtual' mapping is to ease the recovery of books for users who are familiar only with the real library. However, for users who are familiar with interaction on the Internet, this approach doesn't help. We believe that it is more appropriate to provide abstract navigation and representation and offer services like reordering the collection in space depending on search criteria.¹⁷

In agreement with A. Manguel, who thinks that a book is distinguishable from others by its cover or shape as much as by its authors or title,¹⁸ we decided to use pictures of books to offer visual information. Indeed, it is widely known that libraries' patrons can discover interesting bibliographical information by casual glances at the book collections. This 'visual heuristic' is only one example of the 'many complex information tasks [that] can be simplified by offloading complex cognitive tasks onto the human perceptual systems'.¹⁹

Abstract navigation is also important to minimise moves in a 3D context and therefore ease the training of inexperienced or occasional users. In a real or 'virtually real' library, moves are often necessary to find the shelves containing the book of interest, to pick it up and to go back to a reading desk to work on it. These steps can be simplified in virtual worlds.

THE TWO BASIC TOOLS

The research tool

In the book selection process, a 3D interface avoids browsing through long lists of textual information and also offers additional visual information via images of the books themselves. With the right metaphor to represent the collection, a 3D interface also makes it possible to connect browsing in the library with the widening or refinement of search criteria. We set up an 'on the fly' generation of a new way of organising the library at each user's request. To do so, books are classified in different groups: those that are relevant to all the search criteria, those that answer some criteria but not others, continuing to the group that are relevant to none of the criteria. We then organise the groups geometrically and offer the user the means of switching from one group to another, either to refine or to widen their query. This manner of building and submitting the results to the user enhances the 'visual heuristic' for finding books. Indeed, it makes it possible to gather books *not* relevant to all criteria besides the relevant group, allowing users

to see an interesting or complementary book just by taking a quick glance around. This serendipitous discovery of books is often employed in real libraries but is not, unfortunately, possible with standard digital libraries.

One of our main concerns beside usability is the flexibility of our interface. It is therefore necessary to allow display using different representations. For example, we could show a cylindrical or sphere shaped collection as well as other metaphors like semantic maps, cone trees,

¹⁵ See S. Card, G. Robertson and W. York, 'The WebBook and the Web Forager: An Information Workspace for the World-Wide-Web' in *Proceedings of ACM CHI'96*, Vancouver, April 1996.

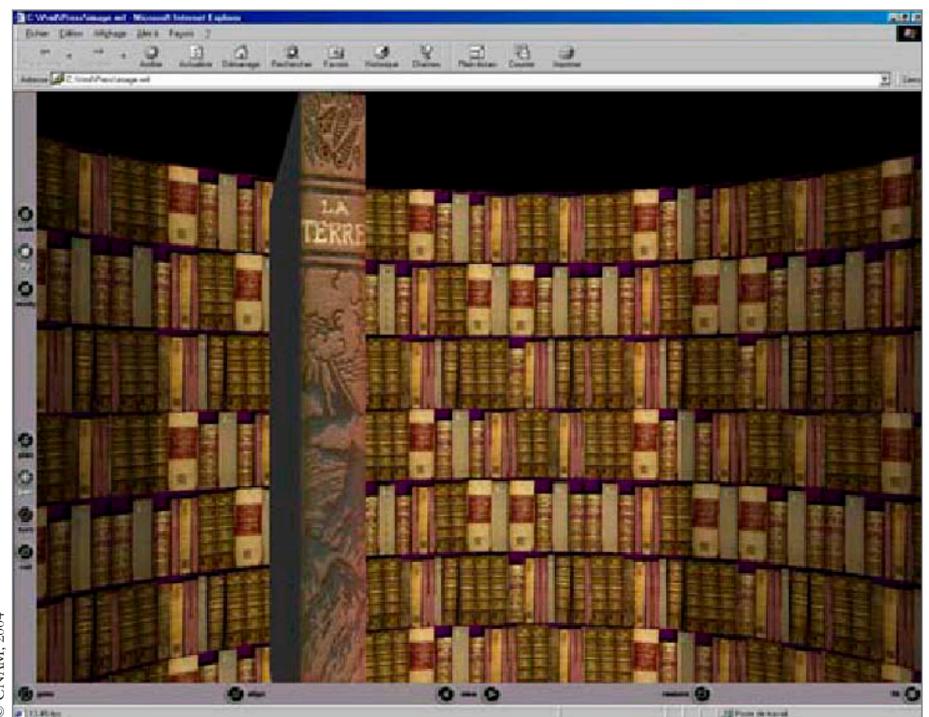
¹⁶ For a comparative view, see E. Fox, N. Kipp et al., 'Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources' in *D-Lib Magazine*, September 1997, and P. Plénacoste, E. Lecolinet, S. Pook, C. Dumas and J-D. Fekete, 'Zoomable and 3D Representations for Digital Libraries' in *Proceedings of IHM-HCI 2001*, Lille, September 2001.

¹⁷ See P. Cubaud, C. Thiria and A. Topol, 'Experimenting a 3D Interface for the access to a Digital Library' in *Proceedings of ACM DL'98*, Pittsburg, USA, June 1998.

¹⁸ A. Manguel, *A History of Reading* (Viking Penguin Group, 1996).

¹⁹ M. Hearst, G. Kopec and D. Brotsky, 'Research in support of digital libraries at Xerox Parc. Part II', *DLib Magazine*, June 1996.

Fig 2: The cylinder-shaped collection



© CINAMI, 2004

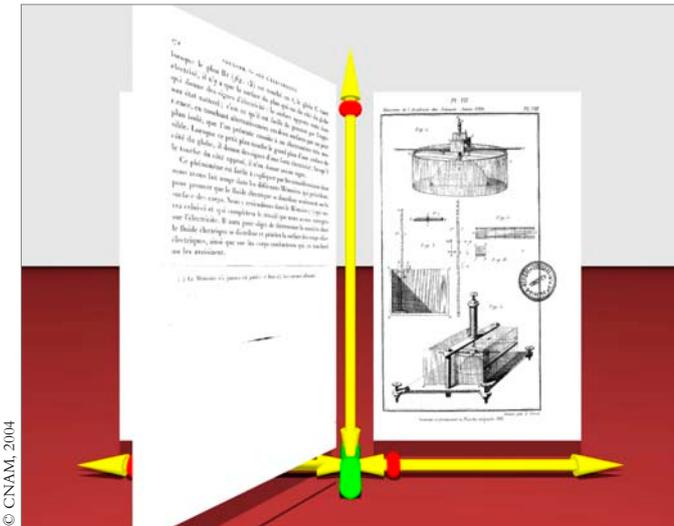


Fig 3a: The book metaphor



Fig 3b: Screenshot of the VRML prototypal application

and 'virtually real' representations. The idea is to display the collection of a particular library with a default metaphor chosen as the most appropriate by the librarian. At the moment, only cylindrical and spherical collections are generated 'on the fly' in VRML file format by a CGI script (see Figure 2). A default sort method for books would also be chosen when entering the library for the first time, but the user could change those settings to increase his ease of use or efficiency.

The reading tool

Our first try was to represent, using a scripted VRML file, the collection and the books enclosed in 2D transparent windows (see Figure 3b).²⁰ We thought that novice 3D users could benefit from this well-known 2D metaphor for organising a 3D environment. However, it appears that these containers are too limiting compared with 3D metaphors.

Using 3D models instead of 2D components is the best solution to obtain more meaningful visual effects. We do not necessarily need a 'virtually real' book but, instead, a book metaphor (Figure 3a) that is more practical and comprehensible at first sight. Interactions with this object are easier to understand and are much more powerful. For example, a tool can be provided

for rolling pages automatically as happens in Web books.²¹ It is an intuitive way to acquire the structure of a book and, as with a real book, it provides a quick means of finding pages of interest (those with images or mathematical formulae, for example). Interacting with 3D objects could lead to a sophisticated but powerful interface. Thus, we have abandoned this first VRML prototype for the book metaphor. The first 3D interface, developed in VRML and Java,²² was presented on the CNum site as an alternative solution to the usual HTML-based reading. However, the difficulties when programming interactive behaviours with this language led us towards some 3D API (OpenGL) and middleware technologies, e.g. Virtools (<http://www.virttools.com/>) and Criterion Renderware (<http://www.renderware.com/>).

INTEGRATION METAPHORS

The activities of browsing through collections and reading content are separated in most present Web-based digital library interfaces. This context break induces longer apprenticeship and navigation time within the interface. Hence, we also studied how 3D interaction metaphors could be used to provide a continuous navigation space for these two tasks.²³ Two options are possible for integrating brows-

ing and reading activities: first, to mix their content; second, to separate them in space and use a 3D metaphor to switch from one to the other. We have sketched such navigation metaphors. The figures below were described and rendered using the ray-tracing program POV.²⁴ In the three integration metaphors, the basket (shown in the three Figures 4, 5 and 6) is the only common tool, always visible and in the foreground. It serves as a link between both interfaces. It is used to group the pre-selected books chosen in the collection interface, to access them in the reading interface, and also to save the user's session.

The cockpit metaphor

The scene's background comprises the collection (Figure 5). The basket and the opened books are in the foreground. They are always visible and are not affected by the user's navigation. In a similar fashion to a pilot flying over a landscape while

²⁰ This prototype application in VRML can be seen on our Web site: <http://cnum.cnam.fr/vrml> or, for more information, see P. Cubaud and A. Topol, 'A VRML-based user interface for an online digitalized antiquarian collection' in *Proceedings of ACM Web3D 2001*, Paderborn, Germany, February 2001.

²¹ See Card *et al.* (1996).

²² See Cubaud *et al.* (1998).

²³ See P. Cubaud, P. Stokowski and A. Topol, 'Mixing Browsing and Reading Activities in a 3D Digitalized Library', *Joint Conference on Digital Libraries 2002*, Portland, Oregon, USA, June 2002.

²⁴ Some animations were also computed and are accessible on our Web site: <http://cnum.cnam.fr/>.

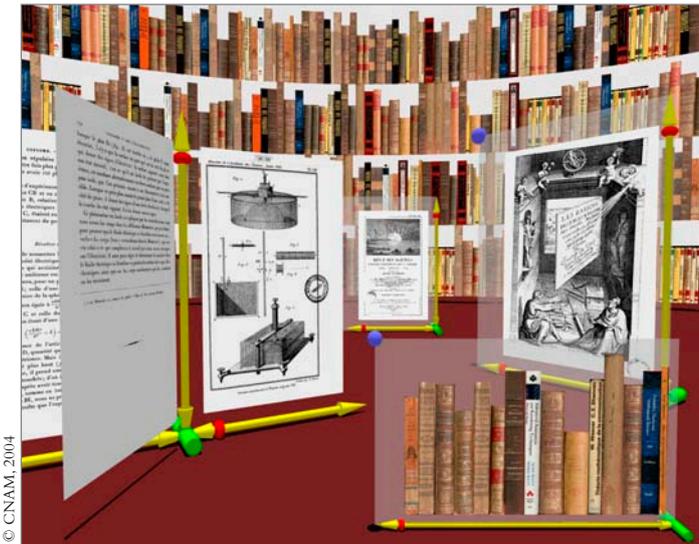


Fig 4: The cockpit metaphor

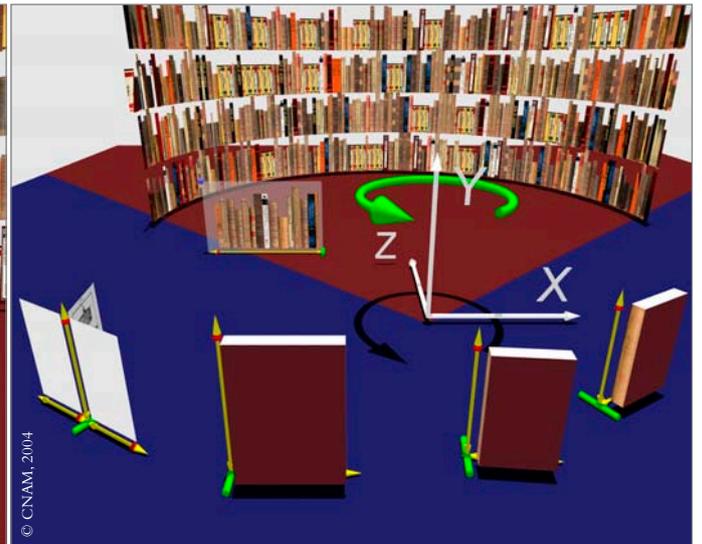


Fig 5: Horizontally structured workspace

seeing data in his eyepiece, in this interface the user navigates through the collection and sees the books. The main problems are the overcharged screen and the difficulty in seeing and accessing the collection behind opened books. Another issue is to maintain a maximum resolution for both tools. Hence, two separated areas in the same interface seems a good choice and, to avoid switching jerkily from one to the other, we considered two other integration metaphors.

Horizontally structured workspace

The space is separated into two horizontal zones (Figure 6). The user can switch from one to the other by rotating around the Y-axis. In the first zone, the collection is rendered and the user can move the point of view. In the second one, the books are always displayed on the ground. Hence, when the user moves in the collection zone, the ground and books also move behind to simulate their immobility. With this metaphor, the user can also create several reading environments depending on their orientation within the reading area and the view angle of the camera.

Vertically structured workspace

With the previous metaphor, after moving in the collection zone, some previ-

ously opened books in the reading zone can be hidden. Moreover, half of the space – under the ground – is not used. To avoid these problems and limitations, we can split the 3D space into two vertical zones. Moving the point of view in the collection zone will not affect the view in the reading zone. To switch from one zone to the other, a rotation around the Z-axis can be used. In the reading zone, the user can still create several reading environments by rotating around the Y-axis. Of course, the user is immersed in the scene and will understand navigation more intuitively than as described here.

CONCLUSION – FUTURE WORK

After a design phase during which we have studied the behaviours of the 3D metaphors within our 3D interface,²⁵ we shall now evaluate our new prototype with a panel of regular readers of the ABU and CNum. We will also work on the annotation interface in order to provide users with the full service in 3D. With CNum this third interface could be replaced by using a standard word-processing tool. Within a 3D environment it must be a specialised interface in order to avoid switching context between 3D and 2D interaction metaphors. A solution would be to enclose 2D appli-

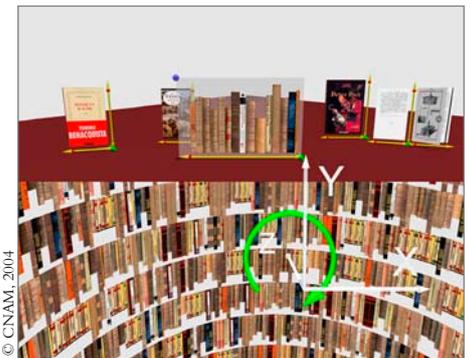


Fig 6: Vertically structured workspace

cations (e.g. word-processing software or an HTML browser) within the 3D scene as we have done in a previous project,²⁶ however, today's situation is the opposite (rendering 3D scenes in 2D windows). We believe that low-cost 3D graphic cards will lead to full 3D interfaces. In such an environment, 3D models (our library for instance) and 2D applications (word-processing tools for example) could share the same common interaction metaphors.

²⁵See P. Cubaud, P. Stokowski and A. Topol (2002).

²⁶See A. Topol, 'Immersion of XWindow Applications into a 3D Workbench' in *Proceedings of ACM CHI'2000*, The Hague, Netherlands, April 2000.

NEWS FROM DIGICULT'S REGIONAL CORRESPONDENTS

GREECE



The Cultural Olympiad

The Olympic Games are more than just a sporting event. They are a major cultural event and a major cultural forum. The Cultural Olympiad <http://www.cultural-olympiad.gr>, is an initiative to promote cross-cultural events and activities that are original and highlight co-operation between cultures. They communicate the notion of solidarity, peace and social cohesion, are articulated both at national and international levels, and seek to raise awareness and mobilization.

ITALY



The Open Access Model in Italy

The Italian situation on digitisation is characterised by important national projects, focused on converting analogue texts into digital image formats, while higher education and research institutions are in the process of constructing many distributed open access information databases. In Italy only universities appear to be interested in supporting and using the open access model while libraries and cultural heritage institutions are unaware of it or do not use it.

Why have digitisation projects in higher education and research institutions started using the open access model? In order to address why this should be the case one needs to ask what open model digitisation really means for universities in Italy. This question has been the subject of two recent seminars, the first held in Parma on 22 November 2003,²⁷ and the second in Florence on 10 February 2004.²⁸

Broader access

The importance of better modalities for the searching out and retrieval of high-quality articles and other digital material as aids to learning is a key issue for Italian academia. A broader level of access is the principal aim of institutional databases held in Italy, with the result that researchers want to store literature such as technical reports and annual departmental reports (low impact publications) but also some articles and learning objects.

The experience accumulated so far has witnessed the involvement of university libraries²⁹ or at least the desire to integrate institutional databases with university libraries' OPAC, linking library catalogues and e-print deposits. An interesting development has been the trend towards the creation of consortia for the provision of services based on institutional deposits and the creation of an infrastructure that is also available to more narrowly based institutions with small amounts of stored content. CILEA (<http://www.cilea.it/>) and CASPUR (<http://www.caspur.it/>), consortia of the North and the Centre of Italy respectively, are two academic consortia that have embarked on a collaborative venture for the provision of this kind of service. In this way, researchers can improve the visibility and the impact of their results at the national and international level.

Another trend is the creation of meta-search engines, to allow the combined searching of different databases using different content management tools. An example of a software toolkit of this kind is OpenDLib (<http://www.opendlib.com/>), which has been developed by the international laboratory D-Lib Competence Centre and the DELOS Network of Excellence. The D-Lib group has tested the software with about half a million documents, including INRIA (Institut National de Recherche en Informatique et en Automatique) digital objects, e-print archives of Florence, Bologna and Trento universities, and CNR (National Council of Research) institutional literature deposits. OpenDLib is not just a search engine but also a value-added service offering benefits such as preservation, access and access control, open peer review, and personalisation of interfaces. It is the first example of the re-use of digital objects in different archives, offering the possibility of new models of combined documents (such as images and video with texts).

Issues

Most of the actual efforts for creating institutional deposits have been devoted to technological infrastructure and there has been a general lack of attention and care paid to organisational and legal issues. This is one of the principal reasons why Italian institutional databases still lack content.

When libraries have promoted the setting up of e-prints databases they have not had enough power to stimulate

²⁷ 'Biblioteche digitali per la ricerca e per la didattica', Università di Parma in collaboration with Casa della Musica, at: <http://www.aldus.unipr.it/master/>.

²⁸ 'L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali', Università di Firenze, at: <http://epress.unifi.it/>.

²⁹ The universities involved in open access repositories are Bologna, Firenze, Padova, Parma, Trento.



submission of digital objects by researchers. Often, promotion has been weak or completely absent. In the cases where there has been an organisation for stimulating and promoting the submission of content, poor attention to legal issues such as IPR (Intellectual Property Rights), plagiarism and privacy was an obstacle or a deterrent to researchers wanting to share their content in open access.

The most successful e-print repository is one that has been related to the Anagrafe della Ricerca (University Research Registry) and to local evaluation of research criteria: if you want to be evaluated, you have to submit your articles to the institutional repository. The lesson to be learnt is that open access is a cultural change that involves the University in its entirety. This means starting at the level of single researchers and libraries, but, to be successful, it has to reach the political and administrative organisation of the University at the top.

At the moment, the top level of Italian universities has not been involved in open access. This may be the reason why the most important characteristics of the open access model – to pay to publish, not to pay to access – is not generally understood to be a revolutionary new publishing method and, also in the case where libraries were leading the way, it is not seen as a solution to the serial price crisis, in competition with traditional print or digital academic publishing. This means that researchers using institutional repositories for submitting pre-prints then publish articles in traditional print journals.

The open access model has not had any significant impact on the research evaluation process and the assurance of quality of articles and digital objects. Few institutional repositories in Italy have organised a validation or peer review process. Career advancement and recruitment of young researchers are still based on print

(and only print) publications and in particular are still based principally on impact factors. It is hoped that the value of these important publishing alternatives will soon be understood, but at the moment universities, government and key experts in all subjects are very conservative in their judgement of an e-article's publication value.

Conclusion

In conclusion, open access is seen in Italy as a solution to broader access. For most of the developers of institutional repositories it is more a matter of faith, without a serious budget plan supporting the project and without any connection to solve periodical price crisis issues. Also, if the open access model is seen as solving academia impact problems, its potential for society more generally is unknown. Researchers still continue to privilege peer-to-peer communication, often using a technical language that is incomprehensible outside the discipline. To promote the diffusion of scientific knowledge by stimulating open access the Government could stimulate better communications between society and academia. However, so far this has not emerged as a pressing issue. Is open access a sector or mass phenomenon? It can be said that open access will be successful when it has become a mass phenomenon. In Italy, however, this will need time to happen.

LITHUANIA

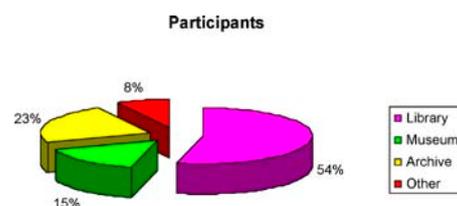


Digitisation courses: Lithuania, March–April 2004

Three-day intensive digitisation courses for the professional community of librarians, archivists and museum specialists took place in Vilnius on 30 March – 1 April and attracted 34 professionals from different

parts of Lithuania. The event was organised by the *Integrated Training Center for LIS Specialists in Lithuania* (<http://www.kf.vu.lt/~btmc/>) and funded by Training Centers Initiative of the Open Society Fund.

Integrated Training Center for LIS Specialists in Lithuania was founded in 1999 and consists of several partners: Vilnius University Faculty of Communication (<http://www.kf.vu.lt/>); Martynas Mazvydas National Library of Lithuania (<http://www.lnb.lt/>); A. Mickevicius Public Library of Vilnius County (<http://www.amb.lt/>); Klaipeda University Library (<http://www.ku.lt/libr/>) and Microlink Networks (<http://www.microlink.lt/>). The training centre offers a wide range of courses for library and information specialists that cover a variety of knowledge and skills from information technology to library management and information policy. VU Faculty of Communication, the main organiser of digitisation courses, expresses an integrated approach to the museum, library and archive professions and for this reason promoted participation of all specialists from Lithuanian memory institutions already involved in or just starting digitisation initiatives.



An experienced lecturer and digitisation specialist, Adolf Knoll from the Czech National Library in Prague, provided an extensive survey of diverse digitisation topics. He is the deputy director of National Library for strategic planning, research and technological development. His rich experience in digitisation includes launching and running diverse digitisation initiatives in the National Library of the Czech Republic, co-ordinating national digitisation programmes on digital access to rare library materials and the digital library,



© Zinaida Manzuch, DigiCULT Regional Correspondent, 2004



Participants at the Vilnius digitisation course with A. Knoll (third from the left in the first row)

and international projects for research and technological development. He is a member of the UNESCO Memory of the World³⁰ programme Advisory Sub-Committee on Technology, author of multiple papers and presentations and an experienced lecturer who has provided training in various countries including those funded by OSI Training Centers Initiative (Kazakhstan, Moldova, Mongolia, Ukraine, and Serbia).

In the series of lectures, a variety of digitisation issues and processes were discussed. Presentations focused on three major topics:

1. Strategic issues of digitisation such as defining the need for digital access to cultural heritage, establishing long-term strategy and a digitisation 'philosophy'.
2. An overview of digital imaging, covering a variety of image formats, compression techniques, current imaging hardware and software and its critical evaluation.
3. Metadata creation techniques, platforms and underlying issues.

All the material was illustrated by experience from projects in the National

Library of Czech Republic and its partners, the results of research by Knoll and his colleagues.

To summarise the experience of participants and track the state of the art in digitisation, the attendees completed questionnaires. The answers revealed a high appreciation of the event and of the competence of the lecturer. However, it became clear that there is a lack of training and information flow between institutions involved in digitisation. The results of questionnaire analysis clearly showed the need for digitisation training initiatives focusing on specific topics and issues and aimed at developing practical skills.

Copyright protection

On 19 February 2004, the Lithuanian Minister of Culture Roma Zakaitiene initiated a discussion about copyright protection in the new digital environment. The main topic of the event was the issue of providing a reward to authors for the use of their creative works for personal purposes. Representatives from Parliament Committees of Education, Information Society Development, INFOBALT

Association, Lithuanian Association of Art and other organisations took part in the discussion. Participants emphasised the importance of the Copyright Law (1999) and Regulation on author reward for audiovisual or phonogram works' reproduction for personal usage (which came into force on 1 January 2004) for effective protection of copyright in Lithuania and successful implementation of European legal norms.³¹

New activities

Having completed successful projects on electronic publishing of cultural content in 2003, Lithuanian museums are planning to undertake new initiatives. The Association of Lithuanian Museums (http://www.museums.lt/Index_En.htm) has launched a significant activity programme for 2004–2005. A working group was established for the creation of the *Literary Lithuania* portal, which aims to provide comprehensive information on literary life in Lithuania and the cultural heritage of Lithuanian writers preserved in the state libraries, museums and archives. Experts from Lithuanian museums, the *Institute of Mathematics and Informatics*, and the *Information Centre of Samogitian Cultural Association* will contribute towards the implementation of the project.

The *Lithuanian Art Museum* (<http://www.ldm.lt>) is implementing Research, Conservation and Restoration of Museum Assets in Lithuania. An interactive electronic database, regularly updated and maintained by museum experts, is an expected outcome of the initiative. These plans will hopefully be realised in the future, and today Lithuanian citizens and tourists are already enjoying digital representations of cultural assets held in Lithuanian museums. In

³⁰ For more about Memory of the World, see the interview with UNESCO's Aziz Abid, also in this issue of *DigiCULT.Info*.

³¹ Information source: 'In the face of new technologies it is necessary to safeguard copyright', Minister of Culture states. Ministry of Culture Press release, 19 February 2004: <http://www.lrkmlt.lt/index.php?ItemId=20641> (in Lithuanian).

2003, the *A. Baranauskas and A. Vienuolis Memorial Museum*, in co-operation with the *Lithuanian Art Museum and Information Centre of Samogitian Cultural Association* (<http://samogitia.mch.mii.lt/index-en.htm>), completed two electronic publishing projects – *Narrow Gauge Railway* (<http://www.baranauskas.lt/index-en.htm>) and *Horse Museum* (<http://www.arkliomuziejus.lt/index-en.htm>), both available in English and Russian – which unfold valuable cultural assets of the Anyksciai district. A narrow gauge railway in Anyksciai is the longest railway of this type in Europe and was built at the end of the nineteenth century. In 2002–2003, when funding was obtained from PHARE 2000 ESS, it was reconstructed and became a popular tourist site. The *Horse Museum*, situated near Anyksciai, is a unique institution in Lithuania introducing the significance of horses to Lithuanian economic development and a picturesque cultural site offering various attractions for tourists. Both museums are departments of *A. Baranauskas and A. Vienuolis Memorial Museum*, so all the important information about museum collections, services and links to its subdivisions can be found at <http://www.baranauskas.lt/index-en.htm>.

Due to the collaboration of *Palanga Botanical Park, Lithuanian Art Museum and Information Centre of Samogitian Cultural Association*, extensive information on Lithuanian amber and the famous Lithuanian resort Palanga and its heritage is available at the new educational Web site *Amber in the Evolution of Lithuanian History and Culture* (http://www.pgm.lt/index_en.htm).³²

³² Information provided by Danute Mukiene, Head of the Information Centre of Samogitian Cultural Association (<http://samogitia.mch.mii.lt/>). Information sources: <http://www.baranauskas.lt>, <http://www.pgm.lt>.

THE NETHERLANDS



Half a million pictures on the Web site of Nationaal Archief

A new image bank has recently become available on the Nationaal Archief Web site (<http://www.nationaalarchief.nl/>). This image bank contains 500,000 pictures from 1890 to 1990. The collection gives a unique view on small and large events, ranging from politics to royalty, and daily life in The Netherlands.

The largest part of the collection consists of press photographs from a former press agency called ANEFO. The collection also contains photographs taken by the photographer Willem van der Poll, while the oldest photographs come from the Ministry of Transport, Public Works and Water Management. In June over a thousand photographs of Dutch national football taken between 1898 and 1940 were added to the image bank.

As well as the facility to retrieve pictures, the image bank offers a few extras. Users can create their own pages and add favourite photographs to their own albums. Unfortunately, the quality of the descriptions is not very good; for example, in many cases it is not possible to identify the persons in the photographs. With this in mind, an option has been built in through which the user can enter additional remarks or descriptions. The image bank does not yet contain high-resolution scans; these will be added in the future.

INCCA: The International Network for the Conservation of Contemporary Art (<http://www.incca.org/>) involves the knowledge sharing of a large number

of leading modern and contemporary art museums and institutions world-wide. The Netherlands Institute for Cultural Heritage (ICN, <http://www.icn.nl/>) and TATE in the UK (<http://www.tate.org.uk>) are co-organisers of this project.

With the financial support of the European Union, INCCA started in 1999 to create a knowledge and information platform for conservators and curators of modern and contemporary art. The aim was not only to disseminate the usual types of information (such as publications, project descriptions and congress announcements), but also to make INCCA a knowledge management tool that supports the daily practice of preservation.

To this end INCCA developed a central registry of metadata records, which describes resources of information on the collecting, preservation and presentation of contemporary art, as well as the information collected from artists on their working processes and artistic intentions. The INCCA Database for Artists' Archives has the following characteristics:

- The XML data input is delivered from remote network environments to the central registry with the support of an Adlib software program.
- The program is based on the Dublin Core metadata standard (enriched with a number of qualifiers).
- The program includes a controlled vocabulary for keywords and agents.
- The descriptive elements are adequate for descriptions of a wide range of information sources.
- Links are included to related information resources published on the Internet.
- Records include data about the providers, i.e. people and/or institutions who are keepers of the source material.
- The system is dynamic in that data are added in the context of a professional activity, such as an exhibition, restoration or project.



INCCA is not an archiving system as such, but in due time the network will generate a large quantity of references to the partners' archives and gradually evolve towards a metadata depository of knowledge and information in this field.

The future target is to facilitate the semi-automatic acquisition of INCCA metadata from the museum's collection management systems, establishing a cumulative effect on the increase of content. More information on the project and its tangible outcomes can be found at the INCCA Web site: <http://www.incca.org/>.

Digital Community of the Past

In January 2004 DIVA, the Dutch Association for Record Management and Archives (Vereniging voor de documentaire informatievoorziening en het archiefwezen: <http://www.divakoepel.nl/>), launched its pilot study 'Digital Community of the Past' (Digitaal Genootschap van het Verleden, DGV). The DGV uses a Napster-type environment to exchange historical information. The core of the system is a peer-to-peer network in which both individuals and cultural heritage organisations can exchange digitised historical information. DGV is supported by a working group of leading professionals in the field of cultural heritage and digitisation. The project is financed by the Ministry of Economic Affairs.

In the past, DIVA, an umbrella association of cultural heritage organisations, carried out a feasibility study into interest in the DGV application among potential users, the technology required, and the costs of operating this digital community. This study showed that both users and suppliers are enthusiastic about the idea of such an interactive application, if it is easy to use.

Following the encouraging findings of this study, the go-ahead was given for three pilot studies to be carried out alongside the DGV. Each pilot study investi-

gates specific issues, such as the relationship between the organisations and individual users, the role of historical societies, and the participation of cultural heritage organisations. In addition, several network models will be tested. DIVA is currently developing a Web site for the Digital Community of the Past. This site will keep you up to date on the progress of the project and the pilot studies. The working DGV will be ready by the end of 2004; the project continues until the end of 2005.

For more information, contact Frans Hoving (project co-ordinator at DIVA) at: fhoving@divakoepel.nl.

Present Past: making Dutch history accessible for a new audience

The Dutch organisation Present Past (Actueel Verleden, <http://www.actueelverleden.nl/>) is a meeting-place of past and present. Present Past aims at increasing historical awareness by making Dutch history easily accessible. Its activities include organising the National History Week, producing a television series and issuing a newspaper on Dutch history.

Present Past was founded in 2002 by cultural organisations such as the National Library of the Netherlands (<http://www.kb.nl/>), the National Archives (<http://www.nationaalarchief.nl/>) and The Rijksmuseum Amsterdam (<http://rijksmuseum.nl/>), and is made possible with a grant from the Ministry for Education, Culture and Sciences. The Editor-in-chief is Wim van der Weiden, chairman of the European Museum Forum.

Our primary audience consists of people who do not usually visit museums. By beginning in the present, focusing on personal stories and creating new means of telling these stories, we will encourage our audience to look back to the past, giving them a better view of the present, as well as the future. Starting primarily as a virtual meeting-place and organiser of tem-

porary events at different locations, Present Past will eventually develop a physical visitor centre in The Hague.

This year, Present Past will launch several projects, linked by the theme 'Identity: Typically Dutch'. Going online this June is a Web site providing historical backgrounds to the news-related topics and enabling visitors to tell their own stories using advanced interactive applications. A network of partners and specialists will help visitors to find answers to questions related to this theme.

SERBIA AND MONTENEGRO



The Third International Conference 'New Technologies and Standards: Digitization of National Heritage 2004' was held in Belgrade, Serbia and Montenegro, 3-5 June 2004. For more information, please visit <http://www.ncd.matf.bg.ac.yu/conferences/en/CallForPapers.html> or look in DigiCULT's Events Database at <http://www.digicult.info/pages/events.php>.

DigiCULT Regional Correspondents:

Efthimios Mavrikas	Greece
Anna Maria Tammaro	Italy
Zinaida Manzuch	Lithuania
Eelco Bruinsma & Johan Oomen	Netherlands
Zoran Ognjanovic	Serbia and Montenegro



RECOGNISING ADVANCES IN DIGITAL PRESERVATION

As more and more digital resources are being produced and used, the work of digital preservation becomes progressively more important in order to prevent a widening gap in our cultural heritage for future generations. The shortlist for a new Digital Preservation Award has been announced by the Digital Preservation Coalition (<http://www.dpconline.org>) and the Pilgrim Trust (<http://www.thepilgrimtrust.org.uk/>). The award recognises innovation and achievement in digital preservation and aims to encourage and highlight creative approaches to furthering the digital preservation agenda. It focuses on 'born-digital' resources rather than those using technology for preservation or conservation purposes and was awarded on 22 June to a project that demonstrates leadership and advancement in the digital preservation area.

Adam Rusbridge, the Digital Preservation Technical Analyst for ERPANET, considered the shortlist:

This is the first set of the Pilgrim Trust Conservation Awards (<http://www.consawards.ukic.org.uk/>) to include an award for digital preservation. Out of nine applications, five have been short-listed.³³ They vary in both scale and direction and this article presents a comparative critique of the shortlist.³⁴

The application form³⁵ states that the award is aimed at projects that:

- Recognise leadership and achievement.
- Highlight the issues posed by preserving electronic information in the long term.
- Focus on born-digital resources.
- Advance the theory, practice and understanding of the surrounding issues.

- Test the theory of a particular digital preservation strategy.
- Assist the long-term storage of electronic objects.
- Illustrate an innovative piece of thinking.
- Demonstrate benefit for the UK.

THE DIGITAL ARCHIVE³⁶

In 1999, the British government released a mandate requiring all records to be stored and retrieved electronically by 2004.³⁷ The National Archives (<http://www.nationalarchives.gov.uk/>) have developed an archive currently capable of storing 100 terabytes of data for this purpose. The successful implementation of the archive is a significant achievement; however, continued management, maintenance and development is crucial to its success. The system is compliant with current government standards and is based upon solid software engineering and project management methodologies. A variety of file formats can be accepted into the archive; in particular, the National Archives have collaborated with the Internet Archive's Wayback Machine (<http://web.archive.org/>) to appropriately archive dynamic Web sites and their associated databases. The National Archives have taken first steps towards implementing a preservation strategy, although remaining flexible enough to allow either migration or emulation where required. It is expected that access and delivery will become Web enabled (it is currently limited to public search rooms).

The Digital Archive directly benefits the UK. It is a British legal require-

ment for this information to be stored and the cost benefit of storing electronic records digitally will benefit the UK public. This project is likely to have a positive impact among other British institutions. The success of a large-scale, high-profile archive such as this is likely to encourage others to begin their own archiving projects and it may be that such organisations will base their ideas upon those implemented here. This is not without merit: as this essential work needed to be performed for legal reasons the techniques implemented quite rightly prioritise security over innovation.

JISC CONTINUING ACCESS AND DIGITAL PRESERVATION STRATEGY³⁸

The Joint Information Systems Committee (JISC) has supported many digital preservation projects in recent years. This strategy, in force between 2002 and 2005, ensures that JISC will continue to play an active role in digital preservation. JISC outlines how it intends to promote the management and preservation of digital materials, encourage the use of standards and good practice, and discuss many of the areas in which research and development is required, focusing on all forms of digital information that must be stored.

³³ Listed at <http://www.dpconline.org/graphics/awards/shortlist.html>.

³⁴ The views presented in this article are those of the author and are in no way affiliated to the awards.

³⁵ The Digital Preservation Award application form can be viewed at http://consawards.ukic.org.uk/appform_dpa.pdf.

³⁶ <http://www.pro.gov.uk/about/preservation/digital/archive/default.htm>

³⁷ See <http://www.archive.official-documents.co.uk/document/cm43/4310/4310-05.htm> for more detail.

³⁸ http://www.jisc.ac.uk/index.cfm?name=pres_continuing



While the strategy highlights and presents many digital preservation issues, it does not include intense analysis of the topics. This tends to be performed by the projects it supports. These projects are often very valuable to the digital preservation community and produce both practical and theoretical outputs that can be further developed. The strategy not only provides a base from which JISC can propose funding and projects, but also a starting point from which UK institutions can develop ideas. JISC is promoting higher education research and supporting best practice products. In this respect, this strategy will help JISC develop a community and assist it in realising its goals.

NLNZ PRESERVATION METADATA EXTRACT TOOL

The Preservation Metadata Extract Tool is the latest release in the National Library of New Zealand's digital preservation strategy. The tool 'programmatically extracts preservation metadata from the headers of a range of file formats'³⁹ and shows that technical metadata can be extracted automatically and separately from semantic metadata. Metadata information is collected according to the NLNZ Preservation Metadata schema; however, it is possible to add further schema options. Although NLNZ is distributing the tool, there seems to be little publicly available documentation describing its underlying design in detail. The project team will have raised and answered many practical questions and the development and implementation cycle is of interest. Although time consuming to produce and release, it is likely that this information will be beneficial to the digital preservation community.

Also of interest is the recent announcement of the Open Source jHove application (<http://hul.harvard.edu/jhove/jhove.html>). Like the NLNZ tool, jHove

extracts technical information regarding a set of file formats. The benefits of the open-source paradigm along with the overlapping fields of work may result in a strong developmental model.

The National Library of New Zealand has a clear direction in which it is heading. Its metadata schema is very clearly defined and aimed at the Library's institutional needs. This tool builds upon the metadata schema and highlights NLNZ's advance into preserving its digital heritage.

WELLCOME LIBRARY/JISC WEB ARCHIVING PROJECT ⁴⁰

In March 2002, the Wellcome Library and JISC commissioned a feasibility study into Web archiving, intended to analyse the challenges presented by current technologies and provide a set of recommendations for how these may be approached. The study consists of two reports, investigating the technical and legal problems presented. Although little practical work was performed, the report evaluated a variety of Web archiving initiatives and provides both an introduction and an in-depth look at the issues discovered.

The value of evaluative reports such as these should not be overlooked. Using these documents, institutions will have a greater understanding of how to generate practical Web archiving solutions, and be aware of organisations to contact for further research. The small scale of the project but wide audience for the results highlights the benefit that can be obtained by performing clear, well-directed evaluative work when a gap in a subject area is perceived.

THE CAMILEON PROJECT

CAMiLEON (<http://www.si.umich.edu/CAMILEON/>), a partnership between Michigan and Leeds universities,

ran from October 1999 to late 2003. The group successfully developed and evaluated a variety of digital preservation strategies. A number of high-profile applications were produced alongside a selection of respected publications. In particular, the work on the BBC Domesday project received much press attention⁴¹ and their innovative Migration on Request concept remains under much discussion among the digital preservation community.

The work performed is of direct benefit to the UK: the team has raised the profile of the digital preservation community here and has provided the foundation for a number of areas in which further development can proceed. Migration, emulation, programming practice, and the OAIS model have all been discussed and evaluated. CAMiLEON was not only limited to theoretical work. The Domesday project overcame both hardware and software problems, proving a large emulation project was feasible, while the Migration on Request vector graphic prototype successfully proved the feasibility of the Migration on Request concept. Importantly, the source code for this was released together with extensive documentation. With a relatively small team, CAMiLEON has given leadership, attracted public attention, advanced both theory and practice, and highlighted many of the issues involved in digital preservation today.

WHO DESERVES TO WIN?

The successful applicant should serve as a good example to the rest of the community. This is difficult to determine as

³⁹ See http://www.natlib.govt.nz/files/ProjectDescription_v3-final.pdf.

⁴⁰ http://www.jisc.ac.uk/index.cfm?name=project_webarchiving

⁴¹ An interview with Paul Wheatley of the CAMiLEON project appears in Issue 4 of *DigiCULT.Info* and is available from <http://www.digicult.info/pages/newsletter.php>.



all of the short-listed applications achieve this. The Digital Archive is an essential, legally required piece of work. The JISC Continuing Strategy will ensure that the problems of digital preservation are not neglected in the near future. As a management and funding committee promoting higher and further education research, they intend to develop a community and allow this community to realise their ambitions. The Preservation Metadata Extraction Tool is a useful software utility and its value will be realised over time. The Wellcome Trust/JISC Feasibility Study produced results useful to a wide audience and will be of benefit to many UK institutions, such as the Digital Archive. Small-scale projects such as this should be encouraged and followed up with further practical investigations. However, it is the CAMiLEON project that I feel is the most suitable to win. It has advanced both theory and practice, highlighted many issues of digital preservation and attempted to provide solutions. Future projects producing results comparable to these can only be beneficial for us to preserve our cultural and scientific heritage.

FINAL NOTE

The Digital Archive was announced as the winner of the award at the British Library on 22 June. The CAMiLEON project was awarded a special commendation certificate. Congratulations must be extended, not only to the Digital Archive and the CAMiLEON Project, but also to all the short-listed entrants for their excellent respective submissions.

More about the award and the MDPC can be found at their Web site <http://www.dpconline.org> and, for more information about the Pilgrim Trust Conservation Awards and the full shortlist, please visit <http://www.consawards.ukic.org.uk/>.

EDITION PRODUCTION TECHNOLOGY (EPT) AND THE ARCHWAY PROJECT⁴²

UNIVERSITY OF KENTUCKY: KEVIN KIERNAN, PRINCIPAL INVESTIGATOR, DEPARTMENT OF ENGLISH; ALEX DEKHTYAR, CO-PI, DEPARTMENT OF COMPUTER SCIENCE; JERZY W. JAROMZCYK, CO-PI, DEPARTMENT OF COMPUTER SCIENCE; DOROTHY CARR PORTER, RESEARCH IN COMPUTING FOR HUMANITIES; IONUT EMIL IACOB, DEPARTMENT OF COMPUTER SCIENCE

The ARCHway project (Architecture for Research in Computing for Humanities through collaborative research, teaching and learning) is an unusual collaboration of scholars and students in computer science and the humanities who seek to identify and solve problems of mutual importance in building image-based electronic editions of significant cultural materials. To accomplish these goals, ARCHway is developing a workbench of integrated tools, called Edition Production Technology (EPT). Using its underlying programming platform, Eclipse (<http://eclipse.org/>), we are creating new editing tools, integrating those tools already developed under other projects, and then using these tools to prepare image-based electronic editions.

BACKGROUND

The concept of electronic editing tools began under the Electronic Beowulf project, when the need for these tools became evident, and continued under the

Digital Atheneum project with the development of a Glossary Tool.⁴³ The production of focused, well-designed, XML-based editing tools began in earnest in Autumn 2002 under the Electronic Boethius project (see Figure 1).⁴⁴

In this phase the *Electronic Boethius* team, under the programming leadership of Ionut Emil Iacob, created a Java prototype EPT called the Electronic Editions Environment, or E3, which integrated the Glossary Tool with text and an XML tagging capability to facilitate searching of text and glossary. During this period the team also developed several stand-alone tools in Java, designed to incorporate images, which E3 also integrated.⁴⁵ Thus, while we had a prototype of EPT in E3, the continual requirement to integrate emerging new tools posed a significant programming problem.

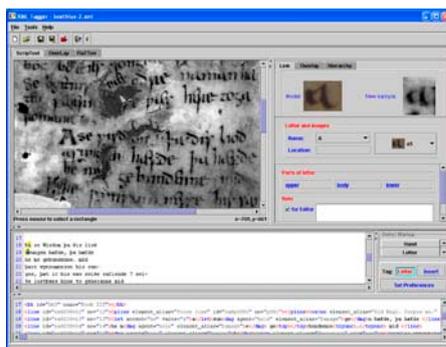
⁴² ARCHway is supported by the National Science Foundation under Grant No. 0219924, awarded pursuant to the authority of the NSF Act of 1950 (42 U.S.C. 1861 *et seq.*), which is subject to GC-1 Grant General Conditions (10/98) and is made in accordance with the provisions of NSF 98-63, 'Information Technology Research.'

⁴³ For details of ARCHway, and these other projects, follow the links at <http://www.rch.uky.edu>.

⁴⁴ The *Electronic Boethius* Project is funded by a Collaborative Research Award from the National Endowment for the Humanities and the Andrew W. Mellon Foundation, with images provided by The British Library and the Bodleian Library, Oxford.

⁴⁵ Chengdong Li was responsible for programming most of the tools that manipulate images. The *Electronic Boethius* Project is funded by a Collaborative Research Award from the National Endowment for the Humanities. It received additional support for the development of these editing tools from the Andrew W. Mellon Foundation. The British Library and the Bodleian Library, Oxford, provided the digital images for this project.

Fig 1 - *Electronic Boethius Electronic Editions Environment (E3)*



THE INTEGRATED PLATFORM

When the ARCHway project began in Spring 2003, Jerzy W. Jaromczyk, one of two co-Principal Investigators (PIs) in Computer Science, advised us to reprogram these stand-alone tools and develop all new tools using the Eclipse programming environment, an open-source platform originally developed by IBM, now broadly used and actively enhanced by the open-source community.⁴⁶ Although this move required the reprogramming of our editing tools, Eclipse suited the needs of ARCHway and the *Electronic Boethius* in two ways. First, it provided an effective software architecture for the production and deployment of the EPT, which is now organised as a set of distinct plugin tools that work together through Eclipse. Each tool is responsible for a specific editing or administrative task, and the tools can also work with one another through the platform, effectively borrowing functionality and allowing for the creation of new tools without having to reprogram established functions.⁴⁷ Second, Eclipse is an ideal environment for the teaching and learning aims of ARCHway. The EPT's plugin design makes it possible to assign individual tools to research assistants, to computer science students as class projects, and to teams of computer science and humanities students collaborating on Master's and Informatics projects.

The programming teams use the appropriate Application Programming Interfaces (APIs) available in the Eclipse Plugin Development Platform to build all editing tools. To ensure that these tools are in fact useful for editors, the programmers always work under the guidance of the editor/PI and in consultation with humanities research assistants. The programmers can integrate tools-in-progress into the EPT, because Eclipse comes with a Concurrent Versioning System (CVS), which permits multiple users to modify the same files without overwriting one another's work.⁴⁸

CVS is especially valuable for ARCHway, as everyone works collaboratively on both programming and editing. CVS also contributes in important ways to the teaching and learning goals of ARCHway, because for testing and grading purposes the Principal Investigators require continual, reliable access to the most recent tools and editing projects.

Another advantage of Eclipse is that it helps achieve uniformity, adaptability and extensibility in areas that can raise high hurdles for complex humanities computing projects. The Graphical User Interfaces (GUIs) developed under Eclipse are attractive and uniform and humanities editors can easily configure them without new programming support by using XML configuration files. Eclipse works across operating systems, allowing Windows and Linux (and Macintosh, to some extent) to support the same tools with the same native appearance. The programming platform also provides automatic updating for the emerging tools, a critical capability that encourages the refinement and expansion of features as well as the correction of programming errors or bugs in the EPT. In the long run this capability will enable automatic online upgrades to the completed electronic editions. With its strong support for XML and its high-quality imaging capabilities, Eclipse is thus in many ways ideal for the development of image-based electronic editions.

EDITION PRODUCTION TECHNOLOGY

The EPT now consists of three software layers: one for editing and administrative tools, one for middleware, and one for data management (Figure 2).

The editing and administrative tools provide the functionality for managing projects and editing primary resource images and text, just as the presentation tools will eventually provide the functionality for using the completed image-based editions in interactive displays and search-

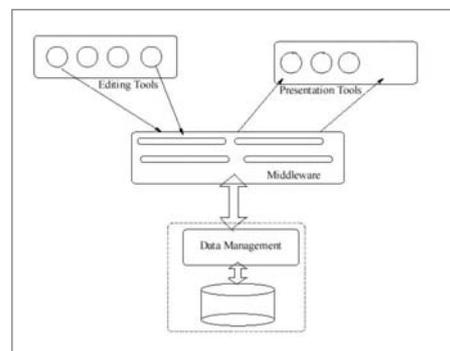


Fig. 2 - ARCHway Model for Edition Production Technology (EPT)

ing facilities. The middleware layer, under the guidance of Jaromczyk and the other Computer Science co-PI, Alex Dekhtyar, provides the utility plugins that allow the upper-level tools to communicate with each other and share image-enriched information of all kinds from the data management layer, Dekhtyar's domain. The data management layer contains the routines devoted to storage, maintenance and retrieval of the information from the image-based electronic edition. The utility plugins in the middleware layer provide functionality that is shared with the editing and administrative tools. The Project Explorer organises current projects and completed editions, and provides a logical view of all project files regardless of their physical location. The Data Source Layer acts as a middle ground between the EPT editing tools and the project files and provides the physical location of the project files for the editing tools. One utility plugin that the editor uses is the Keyboard, set by default to the Old English character set. Like all the plugins, the Keyboard plugin is easily modified in configuration files to support character sets for other languages and other projects.

⁴⁶ For more information, see Eclipse Platform Technical Overview, Object Technology International, Inc., February 2003 (<http://www.eclipse.org/whitepapers/eclipse-overview.pdf>).

⁴⁷ For a more detailed description of the software, and the ARCHway project in general, see Kiernan *et al.*, 'The ARCHway Project: Architecture for research in computing for humanities through research, teaching, and learning', forthcoming in *Literary and Linguistic Computing*.

⁴⁸ For details about the open-source Concurrent Versioning System, visit the homepage at <http://www.cvshome.org>.

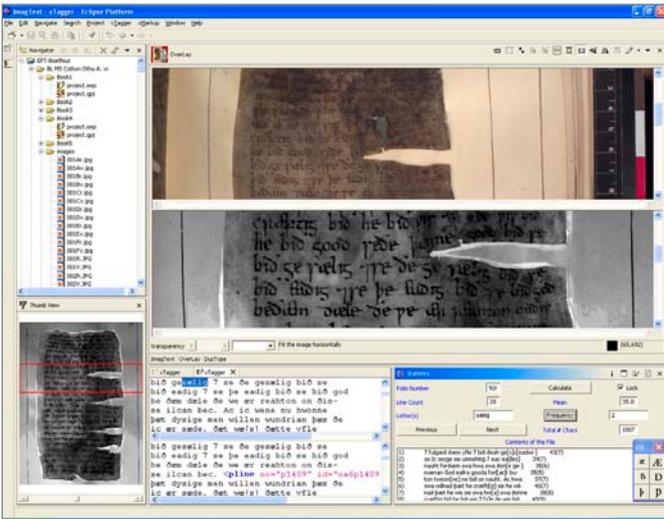


Fig. 3 - OverLay perspective with xTagger and Statistics tools

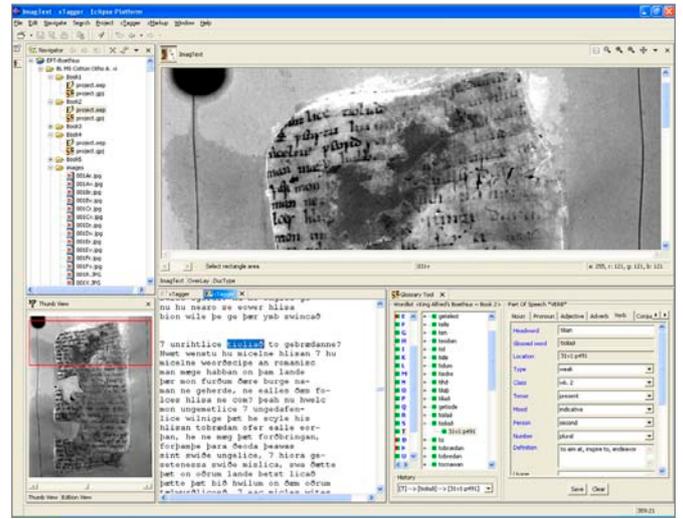


Fig. 4 - ImagoText perspective with xTagger and Glossary tools

THE EDITING TOOLS

The main tools originally developed under the *Electronic Beowulf* project include a ScripText environment for integrating images and text; a Glossary Tool for building comprehensive glossaries from an XML text file; a Tagger for inserting XML markup, based on the images, in the text file; a DucType tool for paleographical description, analysis, and encoding; and an OverLay tool for comparing and encoding multiple images of a folio taken under different lighting conditions. Although first developed in the Electronic Editions Environment (E3), the tools held a somewhat precarious existence as independent, stand-alone programs. Under ARCHway the reprogrammed tools now form the basic editing toolkit for the EPT. We are continually developing new tools under both projects and integrating them into the EPT. An editor can organise these tools in any desirable combination, called ‘perspectives’ (see Figure 3), to suit different editing and administrative tasks.

The editor can save perspectives and navigate between different perspectives, to perform any number of editorial and administrative tasks in the same project (see Figure 4).

USING THE EPT

To keep the EPT as adaptable, extensible and interoperable as possible, ARCHway is using in its testbed three important manuscripts from the British Library, *Beowulf*, Alfred the Great’s Old English translation of Boethius’s *Consolation of Philosophy*, and Aelfric’s *Lives of Saints*. One by one and as a group these fire-damaged manuscripts⁴⁹ present editors with widely different editing problems and computer scientists with equally challenging technical problems. The Electronic *Beowulf* at once serves as our guide for a fully functioning image-based electronic edition (see Figure 5) and as a ‘legacy document’ we expect the EPT to transform in the course of time.

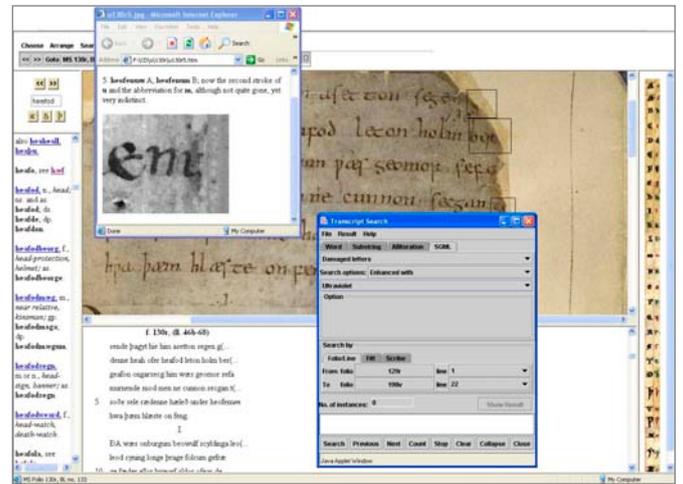


Fig. 5 - Electronic *Beowulf* with image, text, glossary, textual note, and search facility

While we are creating the EPT for editing Old English manuscripts, the ARCHway project has as its long-term goal the general purpose of contributing ideas and practical solutions for preserving and propagating any hand-written materials from the vast and varied heritage of world culture. We believe that ARCHway’s EPT architecture for building image-based electronic editions is an effective model for achieving these ends.

⁴⁹ Andrew Prescott, ‘Their Present Miserable State of Crementation’: the Restoration of the Cotton Library. Sir Robert Cotton as Collector: Essays on an Early Stuart Courtier and His Legacy, C. J. Wright, ed. London: British Library Publications, 1997, pp. 391-454.



CWIS – COLLECTION WORKFLOW INTEGRATION SYSTEM

ADAM RUSBRIDGE,
ERPANET TECHNICAL ANALYST

The National Science Foundation funded CWIS Open Source software (<http://scout.wisc.edu/Projects/CWIS/>) is a tool to ‘assemble, organise, and share collections of data about resources [...] conforming to international and academic standards of metadata.’ Many institutions and organisations maintain portals of resources that can be accessed by a user. This software enables staff to manage and control the information with greater ease, and provides a user with a higher degree of interactivity with the information. The system requires a Linux, Apache, MySQL and PHP (LAMP) server configuration to run and is licensed under the GNU General Public License.

The system has been designed with resource discovery in mind. It conforms to the OAI Protocol for Metadata Harvesting and allows an RSS feed to be generated and updated. The OAI protocol provides search engines with information about the collection, and the two features combined improve the visibility of the content and the ease with which a user may discover specific and appropriate information.

To manage the portal, a clean administration interface is provided. Administrators can edit the System, OAI or RSS configurations, alter the metadata configurations (to unique specifications if necessary), edit and track user accounts, and import and export data. Naturally, the adding and editing of resources is a central part of the system.

When searching for information, it is common for a user to spend a large amount of time browsing through extraneous or sub-standard resources. Rating and Comment features provide peer reviews of the value and content of the resource. These lightweight additions can be very useful to a user but can simply be ignored if not needed. A recommendations feature is provided to a user dependent upon how they have rated other resources. Finally, it is possible to set up a continuous search function providing a weekly e-mail containing new resources that have been added according to a certain set of criteria.

There are many techniques that can be used to generate a resource collection and it is difficult to classify a leading method. CWIS contains many features that should be included in a resource collection and I, for one, will be watching this project with interest.

More information about the Collection Workflow Integration System is available from <http://scout.wisc.edu/Projects/CWIS/>.

STEGANOGRAPHY EXPLAINED

The new DigiCULT *Technology Watch Report 2* (available at <http://www.digicult.info/pages/techwatch.php>) has a chapter discussing Rights Management for cultural and scientific heritage institutions. Digital media are particularly susceptible to illegal copying and distribution and heritage institutions who wish to make their digital collections available online should be cautious about presenting digital objects without a way to identify ownership, copyright and usage rights. The *Technology Watch Report 2* identifies various ways in which an institution can protect its digital holdings from fraudulent use; however, this short article concentrates on protecting images using steganography techniques.

WHAT IS STEGANOGRAPHY?

The word ‘steganography’ comes from Greek and means literally *covered writing*. Today it refers to hiding one piece of information inside another, a common example being the invisible watermarking of images. Other steganography methods include invisible inks (for example, lemon juice which darkens when heated), micro-dots (up to one page can be hidden in the space of a dot), and digital signatures (a cryptographic method of verifying information). In a nutshell, steganography is any way of communicating that hides the actual fact of communication itself.

An early example, from the Histories of Herodotus, was the use of steganography in the wax-covered tablets used in ancient Greece for writing. When sending delicate information that needed to escape the notice of sentries, the wax could be scraped off the tablets, a message written on the wood beneath and the wax reapplied with a decoy message. A different method was to shave the head of a messenger and tattoo a secret message onto his scalp, which would be undetected once his hair had grown back. However, the art and science of steganography has come a long way since then! It is not only messages in the form of text and images that can be concealed, but more advanced tools such as agents or programs. This means that, in addition to ‘passive’ data, steganography can be used to carry ‘active’ data that can perform actions. Examples could include initiating a registration program when a file is downloaded, or ‘self-destructing’ if the criteria for owning the file are not met.

A digital watermark is an image, digital signal or pattern that is embedded permanently into another media type and can act as a digital signature for any copies made of that digital object. Watermarks can be visible (for example, a fairly unobtrusive logo which displays property ownership and can dissuade illicit use of the image)



or invisible. Invisible watermarks cannot be seen but can be detected by special algorithms, facilitating tracking of illegal distribution of copyrighted materials. There are two types of invisible watermarks: 'destroyable' watermarks and watermarks that are highly resistant to destruction and can therefore be used to prove ownership.

HOW DOES IT WORK?

Messages in electronic form can be hidden within any digital media type (images, audio files, or even other text messages) by exploiting the storage capacity within inessential pixels or lines of code in a way that does not alter the look or sound of the original file.

In steganography, two types of file are usually used when embedding data into an image. A 'container' file (for example, a digitised work of art) is the packaging for a secret 'message' file (the content of which could range from a digital watermark to a military transmission). Data describing the message are then embedded into the least significant bits (LSBs) of the container file. This technique works particularly well on images which have subtle colour gradations, as the bits which are altered to store the message change the colour of the pixel so little as to be imperceptible to the human eye. This method is sometimes also known as 'wrapping'. For an example of digital wrappers, see InterTrust's DigiBoxes (<http://www.intertrust.com>). It is important to note that many images are compressed using algorithms in order to reduce the file size for transmission over networks. Image compression techniques that are 'lossy' such as JPEG are not suitable for 'wrapped' data, as much of the original information cannot be recovered and this risks damaging the hidden message. Lossless compression such as GIF should be used instead.⁵⁰

In watermarking, a visible watermark (which is not strictly steganography as it

does not hide itself) can simply be added to the image itself, with no hidden message. A visible watermark can act as a deterrent to misuse; however, it does interfere with the study of the image and can be relatively easily removed using most image manipulation software. Invisible watermarks are less obvious and therefore more difficult to remove.

To find out more about steganography, look at the NINCH Guide to Good Practice: Rights Management (<http://www.nyu.edu/its/humanities/ninchguide/IV/>) or, to find a suitable tool, try the Digital Image Steganography and Digital Watermarking Tool Table (<http://www.jitc.com/Steganography/toolmatrix.htm>).

Screenshot Memory of the Netherlands website



© Memory of the Netherlands, 2004

MEMORY OF THE NETHERLANDS (HET GEHEUGEN VAN NEDERLAND): A PROGRAMMATIC APPROACH TO THE DIGITISATION OF CULTURAL-HISTORICAL RESOURCES IN THE NETHERLANDS

PAUL DOORENBOSCH, MA
KONINKLIJKE BIBLIOTHEEK,
NATIONAL LIBRARY OF THE NETHERLANDS

INTRODUCTION

Memory of The Netherlands (<http://www.memoryofthenetherlands.nl>) is the Dutch national digitisation programme. The programme aims at making available a large amount of cultural-historical collections to a wide audience, in an efficient way and in a relatively short period of time (three to four years). Another aim of the programme is to develop educational applications for a number of the collections. The collections are contributed by various heritage organisations, including museums, archives and libraries. Currently more than 45 organisations are involved. Memory of The Netherlands has been available on the Internet since May 2003. By the end of

2004 (the start of the second phase), it will contain 750,000 images and text pages, 250 hours of moving images, and 100 hours of audio. By then, at least twenty educational applications will have become available, too.

A BRIEF HISTORY

In 1999 the Koninklijke Bibliotheek (<http://www.kb.nl/>) submitted a project proposal to the national government. This proposal formed an important step towards the creation of a digital collection of resources on Dutch history, art and culture:

⁵⁰ For more information, see *Steganography: Seeing the Unseen* <http://www.jitc.com/pub/r2026a.htm> or the detailed descriptions available from <http://www.infosyssec.org/infosyssec/cry2.htm>.

Memory of The Netherlands. New to the Koninklijke Bibliotheek was the choice for a general, rather than an academic, audience.

Although the proposal benefited from the experience that was gained in previous digitisation pilots undertaken by the Koninklijke Bibliotheek, a considerable scaling-up was required as regards content and organisation. What makes the Memory special is its programmatic approach, which is large-scale, comprehensive, and has a professional set-up. The Memory covers art, culture and history in the broadest sense. The main focus is on digital content.

The project proposal also addressed the problem of how to develop a stable and scalable technical infrastructure. An additional focal point of the project concerns the acquisition and dissemination of knowledge about digitisation and project management. After reviewing the project proposal, the Dutch Ministry of Education, Culture and Science granted the Koninklijke Bibliotheek a substantial subsidy, with the advice to devote particular attention to the educational component, and to reinforce the digital learning environment in secondary education.

A PROGRAMMATIC APPROACH

Memory of The Netherlands introduces a programmatic approach to the digitisation of cultural-historical materials in The Netherlands. The unifying elements in the programme's content are history, art and culture in the broadest sense. On the organisational level, the programme is characterised by co-operation, professionalism, and large-scale digitisation. The advantage of this programmatic approach is that the processes can be managed in a professional way. Previously, this approach proved effective in *Metamorfoze* (<http://www.metamorfoze.nl>), a large-scale national preservation programme run by the Koninklijke Bibliotheek. The Memory uses a large amount of material and occupies a

strategic position. As such, it can negotiate favourable prices with scanning companies without compromising the quality of the material. The approach of the various sub-projects is highly uniform. More and more materials are becoming available for comparison. This allows a better insight into both processes and costs. As a result, more effort can be put into the development of the technical and organisational infrastructure. Knowledge is better preserved and disseminated. A durable, stable form of access can be guaranteed. Ultimately, all this contributes to providing the public with better services.



Kettingbotsing op het Damrak / Accident on the Damrak (Amsterdam)

Missionaris op Irian Jaya / Missionary on Irian Jaya (former Dutch New Guinea)
Used with permission of Katholiek Documentatie Centrum.



EDUCATION

From the start, Memory has worked on examples of re-use in educational applications of heritage materials. Initially the idea was to create a separate educational interface on the entire Memory collection, but based on experiences from school practice this was abandoned in favour of the development of a number of varied applications.

As their actual usage would be the best indicator of success, the following starting points were drawn up for the applications:

- Compliant with the school curriculum and examination requirements.
- Attuned to existing teaching methods.
- Educationally sound.
- Developed jointly with people working in education and tested in the school environment.
- Attractive to both teachers and pupils.
- Discrete entities.
- Complete with a 'reward' or check system.
- Demanding few technical skills in preparation, use and maintenance.
- Compliant with existing, widely available software and databases.

Further development should be transferred to third parties. Content providers can suggest materials and ideas, but the expertise to develop and maintain applications should come from the educational sector and educational publishers.

GOOD PRACTICE AND LESSONS LEARNED

The first phase of the project will last until the end of 2004. At the moment it is too early to draw any conclusions from our experiences, especially regarding practical use and demand. However, there are some aspects of the process that can already be commented on:

Scanning costs

Scanning costs should go down further to make large-scale digitisation of cultural-historical resources feasible. The large diversity of materials and the special treatment demanded by the unique materials cause staff costs to remain high. The costs of encoding moving images have decreased considerably. The increased use of broadband requires a higher image quality, which will increase the production and storage costs.

Standards

Memory of The Netherlands uses a simple metadata scheme, which accommodates all descriptive and semantic metadata that is available in each organisation. The search form, developed for a wide audience, offers the search options 'who', 'what', 'where', 'when' and 'search all fields'. In many cases, only the last option will yield acceptable results, since the structure and the content of the underlying metadata are heterogeneous and inconsistent.

All metadata is stored in a Dublin Core-like format. Conversion to DC (extended) is planned for the near future. The same holds for the version of the SRU protocol that is used. By upgrading these, we will meet international requirements for interoperability.

Metadata

The creation of descriptive metadata is costly, even when fairly minimal. The quality of metadata is often poor, ambiguous or completely lacking. The development of tools to generate metadata (semi-)automatically seems feasible for text-based materials, but comparable aids for image materials are still in their infancy. The heterogeneous nature of data formats and methods for semantic-based access that Memory uses allow for little precision. To standardise this, and to enrich the metadata by intellectual effort only, is not feasible. In due course, IT tools will have to provide a solution.

Process Control

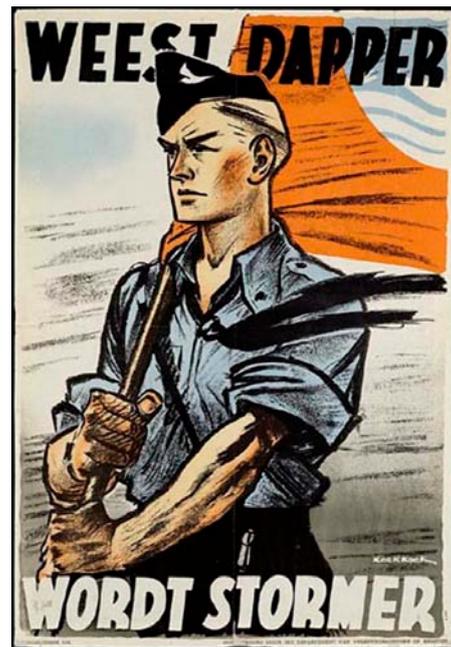
For the final result of digitisation projects it is worthwhile to invest in monitoring the quality of the process. Digitisation of heritage materials does not differ much from other project activities, but co-operation between content experts and technical developers is essential. It pays off to make an extensive analysis of the materials (objects and metadata) during the preparatory phase and to document precisely every stage of the digitisation process. In the end, this will contribute to both efficiency and quality.

Integration of heterogeneous materials

Making heterogeneous materials cross-searchable yields considerable advantages to users. This integration should be realised mainly on the level of the metadata. Storage, subject description, and presentation of the various types of materials each require specific methods, which can be accommodated only in a modular way. In the Memory, text and images have been dealt with in similar fashion. This limits the options for searching and presentation of texts. For this reason, a separate trajectory has been chosen for newspapers, where the integration will be realised only in the end-user interface. Text materials in the Memory (a growing component) will be revised next year to allow for a better presentation.

THE FUTURE

In the future, the Memory's functionality, i.e. the possibilities to search and retrieve the objects, will gradually be improved and extended. Furthermore, a scalable infrastructure has been realised that can be used by future digitisation projects. The



*Weest dapper, wordt stormer (Be brave, be a stormer)
Used with permission of NIOD (Nederlands Instituut voor Oorlogs
Documentatie)*

extensive knowledge and experience that will have been gained and documented will result in a professional organisation, clear agreements, and improved efficiency. The government is willing to support, and continue to support, the Memory of The Netherlands as an integrated part of the digitisation framework in The Netherlands. Both end-users and other parties will benefit from this.

EVENT REPORT: PRESTOSPACE AND AUDIOVISUAL PRESERVATION

RICHARD WRIGHT, TECHNOLOGY MANAGER,
PROJECTS INFORMATION & ARCHIVES,
BBC FINANCE, PROPERTY & BUSINESS
AFFAIRS

SUMMARY

EC Sixth Framework funding in the cultural sector is concentrated in large projects. This concentration of effort has produced many disappointed applicants – and also places a high burden of responsibility on those few successful projects. PrestoSpace (<http://www.prestospace.org/>) is the only Integrated Project (with-in Access to Cultural Heritage) in the preservation area funded by the first FP6 Call. Accordingly, the project is making a sincere effort to determine actual user needs, so that this one project can be as representative of true need as possible. The first public action of PrestoSpace, which formally started on 1 February 2004, was a 1.5-day User Requirements meeting in Amsterdam on 18-19 March. A full house of 70 delegates attended the meeting, held in an appropriate venue: the historic Felix Meritis building. The building is the home of a 250-year old movement fostering international culture and co-operation. The delegates represented large and small audiovisual archives, and the technical sector associated with these archives.

THE PROGRAMME

The programme was a day of presentations and discussion at Felix Meritis, followed by a half-day of film presentations at the Netherlands Filmmuseum.

The programme for the first day covered:

- The purpose of the PrestoSpace project, by Daniel Teruggi and Jean-Hugues



© Richard Wright, 2004

Felix Meritis building in Amsterdam

Chenot of the Institut National de l'Audiovisuel in France.

- The European perspective, from Maurizio Lunghi of the Minerva project.
- The broadcast perspective, from Richard Wright of the BBC archives. The special needs of film, discussed in more detail below.
- The specific work areas of Prestospace: digitisation, restoration, storage and metadata.

The afternoon was devoted to detailed discussion of these areas, with emphasis on determining the real need, not just the academic interests of the researchers.

THE PROBLEMS

PrestoSpace is about preservation of European audiovisual (AV) material. While considerable attention has been

devoted to conventional archive contents, audiovisual materials are the single highest 'at risk' category and have urgent problems. This risk is due to:

- **Fragility:** film, vinyl, shellac (for gramophone recordings) and tape are easily damaged, and in the case of nitrate film also flammable or even explosive.
- **Decay:** chemical deterioration of the media. All AV material is subject to slow (and not-so-slow) chemical change, with magnetic layers coming off polymer bindings, clogging players and in the worst cases destroying the media. Acetate-based material (used in film, video and audio recordings) turns to acetic acid, which not only can destroy the acetate media, but also can attack non-acetate material in the same storage area.
- **Obsolescence:** players for AV formats (such as 6-mm open-reel audio tape recorders, gramophone turntables, U-matic video tape) become obsolete in as little as ten years – a problem that is accelerating with new digital media.
- **Expense:** the problem of chemical decay can only be retarded by conserving materials at low temperature and humidity, which adds the final problem (a problem shared by all archives and all archive contents): money. Proper storage of AV materials is expensive. Proper equipment, staff and preparation of access copies are also expensive in comparison with non-AV media.



© Richard Wright, 2004

From left to right: Daniel Teruggi - PrestoSpace project co-ordinator and Head of R&D, INA, Giovanna Fossati - Filmmuseum, Thomas Christensen - Danish Film Institute, Eddy Goray - RTBF head of New Technologies Department and First project representative

THE SOLUTIONS

European Action

The European Commission has recognised the scale and severity of AV preservation problems, as have the professional societies representing AV collections. There is now a basis for concerted European action on preservation of these materials. An example of this recognition and co-operation has been the various FP5 projects in the area:

- BRAVA (<http://www.ina.fr/recherche/projets/encours/brava/>)
- AURORA (<http://www.ina.fr/recherche/projets/finis/aurora/>)
- DIAMANT (<http://diamant.joanneum.at/>)
- SEPIA (<http://www.knaw.nl/ecpa/sepia/home.html>)
- PRESTO (<http://presto.joanneum.ac.at/index.asp>)
- ARCHIMEDIA (<http://www.ledoux.be/archimedia/>)
- DELOS (<http://delos-noe.iei.pi.cnr.it/>)
- ERPANET (<http://www.erpanet.org/>)
- MINERVA (<http://www.minervaeurope.org/>)
- DigiCULT (<http://www.digicult.info>)
- and FIRST (<http://www.film-first.org/first/>), all contributing to understanding and action.

Broadcast Sector Successes

The other major positive development has been the funding of large-scale preservation projects in the broadcast sector. Broadcast archives are privileged – their contents have high value and high levels of use, and they exist within an enterprise (broadcasting) that is converting all operations from analogue to digital. This situation provides the business case for large broadcasters to invest up to 10 million Euros per year in digitisation/preservation projects.

For example, the BBC (UK) has committed 80 million Euros of its own funding over a 10-year period. RAI (Italy) committed even more resources over a shorter period, when all its radio holdings were digitised in the last three years. RAI and INA (France) have also invested heavily in mass storage systems and the associated digitisation/preservation of video material. There are a dozen or more similar projects across Europe. These internal projects have developed cost-effective assembly-line approaches to AV preservation. The EC project PRESTO documented this work, extended the technology, and generally promoted the ‘preservation factory’ approach. It is the task of PrestoSpace to extend this knowledge and approach, to make cost-effective ‘preservation transfers’ available to

all AV media collections, not just broadcast archives.

SPECIFIC PROBLEMS IDENTIFIED AT THE USER REQUIREMENTS MEETING: THE THREE FS

Format

Eddie Goray of RTBF (<http://www4.rtb.be>) in Brussels reported on the analysis made by the EC project FIRST – a concerted action on the impact of digital technologies on film. Assuming film can be digitised at an acceptable technical standard, how are the digital data to be stored? He showed that high-quality digitisation of a single feature film could require 10 Terabytes (10^{12} bytes) of storage, equivalent to 2000 DVDs or 100 very-high-capacity datatapes. Also there is no agreed data format for high-resolution film digitisation. Therefore storage format and file format are substantial obstacles to the use of digital technology in film preservation. Mass storage also has problems for television programmes. Until now, the permanent record of a TV programme could be held on a single film or tape. Mass storage distributes data across an array of servers or tapes. Archives do not, in general, want to abandon all sense of a single physical entity corresponding to an archive ‘holding’.

Finding

There is also a problem knowing about the content of AV archives. Even for those with comprehensive catalogues, there is a problem knowing what other archives have. It is vital to have a comprehensive European approach to standards and to data sharing, to prevent waste on duplication of preservation activity – and to ensure archives can find, across Europe, the best original material. This problem is most important for commercial material such as cinema films and commercial music recordings. But there are also many examples of finding TV and radio material in archives other than those of the producing company – and, of course,

many radio and TV companies have gone out of business or been disrupted by war. So there was a recognised need for data standards and data sharing.

The host archive, Beeld en Geluid (Netherlands Institute of Sound and Vision, <http://www.beeldengeluid.nl>), has been leading work on AV metadata standards. They are currently reformatting their entire catalogue, to conform to the IFLA representation. They have also produced a comprehensive review and guide to metadata. This work shows what progress can be achieved, and will inform the PrestoSpace metadata activity.



Karin Westerink from Beeld en Geluid

Film

Two presentations were made specifically about the needs of film, by Thomas Christensen of the Danish Film Institute (<http://www.dfi.dk>) and Giovanna Fossati of the Netherlands Filmmuseum (<http://www.filmmuseum.nl/>). Thomas explained how a film archive without film loses all sense of meaning, and showed the steps they had taken in Denmark to ensure a future, for at least the next few hundred years, of film preserved on film. The main action was providing low-temperature and humidity storage – below freezing for the material which has already shown signs of chemical change. Film conservation falls outside the scope of PrestoSpace activity, which was assumed to concentrate on digital processes. Nevertheless, Giovanna showed that there were three areas where she hoped PrestoSpace would provide help

to their film collection:

- **Cost-effective preparation** of access copies: DVD provides new forms of film archive access for both public showings and for commercial sales. However, film needs to be digitised in order to produce a DVD. The PrestoSpace work in cost-effective digitisation is immediately relevant to DVD production by film archives.
- **Storage** for the digital data (digital intermediate format) that film archives create when performing digital restoration. Digital processing has been accepted as a main technology for film restoration and involves making huge digital files as a first step. Archives would like to keep these files, but face the problems mentioned above under Format, i.e. no simple and cheap media on which to store the data.
- **General information on storage**, covering DVD, mass storage for digital intermediates, and covering the evolution of technology and cost. Storage is in the midst of a revolution, where 'spinning discs' are becoming not only the high-performance method of storage but also the cheapest, and include novel methods for achieving security (such as arrays of storage modules) that are unavailable with past AV media. All archives, from paper-based to film-based, will have growing needs for digital storage as a parallel activity to conservation of their originals – and PrestoSpace can provide useful information on the evolution of digital storage.

Daniel Teruggi speaking at the meeting



IMMEDIATE ACTIONS

PrestoSpace has prepared a detailed questionnaire in order to determine the preservation status of European AV material, and to find out what collection holders are doing and plan to do in the future. Some archives received 'pre-view' copies of the questionnaire before the meeting, and some preliminary results were available to guide discussion. Participants were encouraged to complete their questionnaires, and to circulate them as widely as possible. PrestoSpace will collect results in May, and these results will determine the future course of the project. A simplified form of the questionnaire is online at <http://prestospace.org/questionnaire/index.php?sid=3>. All parties with an interest in AV material are encouraged to send in a response. In addition to identifying the requirements of archives, there is a specific section for companies that provide digitisation and restoration services to archives. These are the companies that can work with the PrestoSpace project to launch Preservation Factory services.

THE PRESTOSPACE PLAN:

Better Faster Cheaper

The basic idea of PrestoSpace is that the technical processes involved in AV preservation need to be improved, or precious cultural heritage will be unnecessarily lost. The desired improvement is to make the services higher in quality, faster, and still cheaper. The method is the factory approach: mass production, involving significant investment in automation. This investment is beyond the reach of a small AV collection, hence the need for PrestoSpace to work with the commercial 'facility house' sector. It was very gratifying that several such companies attended the Amsterdam meeting, including specialists in audio, video and film.

Structure of PrestoSpace

Even 'better faster cheaper' preservation is expensive. It would be a fantastic success if



PrestoSpace could reduce digital preservation cost to, on average, 50 Euros per hour of material. But where is a small archive – of say 10,000 hours – going to get the half-million Euros? PrestoSpace has a plan: the increase in funding comes from an increase in access. It is the prospect of new services, new audiences, new products that leads to an increase in grants, donations and sales, or any other new or enlarged funding. In order to provide new access, PrestoSpace has taken a comprehensive view of ‘the digital archive’: all the ingredients necessary to exchange deteriorating media on shelves for accessible and refreshable media on digital storage.

The project intends to provide deliverables (devices, software, reports and recommendations) for preservation processes and management.

Preservation: a fast and affordable data-cine, a contact-less playback tool for audio disks, an automated audio preservation tool, an automated video preservation tool, a manual tape condition assessment tool and an information system for preservation management.

Restoration: a restoration management tool, a defect analysis and description infrastructure, a set of high-level restoration algorithms, a disk-to-disk real-time restoration tool, a film restoration software tool.

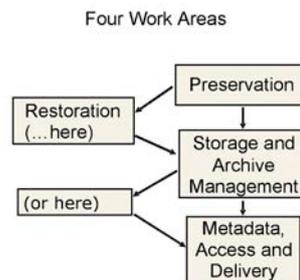
Storage and Archive Management: a Web-guide and software tool for storage planning for audiovisual preservation, a guide and software tool for business-case planning for audiovisual preservation and organisation, a logistics and quality assurance system for audiovisual preservation.

Metadata, Delivery and Access: a semi-automatic description tool, an export system for delivering preservation results to medium and large archives, a turnkey system for delivering preservation results to small archives.

For all this work, the first stage is clarification and ratification against genuine user requirements. The workshop and the

questionnaire are the method for determining those requirements.

The work of PrestoSpace is shown in the following diagram:



THE FILMS

The following morning we had a programme of films, including an early silent animation about the processes involved (80 years ago) in making and distributing film. One of the machines showed the name of a current PrestoSpace partner: Debrie, from Paris. There were also examples of film restored using digital and analogue processes, at projects in France, The Netherlands and Denmark.

CONCLUSION

It was a very packed day, representing a wide range of interests and views. There was agreement that:

- We all have preservation problems.
- Digital processing has an important role.
- Film is here to stay (providing the manufacture and processing of film can also be maintained).
- PrestoSpace can provide a significant, even vital, service if it listens properly to user requirements, and if it provides a structure for comprehensive European co-operation.

Finally, we all agreed the problems were larger than any one sector, viewpoint or institution. It was very encouraging to have such a breadth of experience and

interests gathered under one roof, especially such a distinguished roof. The strong co-operation between archives, research and the commercial services sector is the hope of PrestoSpace, and the hope for a brighter future for audiovisual preservation.

This reviewer would like to give special thanks to Brigit Hoomans and Karin Westerink of the Netherlands Institute of Sound and Vision, who were the principal organisers not only of the conference but also of the questionnaire and who put enormous effort into making this a productive meeting.

PRESTOSPACE CORE PARTNERS:

Institut National de l'Audiovisuel (INA), France (co-ordinator)
 British Broadcasting Corporation (BBC), UK
 Radiotelevisione Italiana (RAI), Italy
 Joanneum Research (JRS), Austria
 Netherlands Institute of Sound and Vision _ Beeld en Geluid (B&G), The Netherlands
 Oesterreichischer Rundfunk (ORF), Austria
 University of Sheffield, Computer Science Dept, UK



USING SPATIAL KNOWLEDGE TO CLASSIFY METADATA

ADAM RUSBRIDGE,
ERPANET TECHNICAL ANALYST

To date, the creation of metadata remains a bottleneck for many institutions collecting digital material. It is, therefore, natural that there are increasing levels of research in this area. The automated extraction of technical metadata is achievable for many file formats as this information is encoded into the specification of the format. Several tools are under development to cater for this task, for example, the National Library of New Zealand Preservation Metadata Extract Tool,⁵¹ OCLC's PREMIS (PREservation Metadata: Implementation Strategies)⁵² and the work of the Global Digital Format Registry.⁵³ While impossible to guarantee, it would not be surprising to see these tools achieve a high level of accuracy and reliability.

Extracting semantic metadata from the content of documents is a different matter. Existing file formats tend not to contain semantic markup and it remains a difficult task to distinguish between sections of a text-based document. Various techniques are under investigation. This article describes a technique for extracting semantic metadata from documents produced in the PostScript (PS) format.

PostScript is both a simple programming language and a page description format, designed to allow powerful graphics capabilities.⁵⁴ When text is stored in a PostScript file, it tends to be stored as plain text surrounded by method calls initialising the appropriate display configuration. This includes the xy page location where printing should begin, the font name, font size, and font widths. Printing a PostScript document requires an interpreter, often GhostScript (<http://www.cs.wisc.edu/~ghost/>), to translate this code into printable commands.

Several methods have been proposed to extract metadata from a document. A popular method of classifying documents is to employ statistical frequencies of words to categorise elements; however, this method is more appropriate for document summarisation tasks. An alternative is to use the spatial knowledge we have of documents to classify certain elements; for example, a title generally appears at the top of a page and is in a larger font size. This idea has been used to implement a metadata generation system at the US National Library of Medicine, and is discussed at length in 'Knowledge Based Metadata Extraction from PostScript Files' by G. Giuffrida *et al.*⁵⁵

The technique requires extraction of text from a document and associating information about the font, metrics, and xy location to each line. A rule set can be applied to these strings (a string is a complete horizontal row of text containing no line breaks) to produce increasingly accurate candidates for a particular element.

Open source PostScript to text converters already exist which redirect text from the printer to a text file. Pstotext (<http://www.cs.wisc.edu/~ghost/doc/pstotext.htm>) is arguably the most stable, although prescript (<http://www.nzdl.org/html/prescript.html>) is also of note. To output text, each of these contain small PostScript programs that override the output methods of a GhostScript processor and redirect the text. The pstotext program also utilises font and metrics information applied to each fragment of text to ensure the document is reconstructed correctly.

It is possible to extend this program to output text and its additional information at a string level. It is then possible to

apply rules to a set of strings determining additional implicit properties. Determining and refining such a rule collection allows the classification of items such as title, author, date of publication, abstract and table of contents. Such a rule set may require a document to contain a certain layout to be recognised, but as the rule set can be extended additional configurations can be interpreted.

A simplified rule set for title identification may find that the title:

1. is generally found on the first page;
2. precedes the abstract or introduction;
3. contains the largest text on the page.

Attractively, this rule identification system can be applied to multiple file formats, assuming the correct information can be output. For example, the Portable Document Format (PDF) is built upon the PostScript format and contains similar font and metrics information.⁵⁶ Additionally, this work can be combined with alternative metadata generation and document summarisation tools, allowing the time and technical requirements of ingest to be reduced and streamlined. As a collaborative technique, this is very promising for digital library systems and has potentially huge benefits for digital collections of all kinds.

⁵¹ This tool has been submitted for the Digital Preservation Award 2004, a new award described in 'Recognising Advances in Digital Preservation', also in this issue. For more information, see http://www.natlib.govt.nz/files/Project%20Description_v3-final.pdf.

⁵² <http://www.oclc.org/research/projects/pmwg/>

⁵³ <http://hul.harvard.edu/gdfr/>

⁵⁴ For PostScript specifications, see <http://partners.adobe.com/asn/tech/ps/specifications.jsp>.

⁵⁵ Article full text available at <http://citeseer.ist.psu.edu/385845.html>.

⁵⁶ For PDF specifications, see <http://partners.adobe.com/asn/tech/pdf/specifications.jsp>.



NEWS FROM AROUND THE GLOBE

DIGITAL RECORDKEEPING INITIATIVE

Australia and New Zealand's public record institutions are collaborating to establish a huge digital library. The Digital Recordkeeping Initiative will develop a set of rules for creation, use and storage of digital records, and it is hoped that it will become a benchmark for international standards.

For more information, see <http://www.computerworld.com.au/index.php/id;1613096940;fp;16;fpid;0>.

DIGITISATION BEGINS AT NATIONAL FILM ARCHIVES OF INDIA

A new centre is being built for the preservation of film at the start of the NFAI's (<http://www.nfaipune.nic.in/>) digitisation programme. The centre in Pune, India, will take steps to begin copying film reels and preserving those in danger. The archives intend to publish a catalogue of all new films on their Web site.

See the full story at <http://www.hinduonnet.com/thehindu/holnus/002200404211180.htm>.

RECOVERY OF NAMIBIA'S HISTORICAL MEMORIES

The Archives of the Anti-Colonial Resistance and the Liberation Struggle Project (AACRLS) held a meeting to discuss how best to identify, locate and secure oral and physical historical materials, and to store them in the National Archives. Their aims include repatriating records held by former colonial rulers.

For more information, visit <http://africa.com/stories/200404261126.html> or see the IFLA General Conference and Council paper at <http://www.ifla.org/IV/ifla69/papers/091e-Namhila.pdf>.

FREE ACCESS TO IMAGES OF SPACE EXPLORATION

NASA (<http://www.nasa.gov>) is seeking a partner to develop a comprehensive database of historical and educational material, including more than 115,000 film and video titles and millions of still images documenting the history of America's space programme. These materials will be digitised and published online for public research.

For more information, see <http://www.spaceref.com/news/viewpr.html?pid=14077> and http://www.usatoday.com/tech/news/2004-04-22-nasa-media-archive_x.htm.

MAPPING RECORDS USING A BEAM OF LIGHT

Researchers in California have discovered a way to read the grooves of fragile recordings using a beam of light. This method imposes no risk to the disks and can even reassemble data held on disks that are too damaged to play, or even broken into pieces.

The full story is available from http://www.pcpro.co.uk/?http://www.pcpro.co.uk/news/news_story.php?id=56778.

BLUE SHIELD

ICA has now begun presiding over the International Committee of the Blue Shield (ICBS), the cultural version of the Red Cross, which works to protect threatened cultural heritage all over the world.

For more information, see <http://www.ifla.org/blueshield.htm>.

DIGITISATION IN THE BRUNEI NATIONAL ARCHIVES

The restoration, reproduction and digitisation of thousands of photographs in the Brunei National Archives will begin soon, thanks to the signing of an agreement between the Brunei Museum's Department and Memorable Impressions Sendirian Berhad. The National Archives store thousands of Brunei's historical records and photos for research purposes. One of the records consists of more than twelve thousand photographs, many depicting epic moments in Brunei's history.

For more information, see <http://www.brunei-online.com/bb/tue/apr20h10.htm>.

ISLAMIC TEXTS UNDER THREAT

In Africa, thousands of Islamic texts are decaying due to lack of preservation. Scholars have already rescued tens of thousands of pages, placing them in safe storage, but an estimated three million manuscripts, of incalculable value to the understanding of Islam in West Africa, are owned privately and are at risk.

For the full story, visit <http://www.wtopnews.com/index.php?id=220&sid=189210>.

DIGITISING THE ELGIN MARBLES

The surviving fragments of the magnificent Parthenon sculptures known collectively as the Elgin Marbles are scattered across ten museums in eight different countries. However, thanks to a new digitisation project, the sculptures will soon be reunited in a virtual reconstruction. 3D scanning technologies will increase both access to, and interactions with, these artefacts.



For more detail, see the news story at <http://news.bbc.co.uk/1/hi/magazine/3486476.stm>.

ACCESS TO ARCHIVES IN CHINA

Until recently, archives in China were largely confidential, with even the archivists themselves having restricted access. However, the archives in Guangzhou, the capital of South China's Guangdong Province, will soon be publicly accessible. The city is the first to make nearly all official archives available and is constructing a new building, hoped to be completed by late May.

For more information, see http://news.xinhuanet.com/english/2004-04/19/content_1427008.htm.

LISTENING TO THE PAST, SPEAKING TO THE FUTURE

A report from the UK's Archives Task Force was published in March this year and is hoped to encourage modernisation of archives and bring innovation to the way they are accessed and used. More information is available from the Action on Archives section of the Museums, Libraries and Archives Council (MLA) Web site <http://www.mla.gov.uk/action/archives/atf.asp>.

HOLOCAUST MUSEUM OPENS IN HUNGARY

In April this year, the Holocaust Memorial Centre, which incorporates an old synagogue, exhibit halls and documentation archives, was opened in Budapest on the eve of the sixtieth anniversary of the start of the Holocaust in Hungary. This first museum presenting the different experiences of Hungarians in World War II is an important progression for the nation as, until the end of communism in 1990, little public discussion of the Holocaust was allowed.

For the full story, see http://seattlepi.nwsourc.com/national/apeurope_story.asp?category=1103&slug=Hungary%20Holocaust%20Memorial.

PAPER DISKS

A new storage medium has been announced – a disk capable of storing up to 25 Gb, and made of paper!

More information is available from <http://www.technewsworld.com/perl/story/hardware/33462.html>.

ROMAN ARCHAEOLOGICAL FINDS

A German archaeological site, first discovered by metal-detecting hobbyists, has yielded more than 250 tools and weapons used by Roman legionnaires.

More information is available from <http://news.scotsman.com/international.cfm?id=406222004>.

FILM RECOVERED FROM DELIBERATE CONCEALMENT

Film archivists in The Netherlands have discovered the only known copy of *Beyond The Rocks*, starring Rudolph Valentino and Gloria Swanson. The film, which is in excellent condition, had been part of a collection that was deliberately dispersed and mislabelled and is the only film in which these two stars appear together.

For more information, see <http://www.themediadrome.com/cgi-bin/newspro/fullnews.cgi?newsid1082387066,35246>,

NOTE TO EDITOR – the comma at the end of the URL above is necessary for the link to work.

ONE MILLION IMAGES

PictureAustralia (<http://www.pictureaustralia.org/>), a collaborative Internet-based image bank hosted by the National Library, has reached one million images. Users can search the online collections of 34 organisations in a single process through the PictureAustralia portal.

Visit the Web site for more details.

INVESTIGATION INTO CYBERINFRASTRUCTURE

A national commission on cyberinfrastructure in the humanities and social sciences has been set up by the American Council

of Learned Societies. Research and consultations will result in a report, to be published early in 2005.

More information can be found at <http://www.acls.org/cyberinfrastructure/cyber.htm>.

DEGRADATION OF CDS

Optical disks such as CDs and DVDs are not as long-lived as once thought, even if stored correctly. The thin layer of lacquer that protects the reflective aluminium layer beneath can be damaged by mishandling the disks, and if applied incorrectly by the manufacturers can allow air to penetrate, causing damaging oxidation of the aluminium layer.

For more information, see the news stories at <http://www.cnn.com/2004/TECH/ptech/05/06/disc.rot.ap/index.html> and <http://www.macobserver.com/article/2004/05/06.4.shtml>.

For quick tips on how to protect CDs and DVDs, visit http://seattlepi.nwsourc.com/business/aptech_story.asp?category=1700&slug=Disc%20Rot%20Glance.

DIGITAL DUNHUANG

An ongoing digitisation project will unite virtually treasures from Dunhuang (in Northwest China) that are currently scattered across museums all over the world. Cultural relics to be digitised include images of caves, murals and clay sculptures, as well as research and historical records relating to the region and artefacts. The Buddhist Mogao Grottoes, which made the city famous, were added to the UNESCO

World Heritage List (<http://whc.unesco.org/>) in 1987.

See the full story at http://news.xinhuanet.com/english/2004-05/26/content_1492300.htm.

The use of quantitative methods in the study of prehistorical archaeology began early in Italy, although their application to classical archaeology came later. This pattern is repeated around much of Europe. It was only relatively recently that these methods began to be used in later periods.

CHANGING TECHNOLOGIES CHANGING METHODS: AN INTERVIEW WITH PAOLA MOSCATI

PAOLA MOSCATI HAS WORKED WITH COMPUTER TECHNIQUES IN ARCHAEOLOGY SINCE SHE FINISHED UNIVERSITY AND IS THE EDITOR OF *ARCHEOLOGIA E CALCOLATORI*.
INTERVIEW BY DAISY ABBOTT.

As the use of computer applications became more widespread, scholars have not only begun to take advantage of the mathematical techniques themselves in the analysis of results, but have learned to use information technologies as an actual method of study in archaeology and cultural heritage. My study of Etruscan mirrors and funeral urns used statistical methods applied to these specific objects to classify them from a typological point of view. Studying the objects in terms of iconography also provided very interesting results. Samples were classified using multidimensional techniques such as cluster analysis and factor analysis. The application of statistics allows conclusions to be drawn connecting the materials used (bronze in the case of the mirrors and stone for the funeral urns) and the morphology of the objects. It was the growing use of new technologies that led me to write a book on the subject, *Archeologia e Calcolatori*, which was published in 1987 and became a very important step in my work in this field of study.

ARCHEOLOGIA E CALCOLATORI

As the use of computers in archaeology reached a high level at the end of the eighties, it became clear to me that there was a problem in developing this type of study. In Italy and across Europe there were many researchers who used computer applications to assist archaeological analysis, but there was a distinct lack of a unified editorial reference work. It seemed that the quickly developing technologies and their effects on research methods were often not

fully known by the people using them. It was decided that, to solve this problem, it would be useful to create a new journal, as a product of the National Research Council – *Istituto di Studi sulle Civiltà Italiane e del Mediterraneo Antico* (<http://soi.cnr.it/iscima/>) and the University of Siena – *Dipartimento di Archeologia e Storia delle Arti* (<http://www.archeo.unisi.it/>), which I still edit today. The purpose of the journal *Archeologia e Calcolatori* (<http://soi.cnr.it/archcalc/>) was to make activities in this field more widely known and also to present articles from other countries to an Italian audience and vice versa.

Issue 1 was published in 1990, creating the long-needed editorial point of reference for this work. Articles are published in many languages, providing a multilingual solution. Each issue ended with a bibliography to widen awareness of the scholarly material available, but this was stopped after the tenth issue, as the amount of available texts was increasingly growing and easier to find. Instead, we replaced the bibliographic section with a succinct summary of subject areas: three or four short, high-level articles each based on one specific topic (for example, GIS: Geographic Information Systems),⁵⁷ which include citations and important references for that topic. The journal is a unique publication as it includes the archaeological results as well as the project descriptions and the techniques applied. This is a main feature of the journal, as technology does affect the research methods to achieve results, but is not the main focus itself.



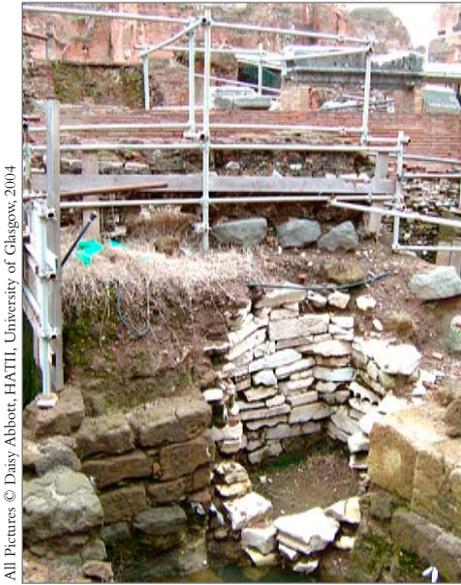
© Daisy Abbott, HATII, University of Glasgow, 2004

Paola Moscati

We publish special issues, for example, conference proceedings (for more information, see <http://soi.cnr.it/archcalc/news.htm>) or monographic issues dedicated to specific theoretical topics. Theoretical approaches are particularly important to the use of computers in archaeology and every five years a special issue is devoted to one topic. The fifteenth issue of *Archeologia e Calcolatori*, to be published later this year, will address 'New frontiers in archaeological research: Languages, Communication, Information Technology'. This topic binds together scientists devoted to the application of diverse computer methods to archaeological data from the phases of data acquisition and representation to their processing, interpretation and diffusion.

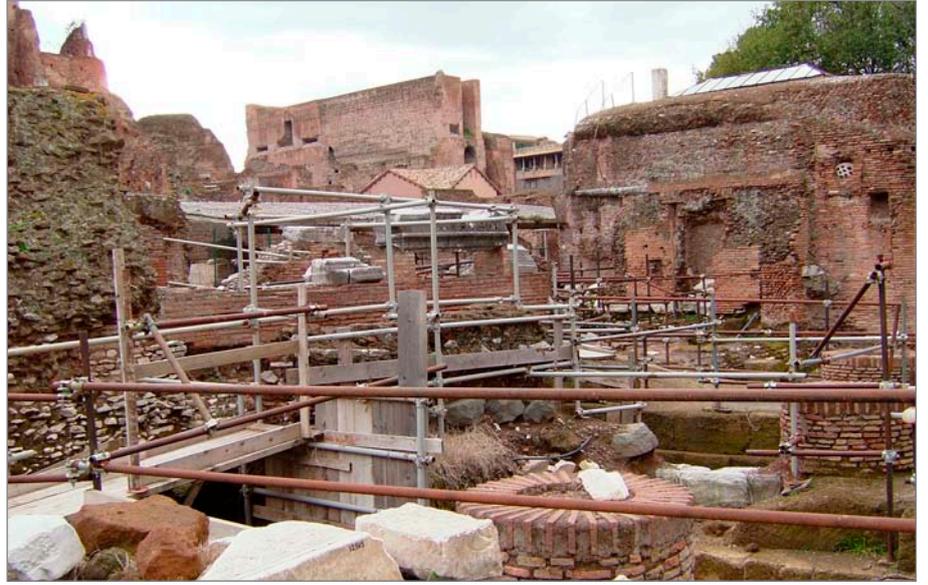
We have kept a database of applications and projects discovered over the years, which, after 20 years, we intend to publish as a retrospective on the field. The data are categorised by both archaeological subject and the computing techniques used, therefore data can be cross-referenced and correspondences between subject and appropriate technologies may be found.

⁵⁷ DigiCULT's third *Technology Watch Report* will devote a section to GIS in the heritage sector.



All Pictures © Daisy Abbott, HATIL, University of Glasgow, 2004

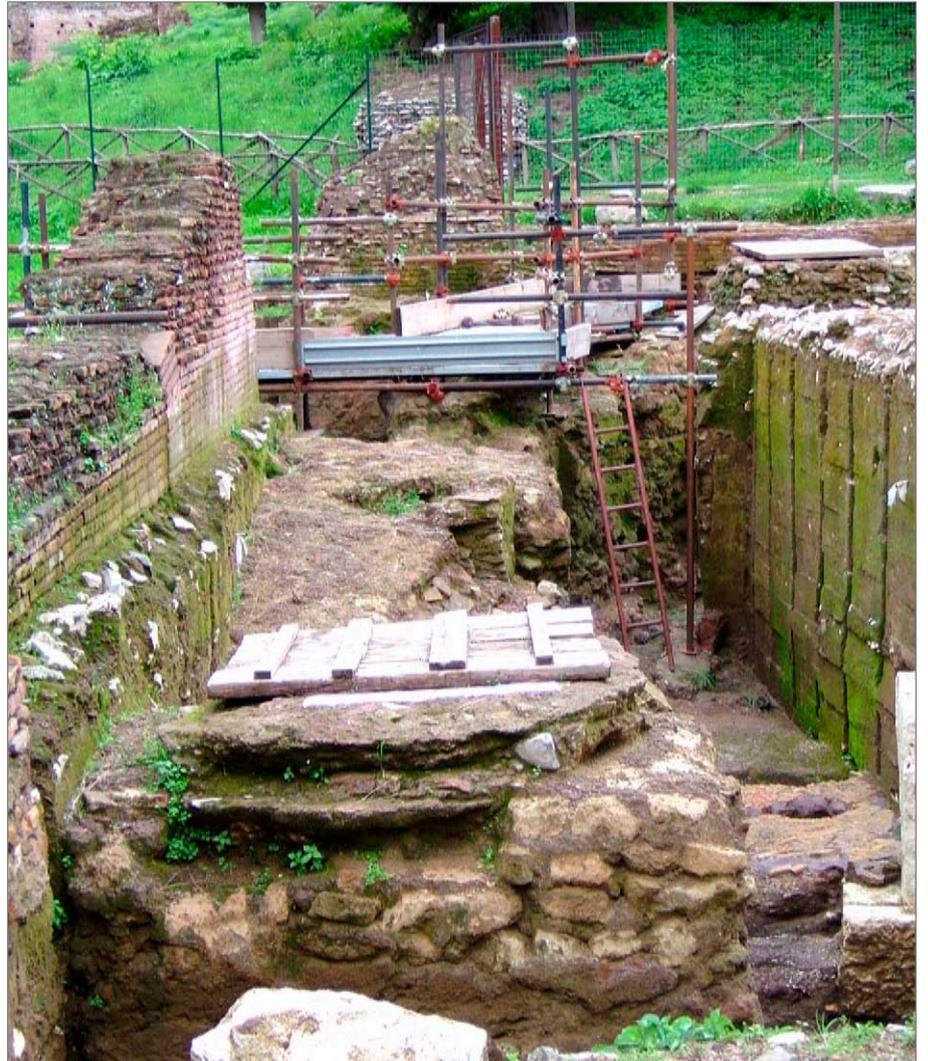
Current excavations at the Roman Forum

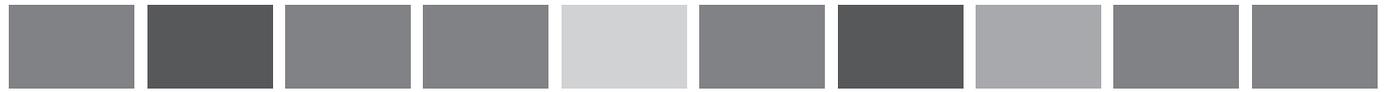


THE CAERE PROJECT

My more recent work has focused on new ways of representing information through text markup rather than databases. The Caere Project (<http://www.progettocaere.rm.cnr.it/>) was started in 1996 in order to use an archaeological information system to study an ancient Etruscan town. Excavations had not been performed in urban areas of the town until the 1980s and the data were recorded in databases and analysed with statistical methods. This year, the excavations have begun again; records have been computerised from the outset and newer GPS and georeference data have been used to complement former excavation diaries. The project integrates all our previous experience in computing archaeology: databases, graphics, digital cartography, GIS, statistical analysis and a new area, the electronic encoding of archaeological texts.

Excavation diaries of the 1980s contain precise information, therefore the differences between storing information in databases and encoding the texts directly became apparent. Databases are reductive; the fields must be pre-defined, whereas





textual encoding ensures no data are lost and tags can be designed as you go along. A DTD was designed for this project (initially in SGML and now in XML) to represent the structure and the content of the excavation diaries. The DTD was successfully integrated with the TEI Lite initiative (<http://www.tei-c.org/>) and led to experimentation with the markup of manuscripts, for example, 'Ruderi delle Ville Romano – Sabine nei dintorni di Poggio Mirteto' by Ercole Nardi (c. 1885) (see the 2003 issue of *Archeologia e Calcolatori* for an article on this manuscript). In an innovative development, GIS was used as a platform to integrate spatial data with the encoded texts, perhaps for the first time in archaeology, as in GIS spatial data are usually linked to databases. This approach has been used within the Caere project to study the distribution of finds in the urban area (see Issue 9 of *Archeologia e Calcolatori* for a collection of many international GIS projects).

Using technologies such as this in archaeology can change and develop the way the data and supporting documentation are interpreted.

CHANGING TECHNOLOGIES

Archeologia e Calcolatori provides a valuable overview of the changes in my area of study over the last fifteen years. Significant developments include the hardware change from mainframe to PCs, computer graphics and, of course, the Internet. From a methodological point of view, scholars now have access to many more facilities than they did in the early days of computing archaeology, and this could lead to an advance of technological approaches over methodological or theoretical standpoints. We must avoid the field becoming more based on technology than archaeology! Methods and results are the most important thing, not the technology used to achieve them, and often the 'wide' solutions offered by the application of some

technologies overlook the scientific part of discovery. For example, VR can provide a spectacular way of visualising data, but is not necessarily rigorous since information is not always made explicit. Another innovative system, GIS, acts as a container for older technologies – it is not the be all and end all of the archaeological study, but can integrate many forms of previous data in useful ways, if systems are designed correctly. Of course, in terms of influence on research methods and approaches, the Internet has had and will have the most significant effect on this field, not only in terms of resource awareness for more people, but in its role as the catalyst for the development of standard languages such as HTML and XML.

Use of IT has driven a change in the approach towards research, providing access to interpretations that are different from those that would otherwise have been noted – a change in the traditional approach that is extremely useful if a scientific orientation is maintained.

I believe that the future will see many more changes in the study of archaeology and the technologies used to achieve results. The primary change will continue to be in communication methods and the use of the Internet, together with the development of multimedia techniques for study and visualisation and more widespread standardisation of data representation. Dissemination methods will also change, with paper publications being used for one type of study while e-publications are exploited for others, and the e-publishing revolution will continue.

Mathematical and computer techniques have so much to offer to the study of any scientific field. It is essential that we understand how best to use them to maintain focus on the subjects at hand and fully exploit new technologies to improve our study methods and results.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM

The International Internet Preservation Consortium (<http://netpreserve.org>) was formed in 2003 to provide a focus for international collaboration on preserving online content. The key aims of the IIPC are:

- to encourage collaborative work in order to identify and develop solutions for the selection, collection, preservation and access of Internet content;
- to facilitate international coverage of Internet content archives;
- to provide international support for preservation and access initiatives.

The Consortium provides a forum for the exchange of expertise and aims to recommend appropriate standards for these collections, developing new standards where necessary. It will also work on the development of tools for the acquisition and archiving of Web sites, concentrating on interoperability and means of access. Raising awareness of Internet preservation issues is also a key goal and the IIPC will run conferences and training workshops as well as producing its own publications.

The Consortium comprises the national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden, The British Library (UK), The Library of Congress (USA) and the Internet Archive (USA). More information can be found on the IIPC Web site at <http://netpreserve.org>.

ERPANET/CODATA SEMINAR FINAL REPORT

The ERPANET/CODATA seminar on the Selection, Appraisal and Retention of Digital Scientific Data was held from 15-17 December 2003 at the Biblioteca Nacional in Lisbon, Portugal. Bringing together over sixty-five participants, the seminar examined the current state of practice in the selection, appraisal and retention of digital data generated from diverse scientific disciplines and through their collaboration. Participants discussed issues relating to changes in methods and interactions brought on by swiftly advancing technologies, and how archival concepts can best be applied to the management and long-term preservation of digital data. The seminar also highlighted some conceptual obstacles that must be overcome before effective collaboration between diverse communities can take place.

The final report produced as a result of this seminar, 'The Selection, Appraisal and Retention of Digital Scientific Data', is now available online from: <http://www.erpanet.org/www/products/lisbon/LisbonReportFinal.pdf>.

More information is available from the ERPANET (Electronic Resource Preservation and Access Network) and CODATA (Committee on Data for Science and Technology) Web sites, <http://www.erpanet.org/> and <http://www.codata.org/>.

NEW CULTURAL HERITAGE REPORTS

Reports presenting current thinking concerning electronic records and the recommendations of the International Council on Archives have recently been made available in the Publications section of the ICA Web site at <http://www.ica.org/>. The reports: *ICA Study 13-1:*

Authenticity of Electronic Records: A Report by ICA to UNESCO, 2002, and *ICA Study 13-2: Authenticity of Electronic Records: A Report Prepared for UNESCO and the International Council on Archives, 2004*, form part of the ICA Studies series.

The Digital Preservation Coalition (<http://www.dpconline.org/>) has published its first annual company report setting out its achievements between July 2002 and July 2003 against its aims. The DPC works to advance digital preservation in the UK and internationally. The report can be downloaded from <http://www.dpconline.org/graphics/reports/index.html#annual>.

New reports have also been published by the Council on Library and Information Resources (<http://www.clir.org/>). Of the many valuable reports available from their Web pages, two are of particular interest with regard to the commercial side of cultural heritage: 'Business Planning for Cultural Heritage Institutions' by Liz Bishoff and Nancy Allen (January 2004) and 'A Survey of Digital Cultural Heritage Initiatives and Their Sustainability Concerns' by Diane M. Zorich. (June 2003). All are freely available from <http://www.clir.org/pubs/reports/reports.html>.

The OCLC offers links to its most recent research presentations from <http://www.oclc.org/research/presentations/> and one particular author featured on this page, Brian Lavoie, has published two more papers relating to digital preservation: 'Of Mice and Memory: Economically Sustainable Preservation for the Twenty-first Century' in *Access in the Future Tense* from the Council on Library and Information Resources (available online from <http://www.clir.org/pubs/reports/pub126/lavoie.html>), and 'Implementing Metadata in Digital Preservation Systems: The PREMIS Activity' in *D-Lib Magazine*, vol. 10, no. 4 (<http://www.dlib.org/dlib/april04/lavoie/04lavoie.html>).

TRAINING OF THE FUTURE ARCHIVISTS AND RECORDS MANAGERS OF EUROPE: THE SITUATION IN THE UK AND IRELAND

ADELE REDHEAD,
GLASGOW UNIVERSITY ARCHIVE SERVICES

A NEW ARCHIVAL TRAINING COURSE AT UNIVERSITY OF GLASGOW

The recent increase in legislation covering records in any media and a number of high-profile scandals highlighting bad records-keeping practice have combined to lead an ever-increasing number of organisations to employ a professional archivist or records manager. This has in turn led to renewed interest in archives and records management as a career within the UK and Irish graduate community. Until this year, interested graduates in the UK and Ireland had the choice of full-time postgraduate courses for archivists and records managers in London, Liverpool, Aberystwyth and Dublin, and distance learning courses from Aberystwyth and Northumbria, but no course in Scotland, and no course that could be said to fully address the new challenges of the Digital Age. All of this has changed, with the University of Glasgow offering an MPhil in Information Management and Preservation (Digital)/(Archives & Records Management).

This exciting, forward-thinking course will provide post-graduate students with training that will give them the full range of skills to work as archivists and records managers in the twenty-first century. The course will specialise in the management of digital records and electronic resources, but will also give the students a



thorough grounding in the basic skills necessary to undertake archives and records management jobs, regardless of the medium. Modules include Archival Theory; Description, Cataloguing & Navigation; Archives, Records & Information Management; Records & Evidence; Records and the Transition to the Digital; Document Encoding; 2D Digitisation, and Management, Curation & Preservation of Digital Materials. All modules are available individually, and may be useful to archivists and records managers for continuing professional development purposes. The course may also be taken part-time over two years. For further details, see: <http://www.hatii.arts.gla.ac.uk/imp/>.

DIGITAL CURATION CENTRE

With increasing numbers of digital documents and datasets being generated by researchers, heritage professionals, and scientists, and a growing emphasis being placed on digital content, curating these examples of society's heritage is becoming a more and more demanding task. Digital documents are vulnerable to the rapidly changing market, where technology can become obsolete, as well as the inherent fragility of digital media. To address these issues, a Digital Curation Centre is to be set up in the UK to support institutions in the storage, management and preservation of digital data.

The Digital Curation Centre (<http://www.dcc.ac.uk>) is jointly managed by the Joint Information Systems Committee (<http://www.jisc.ac.uk>) and the e-Science Core Programme (<http://www.rcuk.ac.uk/escience/>). The DCC is made up of a consortium comprising the University of Glasgow's Humanities Advanced Technology and Information Institute (HATII) and its Information Services (<http://www.hatii.arts.gla.ac.uk/>), the University of Edinburgh (<http://www.edinburgh.ac.uk>), UKOLN at

the University of Bath (<http://www.ukoln.ac.uk/>) and the Council for the Central Laboratory of the Research Councils (<http://www.cclrc.ac.uk/>). Digital curation refers to data archiving and preservation, but also includes a wider spectrum of issues such as the life cycle of the digital document and its management.

The DCC's aims can be summarised as:

- To support practice and expertise in digital data curation and preservation
- To establish a vibrant research programme
- To ensure continuing access to data of scholarly interest
- To promote collaboration between universities and the Research Councils and other organisations or individuals within the community of practice.
- To develop evaluation services for tools, methods and standards, and address future issues such as e-learning and scholarly communication

The DCC is not a digital repository, but a means to unify themes to aid research and build a platform for collaboration.

Funding began in March this year (2004) and the formal launch of the Digital Curation Centre will take place in October. Immediate deliverables include a Web portal, an e-journal, an advisory service (the Helpdesk is already open and can be contacted on digitalcuration@ed.ac.uk), and outreach programmes.

More information and latest news can be found on the Digital Curation Centre Web site at <http://www.dcc.ac.uk>.