DigiCULT.Info

Issue 10 A Newsletter on Digital Culture October 2005 ISSN 1609-3941

CONTENTS

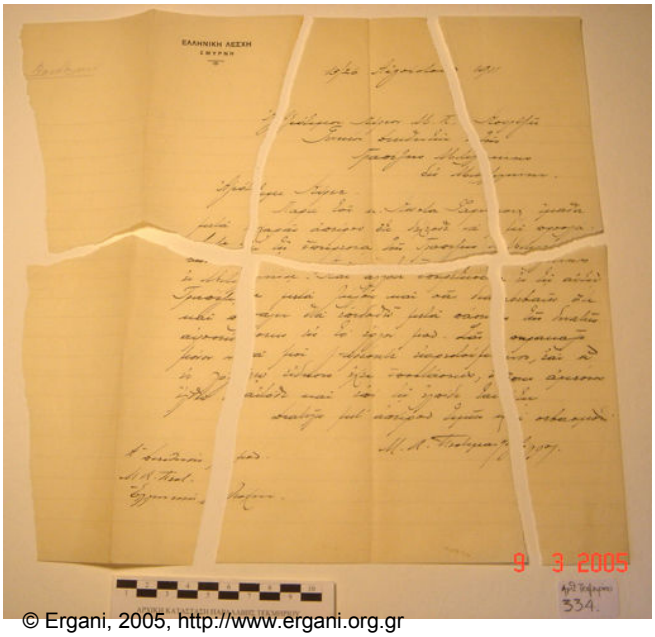
Contents 1
Introduction 1
The Prototype Process: Ideas turned into Experiences 2
LIFEPLUS Project: Real-time Virtual Humans in Augmented Reality Sites 6
News from DigiCULT's Regional Correspondents 10
- Albania 10
- Belgium 11
- Greece 12
- Lithuania 15
- The Netherlands 16

From Elements to Elephants: A Review of Progress in providing Online Access to the United Kingdom's Archival Information 17
Kosmopolis Digital Collection 20
Google Scholar: New Academic Research Tool. 21
Disposal and Archiving of Electronic Records: The DOMEA Concept in Germany 24
After ERPANET 25
Digitisation and Archiving from a Technological Perspective 26
InterPARES 2: A Progress Report 28
The BIRTH Television Archive: The Research and Delivery Portal for Audiovisual Heritage 32
NEMO: Network of European Museum Organisations 35
European Workshop on Culture and Technology: 36
2-D and 3-D Technology Trends for Culture 36
The VICODI Project: Creating a Semantic Web Application for Historians 38

INTRODUCTION

This tenth issue of *DigiCULT.Info* presents a wide variety of articles and information on the latest developments and projects within the cultural heritage field in 2005. With a discussion of the integrity and authenticity of digital records, covered in our first Newsletter and Thematic Issue, and an update of the InterPARES project, DigiCULT has come a full circle. Luciana Duranti, of InterPARES 2 presents general and case studies which identify the issues and challenges of creating digital versions of subjective works, and describes the development of a Metadata Schema Registry, enabling users to select appropriate methods for their own record-keeping.

Records and information management continues to be one of the core interests of the cultural heritage sector. Document Management and Electronic Archiving (DOMEA), based in Germany, presents the administration of electronic records and how they can be used to support business processes, with examples of federal schemes and policies on the eventual disposal of electronic records. A technological perspective on digitisation and archiving is presented by the Department of Information and Knowledge Management of Budapest University of Technology and



© Ergani, 2005, http://www.ergani.org.gr

Economics. This article details the Hungarian goals for national cultural heritage information research projects and summarises projects in this area.

The importance and value of prototyping for projects is an area not yet covered in detail by DigiCULT, and the candid interviews of our first article demonstrate how the Smithsonian American Art Museum used prototyping to gather information on how best to proceed with innovative developments to the access of art materials.

DigiCULT's Regional Correspondents from Albania, Belgium, Greece, Lithuania, and the Netherlands present a summary of work happening around Europe, including a focus on the new digital collection of periodicals made accessible by the University of Patras, Greece, and a fascinating insight into the Historical Archive of the Aegean Ergani which shows how the development of one multimedia collection, in this case the Kourtzi family archive, can demonstrate in microcosm the wider history of a nation from local, personal and private perspectives, enriching historical discourse and research. The Vicodi Project also takes an unconventional approach to scholarship in their development of a new visual contextualisation, interactive, semantic portal for document retrieval on the Internet. Focusing on visualisation and contextualisation, the VICODI team describe how they built a history ontology, what lessons they learned, and how this knowledge can be used in further development of the Semantic Web concept in the humanities. On a pan-European scale, the activities of NEMO (Network of European Museum Organisations) aim to increase co-operation between and promotion of museums of all sizes in Europe.

DigiCULT.Info does not neglect more innovative projects and two exciting articles present practical applications of recently developed technologies. In another Greek project, LIFEPLUS, the latest Augmented Reality technologies are being used to reconstruct and immerse users in an environment of ancient frescoes. The article also considers wider applications of developing these technologies in the arts and humanities, education, and entertainment fields. Johan Oomen, of the Netherlands Institute for Sound and Vision describes the BIRTH Television Archive (http://www.birth-of-tv.org): an innovative Web portal providing access to digitised audiovisual material. This project required co-operation from the archives of broadcasters from all over Europe, and the article presents the technologies and structures that were necessary to release the rich materials held in these archives into a wider arena.

I would like to take this opportunity to thank all of the fantastic contributors we have had over the years, in particular our hard-working Regional Correspondents, who have added so much to the scope, range, and quality of this newsletter. We hope to continue this collaboration well into the future. The DigiCULT events and resources databases are still active, and all DigiCULT materials (back issues of this newsletter, Thematic Issues, and Technology Watch Reports) are all available from our Web site at http://www.digicult.info. In addition, HATII (http://www.hatii.arts.gla.ac.uk/) will produce more value-added content from our previous material, such as Best Practice case studies. We hope that our materials have been, and will continue to be of use to you and your organisations. Enjoy the issue!

Daisy Abbott, Content Editor
Seamus Ross, Editor

THE PROTOTYPE PROCESS: IDEAS TURNED INTO EXPERIENCES

Angela T. Spinazzè

INTRODUCTION

DigiCULT has covered projects dealing with applications of mobile computing in the cultural heritage sector, both in previous issues of DigiCULT.Info and in our Technology Watch series.[1] In February 2005, Angela Spinazzè interviewed Claire Larkin and Michael P. Edson (respectively Special Projects Director and Chief, Information Technology Office) of the Smithsonian American Art Museum (http://www.americanart.si.edu). She asked them to talk about the Renwick Gallery Handheld Education Project, a two-and-a-half-year, two-phase research and development effort that involved the use of a wireless network and on-demand content delivery to visitors through a handheld device. The prototype project was so successful that the museum has decided to pursue the development of a full production system to be launched when the building re-opens in 2006.

BACKGROUND

In January 2000, the Smithsonian American Art Museum closed its doors for a major renovation and enhancement project.[2] In anticipation of building the Luce Foundation Center for American Art, an open-storage facility that will showcase over 3,400 collection objects as part of the new museum, the project team became interested in on-demand and dynamic delivery of information. Initial discussions led to the Renwick Gallery Handheld Education Project,[3] a pilot programme to test the concept. The project began as an investigation into how to deliver as much information as possible to the widest possible audience within the context of an open study/storage area. Museum staff wanted to analyse both the on-demand delivery of content, as well as find out what types of dynamic content visitors were most interested in. Three years later, the museum has committed to the installation of a wireless network throughout its main building and is on its way to a full-scale

[1] See "Handscape: Investigating Mobile Computing in Museums" in *DigiCULT.Info* Issue 3 and the chapter on mobile computing in *Technology Watch Report 2*, both available online at http://www.digicult.info.

[2] In 2001, the Smithsonian American Art Museum began an extensive $216 million (roughly €209m) renovation and enhancement project. The museum is scheduled to reopen on 4 July 2006, commemorating the 170th anniversary of the date that Congress approved construction of the building. More information is available from the museum's Web site: http://www.americanart.si.edu.

[3] The Renwick Handheld Education Project took place from 2002 to 2004. The prototype application was tested in two phases over an eleven-month period. Over 3,000 visitors participated in the testing process.

implementation of handheld devices as one of several access points into its increasingly sophisticated information repository.

How did you become interested in the use of handhelds in the museum?

Claire Larkin (CL): One of the special projects I am currently working on, the Luce Foundation Center for American Art, involves thousands of objects that have stories to tell as well as biographical information about the artists who made the objects. What visitors will see in the Center may not always make sense without additional information, description or discussion. We see our visitors as increasingly information hungry and communication savvy, and the use of handheld technology naturally came up within the context of what we wanted to do – serve up information to the broadest possible audience.

For example, in the Luce Foundation Center, where there will be thousands of objects that require descriptive information to help bring them to life, even a basic descriptive record requires a lot of work to develop – especially when you're talking about thousands of objects. We can prepare this descriptive information and provide access to it easily enough through a kiosk installed somewhere in the Center. We realised, however, that not everyone wants to access information about the objects in this way, that is, by remembering a roster of objects they have seen and then having to go somewhere else in the Center to look them up. There will be visitors who want instant gratification. A question pops into their head and they will want to look it up or have the option to discover the answer at that moment. This is where the idea of the handheld application began. Why not offer up the information through **both** the stationary kiosk and a mobile handheld device? This ties in directly with our mission to make information accessible to the broadest possible audience. This is what we decided to test.

Describe the purpose of the prototype. What were you trying to accomplish?

CL: The project started out as an investigation into how to deliver as much information as possible to the widest possible audience. We envisaged visitors picking up a device in the Luce Foundation Center and using it to get information as they perused the art objects in cases. We wanted to use the handheld device to serve up information about the objects in open storage as well as information about the artists who made the objects.

The Renwick Gallery is a branch of the Smithsonian American Art Museum, dedicated to American crafts from the nineteenth to twenty-first centuries. The exhibition galleries are housed in a historic architectural landmark that was completed in 1861. Selections from the Renwick's permanent collection of contemporary American crafts and decorative arts are displayed on a rotating basis in the second-floor galleries. Popular works include Larry Fuente's Game Fish, Wendell Castle's Ghost Clock, and Albert Paley's Portal Gates.

For more information, see the Renwick Gallery Web pages at:
http://www.americanart.si.edu/renwick/index.cfm

There were two other important features of the experience that we envisaged. The first was on-demand delivery of information. The second was dynamic information. We want the information to be delivered only when the visitor wants to access it. And, we did not want to limit what we could offer by pre-loading the content on to the device. The combination of these ideas was new at the time we started with the prototype project.

The overall purpose of the prototype was to prove that the concept worked well in a museum context. Our initial research indicated that there were no instances in which a museum installed a wireless network and served up information from a remote server to the handheld, on-demand. We really wanted to see if we could provide access to a database of information that people could tap into as they walked through the galleries. It seemed to us that no other museum had tried to develop an application and installation similar to what we were looking for, so we decided to prototype the scenario in order to learn from experience.

What questions did you set out to answer?

CL: We had very specific questions that involved an environment and how people moved through it and how they wanted to ask questions as they moved. For example, we wanted to gather information about how visitors wandered through the gallery spaces, we wanted to know what objects they were most interested in, and what type of information they were interested in viewing. It seemed that a wireless system that accessed an information repository housed on a server would be the best way to do that.

We wanted to learn about what type of content people liked. It is so time consuming to develop content. Early on in the process we were spending a lot of time developing content and one of the things we wanted to know was, 'Is this content that we're spending so much time and effort on the type of content that visitors using a device like this actually want?' We spent a lot of time crafting digital video and audio elements specifically for the handheld experience. We wanted to test the types of content, to determine their value to the visitor.

We wondered about offering more than one content element per object. Does it make sense for each of the content elements to be shorter in length than what is currently offered in the audio-only tour approach? We experimented with this in the pilot. Short audio, short video, a few still images, a short quiz. The scale of the full project involves over 3,000 objects, so we figured that we had better understand our visitors' content needs now, before we invest considerable resources to develop content on such a large scale.

We wanted to see if people liked the format of the content we were putting on the handheld. We wondered if the tiny screen was suitable for the type of content we were delivering. We were concerned about technology and different age groups, so we set out to see if it would have limited appeal or broad-based appeal. And, we wanted to test the technology too. It has changed considerably since we carried out our pilot project.

Can you talk about the sequence of activities?

CL: First, we defined the set of questions we wanted to answer. Second, we set out to find a developer who understood our needs and had an application that could be used in our environment for our specific project. The third step was development of both the content and the testing scenarios, and installation of the wireless network. Stage four was testing everything in the galleries.

That means that you went out looking for an application that already existed?

CL: Yes, we were not interested in developing the technology in-house. We wanted to develop the design, look and feel, and content, but we were not interested in a technology development project.

We issued a Request For Proposals in order to compare competitive bids. We outlined what it was we were looking for in more of a narrative way than a technical way. We suggested a timeline. We asked each respondent to describe their experiences working with the technology in a museum environment, and if they had not done that,

how they would approach delivering it given our scenario.

How did you decide where to install and test the handheld experience?

CL: We didn't have a lot of choice! The museum was closed at the time. The Renwick Gallery is a branch of the museum, in a separate building that remained open while our main museum building was being renovated. We were using the Renwick exhibition spaces and galleries to mount temporary exhibitions and maintain permanent collection galleries. Our first thought was to use the handheld in conjunction with a temporary exhibition, but then we realised that we could run the test longer if we focused on permanent collection installations.

Process is such a big part of contemporary craft, we thought we could illuminate the process, let people meet living artists, and reveal some of the hidden aspects of the objects. As it turned out, it really was a great collection to use for this project. The diversity of subject matter, materials, techniques – it was so useful to us for testing the various aspects of the handheld experience. It was easy to develop a variety of content types and formats that could not otherwise have been delivered in the gallery space.

Who was involved in the project?

CL: The team was comprised of museum staff and several specialists. We had museum staff work primarily on content development, and the look and feel of the presentation – the graphic elements, the colours, the flow between elements. The team always met in person; we made content decisions as a group. We spent a lot of time on the audio and video elements as a group. For example, everyone got a transcript of the film sessions. Team members would review them and then highlight sections they thought were of interest or that fitted with the objectives we had set out for the media clips. We would discuss them and then send the comments back to our editor/videographer (who, in this case, was an outside contractor). The videographer then edited the piece based on our comments. Sometimes he would create a second edited version based on what seemed to work well from his perspective. We had a very productive working relationship and developed a good trust and feel for the content as the project progressed.

Audio elements were recorded by an outside contractor, actors were chosen to read the elements and we ultimately chose the voice. Still images were selected by a member of staff familiar with the Renwick collection. Images were selected from existing stock where possible and then scanned to a variety of specifications to fit the handheld environment. Museum staff also coded the content, then handed it over to the application developer initially, so that we could see what it would look like. Templates were developed and, by the second phase of the project, we worked directly with a content management system that communicated with the handheld device.

What was the timeline for the project?

CL: It took us one year from the time we started talking about a handheld experience to the launch of the first test application. It took longer to find an application developer than we initially anticipated. In the end, we spent six months on this part of the process and that included reading the proposals, selecting an organisation to work with, preparing and agreeing a contract, and so on.

We spent four months creating content. It took us about the same time to install the wireless network. Given the age of the building, issues such as how to gain access to the attic, who would be allowed up into this part of the building, and finding someone to do the work after hours, required a lot of effort. This was all new territory for us to a certain extent. In the end, it took three months from start to finish to install the network.

We tested two applications. The first test ran for seven months and we used twenty handheld devices. The second test ran for four months and we used eighteen devices. We had about a month of downtime in between. The second application extended the first one. It was developed based on feedback and observations from the first test. It was more ambitious and more complex. Over the course of the eleven months of testing, approximately 3,000 people participated in our experiment.

Could you share some details about costs? What costs did you anticipate? What costs came as a surprise?

CL: Some of the costs are obvious, of course. We had to purchase equipment, the software application, and install a wireless network. We paid outside specialists to design and install the wireless network, to develop video clips, record audio elements, and work the desk (the people who actually handed out the devices to the public).

Our in-house team, with a lot of talent, included curators, staff from publications, from information technology, and graphic design. Their time and salaries were not figured into the overall project budget. I would say that the three most significant areas with regard to the project budget were the technology, the environment, and content development.

Two areas emerged as particularly challenging in terms of costs: the installation of the wireless network and staff who interacted directly with our visitors. As I mentioned before, the architecture of the building posed challenges with regard to how and where the network could be installed. I don't think we appreciated the complexities of this part of the process when we first started out. This issue won't be the same for us in the future because the project that is currently going on with the main building means that we can install the network as part of the renovation process rather than after it is complete. It is something to take into consideration, though, for other museums considering a test like this.

Finding the right type of person to interface with the public was important. We hired temporary staff to work the desk, to explain the project, to encourage people to participate, and to help people when something went wrong. As you can imagine, customer service is key, but also an understanding of the goals and objectives of the project were important so that the person communicating with the visitor could feel engaged in the process and compensated, so as to reflect the value of their contribution to the experience. Human resource issues are not always talked about, but it does make a big difference if there is someone there who is good with people.

What about the hardware? Were the devices robust enough to cope with the amount of use they were given?

CL: We didn't lose anything, and they were difficult to break. Equipment not connecting to the network was a bigger issue than hardware breaking down.

Michael Edson (ME): One of the biggest areas of understanding for us came within this context of what happens when the device loses its connection to the wireless network. In a way, it was helpful that these experiences happened within the context of the prototype so that we could learn just what was needed to avoid it in the larger implementation environment.

How were outcomes measured? And how do you plan to use what you learned to move ahead?

CL: We gathered feedback in several ways. There was a short survey on the handheld device that we asked visitors to complete. We conducted exit interviews, and recorded our observations from spending time in the gallery throughout the test period.

ME: I arrived at the museum when the Renwick Handheld Education Project was already under way and, as a fan of the spiral model of software development, the prototype was already a familiar process for me. In this model, the spiral is organised into four quadrants: brain storming/design; construction; publication and testing validation; and real evaluation of what you've accomplished and what you want to do in the subsequent spiral or activity loop. The idea is that each time you pass through an entire cycle of the spiral, you've also passed through all four quadrants.[4]

This model is helpful because it deals explicitly with risk. Also, it acknowledges that you don't know all the answers at the outset, but that through experience the answers will come. So, the greater the investment, the greater the experience, and the better the end-product. As the whole idea of piloting and prototyping is built in to this model, the Renwick project fitted in perfectly.

So, you're comfortable with the prototype process?

ME: Yes. Implicit in the handheld initiatives is that we're working with new technologies and with companies that are new entities too. The technology is new to the world, the companies are new to themselves and, with our museum community and through the spiral model, we're beginning to test relationships. An analogy would be to think of the prototype process as a set of short dates, a 'getting to know you' sort of approach to working both with new technologies – handheld devices and wireless networks – and new people. We needed to get to know the application developers, understand their working methods and philosophies, and one very good way of doing this is to agree to work together on a short-term, small-scale test of our environment and content and their technical expertise and application development skills.

We've now had two 'dates', one each with two different companies and we're planning for a third date that will be the transition from test to full-scale development in the environment we were initially planning for. We've learned something from each situation and we will carry that with us through each subsequent date. This series of short-term, low-exposure projects, as far upstream as possible, I see as fundamental to constructing the team and the approach we need for the longer-term implementation.

Are you pleased with the success rate for each of the steps thus far?

ME: Yes, each iteration has moved us forward in our understanding of both what can and can't be done. I've been grateful for the failures we've had. On a few key issues we really learned a lot. Some of our initial assumptions were wrong and testing identified them to us. In some cases we bought technology that didn't fulfil our needs. These were fantastic mistakes to make two-and-a-half years before opening. If you're going to make mistakes, which we all do, it is wise to do so with some time on your side.

There is something else that we are learning about here too, and that is our level of institutional maturity. I am thinking in terms of the Capability Maturity Model for Software, or CMM for short.[5] This model is used by organisations to identify best practices and to help them assess their own maturity with regard to information technology. The analogy to our situation is that three years ago we didn't know what we didn't know about handheld technology, on-demand and/or dynamic content delivery. We couldn't gauge whether we could implement a handheld application in connection with a wireless network three years ago. Part of this spiral model approach and our investment of smaller sums of money, anywhere between $10,000 and $30,000, has been to gauge our own operational maturity.

How ready are we as an organisation to take on risky projects? For me, this is a very important question. Are we a mature enough organisation to take the next step with the wireless handheld initiatives? It is a big jump from where we currently are in the spiral to where we want to go in the next step (which is a full production mode). Leaving low-risk land to move into a place where we invest substantially more money and expose ourselves to higher risk as we broaden our approach to include more gallery space and more people requires us to be more mature.

In terms of the work we need to do, there are simultaneous activities. While we move ahead with the technological infrastructure to support the use of handheld devices in the new building, we must also move ahead with the construction of a complex meta-structure to support communication between the devices and our information resources. To build and sustain rich content on the handheld side, we want to create a series of well-defined repositories for data and have the handheld communicate with them. While we want to serve up data to the handheld, we don't want to change core practices within the museum.

How are you able to evolve from the Luce Center original idea towards a more pervasive approach to using handhelds in the museum?

ME: The idea of wireless connectivity in the new building was acknowledged as something we ought to do from the outset, but we didn't really know what that meant. Having participated in the prototype project, our understanding of what is and is not possible has grown and we've developed some terrific partnerships that have enabled us to craft a vision for dynamic delivery of information throughout the building using handheld devices. We now know where we want to go and we're trying to realise this vision. Over time, we will allow the handheld application to interact with many of the other information resources we have available and customise those interactions based on a visitor's needs and wants. We want to build a narrative partnership with the public and have our information systems and appliances know what resources are available.

What advice can you give other museum professionals interested in, but perhaps hesitant to try, prototyping?

CL: Think about how the project will affect others in the museum. For example, we needed people with technical skills to support our project. Our information technology staff have a lot of responsibility and we added a lot more with our prototype project. Our internal technical team did a great job, but we did take them away from their regular duties during our project. Also, I mentioned before the temporary staff we brought in to work the desk to hand out the

[4] The spiral model of software development was created by Barry Boehm. The model was influential because it endorsed an iterative approach to the development process. Further details can be found on Wikipedia: http://en.wikipedia.org/wiki/Barry_Boehm.

[5] The Capability Maturity Model for Software (CMM) was developed by the Carnegie Mellon Software Engineering Institute. For more information visit the SEI Web site at: http://www.sei.cmu.edu/cmm/.

devices. I personally spent a lot of time working with people, sitting at the desk myself, to make sure that the experience of the experiment was communicated as much as anything else. I wanted visitors to know why we were testing the appliance, why we were interested in their feedback, and how much we appreciated their time and effort.

You will be amazed by what you observe. I noticed many times that the handheld device got visitors very excited about the objects they were looking at. I almost wish that we could have had docents on hand as well to continue the experience. What we found was that people wanted to ask more questions or dig deeper as a result of having used the handheld.

THAT'S INTERESTING – YOU'RE SUGGESTING THAT IT WAS THE USE OF THE TECHNOLOGY DEVICE THAT GOT VISITORS TO WANT TO TALK TO SOMEONE ELSE?

CL: Yes. In fact, in many instances I observed just that. It was really interesting to learn this.

BRIEF BIOGRAPHIES

Claire Larkin is Special Projects Director at the Smithsonian American Art Museum. She is responsible for two long-term projects that are part of the renovation and enhancement of the museum. The Luce Foundation Center for American Art is a 12,000 square foot open storage facility that will display more than 3,400 art objects from the Museum's collection, and the Lunder Conservation Center will be a 10,200 square foot facility featuring five visible laboratories and studios. Both Centers will be open to the public seven days a week, year round.

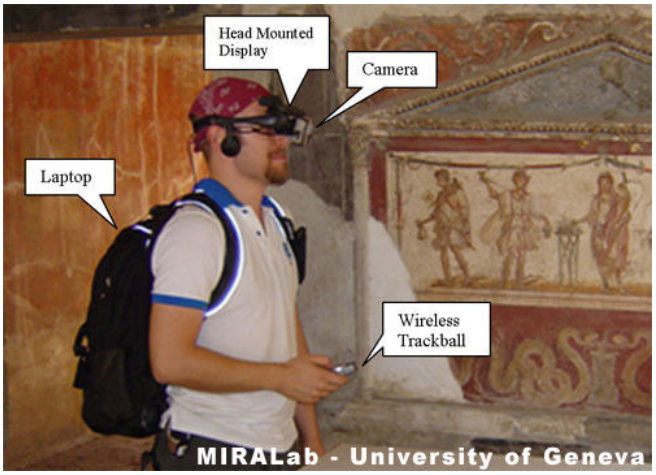
Michael P. Edson is Chief of the Information Technology Office at the Smithsonian American Art Museum. The IT Office includes New Media Initiatives, Computer Support and Data Integration. The office supports all hardware and software systems in the museum and conducts all new media programmes, both for the general public and for classrooms across the country. The IT Office also develops database applications, the Museum's Web site and related information management initiatives.

Angela T. Spinazzè is a consultant and independent researcher. Her experience includes feasibility studies and strategic planning for new technologies, the design of metadata schemas and associated standards, professional development workshops and seminars, project management and research. She works with museums to integrate new technologies into cultural environments. Find out more at http://www.atspin.com.

LIFEPLUS PROJECT: REAL-TIME VIRTUAL HUMANS IN AUGMENTED REALITY SITES

Nadia Magnenat-Thalmann, MIRALab, CUI, University of Geneva
Panos Trahanias, Institute of Computer Science, FORTH, Heraklion, Greece

In this article we are presenting our work in the EU Information Society Technologies (IST) project LIFEPLUS (http://www.miralab.unige.ch/subpages/lifeplus/). The project proposes an innovative 3D reconstruction of ancient frescos through the real-time revival of their fauna and flora, featuring groups of virtual animated characters with dramaturgical behaviour in an immersive Augmented Reality (AR) environment. The goal of this project is to push the limits of current AR technologies, exploring the processes of narrative design of fictional spaces where users can experience a high degree of realistic interactive immersion. Based on a captured/real-time video of a real scene, the project aims to enhance such scenes by enabling realistic 3D simulations of virtual characters to be rendered in real-time. Although initially targeted at cultural heritage centres and sites, the paradigm is by no means limited to such subjects but encompasses all types of future location-based entertainment, e-visitor attractions, e-worker training schemes, as well as on-set visualisations for the television and film industry. This article provides an overview of the project and the technologies employed, and presents the final results based on completed research and development. Professor Nadia Magnenat-Thalmann fulfilled the role of scientific coordinator of LIFEPLUS and Professor Panos Trahanias undertook the administrative coordination.



Mobile AR-Life simulator system © LIFEPLUS, 2004

INTRODUCTION

Since ancient times, images have been used as records of information, events and lifestyles as well as for decorative purposes. The possibility of reviving historical images adds a new dimension to the understanding of our past. The re-creation of historic environments for serious study, education and entertainment is not new,[6] although methods for achieving these objectives have evolved considerably over time. Before the days of widespread books and printing, story-tellers would conjure up visions of events and places, providing their listeners with an impression of reality (often augmented reality) elsewhere in time and

[6] See D.B. Arnold, "Computer Graphics and Archaeology: Realism and Symbiosis" in *ACM SIGGRAPH and EUROGRAPHICS: Interpreting the Past, preconference proceedings*, pp. 10-21, 2000.

space. Theatre, fine art and cinema have added to the richness of the explicit visual experience available to the viewer. They have made interpretations of history more accessible to the general public, but at the same time have narrowed the individual's scope for personalised, interactive experience and visualisation of the description of it. Historical frescos are a unique arrangement of *mise en scène* (stage-setting) elements that enhance the user experience by creating a set of compelling narrative patterns, albeit in a static, two-dimensional way. Mixed realities[7] and their concept of cyber-real space interplay invoke interactive digital narratives that promote new patterns of understanding in various contexts. In the context of cultural heritage sites such as the ancient city of Pompeii, we would like to observe and understand the behaviours and social patterns of living people from ancient Roman times, superimposed on to the natural environment of the city. In industrial environments, we would like to 'employ' virtual workers, for training and maintenance reasons, to assist real ones. Until recently, AR systems had various problems in managing such simulations in a fully interactive way, due to hardware and software complexities in AR enabling technologies.[8] Generally the set-up of such systems was operational only in specific places (indoors or outdoors) or with specific objects that were used for training purposes, making them difficult to apply at different sites. Furthermore, almost none of these systems features full real-time virtual human simulation. With our approach, based on two efficient real-time camera tracking systems, we can quickly set up AR experiences anywhere. With the interplay of a modern real-time framework for integrated interactive virtual character simulation, we illustrate how we enhance the experience in LIFEPLUS with full virtual character simulations at cultural heritage sites.

Real-time virtual Pompeian character simulation in the Pompeian thermopolium © LIFEPLUS, 2004

RELATED WORK

A number of projects are currently based on AR integrated platforms, exploring a variety of applications in different domains such as medical, e.g. ART (Augmented Reality for Therapy, http://mrcas.mpe.ntu.edu.sg/groups/art/); cultural heritage, e.g. the work of Stricker *et al*.[9] and Papagiannakis *et al*.;[10] training and maintenance, e.g. Schwald *et al.*[11] and Wohlgemuth and Triebfürst;[12] and games, for example, Thomas *et al*.[13] Special focus has recently been applied to system design and architecture in order to provide the various AR enabling technologies with a framework for proper collaboration and interplay.[14] An extensive bibliography on current state-of-the-art AR systems and frameworks has been described.[15] However, few of these systems take the modern approach that a realistic mixed reality application, rich in AR virtual character experiences, should be based on a complete VR Framework (featuring game-engine-like components) with the addition of the 'AR enabling technologies' such as Real-time Camera Tracking, AR Displays and Interfaces, and Registration and Calibration.

AR FRAMEWORK COMPONENTS

Our AR-Life platform is based on the VHD++[16] component-based framework engine developed by VRLAB-EPFL (http://ligwww.epfl.ch/) and MIRALab-UNIGE (http://www.miralab.unige.ch/), which allows quick prototyping of VR-AR applications featuring integrated real-time virtual character simulation technologies. The framework has borrowed extensive know-how from previous platforms.[17] The

[7] For more on mixed realities, see P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays" in *IEICE Transactions on Information Systems*, Vol. E77-D, No. 12, pp. 1321-1329, 1994.
[8] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre, "Recent Advances in Augmented Reality" in *IEEE Computer Graphics Applications*, Vol. 21, No. 6, pp. 34-47, 2001.
[9] D. Stricker, P. Dähne, F. Seibert, I. Christou, L. Almeida and N. Ioannidis, "Design and Development Issues for Archeoguide: An Augmented Reality based Cultural Heritage On-Site Guide" in *Proceedings ICAV3D 2001: Augmented Virtual Environments and 3D Imaging,* pp. 1-5, 2001.
[10] G. Papagiannakis, M. Ponder, T. Molet, S. Kshirsagar, F. Cordier, N. Magnenat-Thalmann and D. Thalmann, "LIFEPLUS: Revival of life in ancient Pompeii", *VSMM2002* (invited paper), 2002.
[11] B. Schwald, J. Figue, E. Chauvineau and F. Vu-Hong, "STARMATE: Using Augmented Reality technology for computer guided maintenance of complex mechanical elements", in *Proceedings e2001: eBusiness and eWork,* Vol. 1, Section 1.4, 2001.
[12] W. Wohlgemuth and G. Triebfürst, "ARVIKA: augmented reality for development, production and service" in *Proceedings DARE 2000: Designing augmented reality environments*, pp. 151-152, 2000.
[13] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris and W. Piekarski, "ARQuake: An Outdoor/Indoor Augmented Reality First Person Application" in *Proceedings ISWC 2000: Wearable Computers,* pp. 139-146, 2000.
[14] For more on this, see E. Gamma, R. Helm, R. Johnson and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software* (Addison-Wesley, 1994).
[15] See R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre, "Recent Advances in Augmented Reality" in *IEEE Computer Graphics Applications,* Vol. 21, No. 6, pp. 34-47, 2001.
[16] For more technical information, see M. Ponder, G. Papagiannakis, T. Molet, N. Magnenat-Thalmann and D. Thalmann, "VHD++ Development Framework: Towards Extendible, Component Based VR/AR Simulation Engine Featuring Advanced Virtual Character Technologies" in *IEEE Computer Society Press, CGI Proceedings*, pp. 96-104, 2003.
[17] Such as those presented in G. Sannier, S. Balcisoy, N. Magnenat-Thalmann and D. Thalmann, "VHD: A System for Directing Real-
(cont…)

key innovation is focused in the area of a component-based framework, which allows the plug-and-play of different heterogeneous human simulation technologies, such as real-time character rendering in AR (supporting real-virtual occlusions), real-time camera tracking, facial simulation and speech, body animation with skinning,[18] 3D sound, cloth simulation and behavioural scripting of actions. To meet the hardware requirements of this objective, a single Alienware P4 Area-51 Mobile Workstation was used, with a GeforceFX5600Go NVIDIA graphics card, a firewire Unibrain Camera or USB Logitech Web camera for fast image acquisition in a video-see-through i-glasses SVGAPro monoscopic HMD set-up, for advanced immersive simulation. Our previous efforts were based on a client-server distributed model, based on two mobile workstations. To achieve the requirement of 'true mobility', a single mobile workstation is used in our current demonstrations, after improvements in the streaming image capture and introduction of hyper-threading and graphics processing unit (GPU) calculations in the platform code.

The Pompeian 'thermopolium' © LIFEPLUS, 2004

Real-time markerless camera tracking

Real-time markerless camera tracking presents two main problems, namely the absence of easily recognisable markers and a demand for high-speed computation. Using markerless tracking is often a necessity, as applying markers within the tracking area is neither suitable nor possible (especially on cultural heritage sites). This then requires accurate tracking to be done with only natural features present, often with sharply varying light sources, shadows, motion blur and occlusions. Performing this tracking in real-time necessitates the use of algorithms specially adapted to fast operation, and thus disqualifies many algorithms that are perfectly suitable in offline applications.

The LIFEPLUS application utilises the 2D3 camera tracking solution (creators of the Boujou Software Tracker, see http://www.2d3.com/ for details), based on the approach that the system should be able to self-initialise anywhere within the tracking environment without any intervention by the user. In effect this means that, instead of calculating *relative* changes in rotation and translation (i.e. relative to set markers), we calculate *absolute* rotation and translation for every frame. This has the advantage of avoiding the problem of drift, and also ensures instant recovery if tracking is lost due to excessive motion blur or occlusion. Another innovative camera tracking solution has been developed within LIFEPLUS by FORTH (Foundation for Research and Technology: http://www.forth.gr/)[19] and was integrated in the AR-Life simulator, yielding interactive results.

Experiments and results

Cultural Heritage AR Simulation: Pompeii and the thermopolium of Vetutius Placidus

In order to confirm that our integrated AR framework for virtual character simulation operates in different environments, we tested the system directly in the ruins of Pompeii. However, in order to continue tests, a paper 'maquette' was constructed to represent the actual Pompeii site that we visited for our first on-site tests. This allowed extra fine-tuning and improvement of our simulation and framework, without having to visit the real site numerous times. With the help of the Archaeological Superintendence of Pompeii (http://www.pompeiisites.org), who provided us with all the necessary archaeological and historical information, we selected the 'thermopolium' (tavern) of Vetutius Placidus and conducted our experiments there.

Some video animations of the work of LIFEPLUS are available online at http://netzspannung.org/database/lifeplus/en .

Conclusions and future work

With the current results of our AR Framework, we are able to manage augmented reality, full virtual character simulations (including body, skin, face and clothes) in real-time cultural heritage environments through a markerless AR tracking system. However, there is still room for improvement. First, we are working on improving the real-time camera tracking, in order to achieve higher frame rates in both methods. The 'illumination' registration between the real and virtual scene is being addressed with the introduction of High Dynamic Range Image Based Lighting for virtual character simulations in AR. Finally, the potential incorporation of various interaction paradigms between the real user and the virtual humans in the augmented world is currently being investigated.

Lab 'maquette' AR tests © LIFEPLUS, 2004

Time Virtual Actors" in *The Visual Computer*, ed. Springer, Vol. 15, Nos 7/8, pp.320-329, 1999.

[18] For more on 'skin' technologies, see the chapter on Cultural Agents, Avatars and Personalisation in DigiCULT's *Technology Watch Report 3*, available free online from http://www.digicult.info/pages/techwatch.php.

[19] See also M. Lourakis, "Efficient 3D Camera Matchmoving Using Markerless, Segmentation-Free Plane Tracking" in *Technical Report 324*, Institute of Computer Science, FORTH, 2003.

ACKNOWLEDGEMENTS

Special thanks are extended to Sebastian Schertenleib, Michal Ponder and Tolga Abaci for the VHD++ framework and to Marlene Arevalo and Nedjma Cadi for their design work. The work described was supported by the Swiss Federal Office for Education and Science and the EU IST FP5 programme, in the context of the EU IST LIFEPLUS (15) project.

GOOD PRACTICE GUIDE

Resources for good practice, based on the experience and understanding developed through the UK's NOF-digitise programme, are available at http://www.ukoln.ac.uk/interop-focus/gpg/, supported by BIG (the successor organisation to the New Opportunities Fund, NOF) and the MLA (Council for Museums, Libraries and Archives, http://www.mla.gov.uk/).

The NOF-digitise programme (http://www.nof.org.uk/) provided lottery funding to support digitisation activities and the development of 150 Web sites offering access to cultural heritage resources. These resources are now available from the EnrichUK portal: http://www.enrichuk.net/.

UKOLN and AHDS (http://www.ukoln.ac.uk/, http://www.ahds.ac.uk/) provided technical support for projects through the NOF-digitise Technical Advisory Service (NOF-TAS), working closely with MLA who provided expert advice to the Big Lottery Fund. The wide range of materials and experience gained by NOF-TAS is extremely valuable to the sector and will be of interest to readers working in heritage organisations and projects. To ensure access to this knowledge resource, UKOLN have launched the Good Practice Guide Web site, which offers a range of advice to organisations involved in similar development activities. This resource will be complemented by others based on experiences gained in supporting similar work within the cultural heritage and higher and further education sectors.

LINKING ARMS

National and regional institutions and strategic bodies from across the British Isles have formed a consortium to propose a project that will 'enhance and enrich the United Kingdom's archival heritage; make a core element of the archival gateway available online;[1] and empower new and existing archives users to access their heritage through training and outreach programmes.'

Linking Arms aims to create new digital resources as well as making many existing online materials and repositories accessible through a single portal. MLA (http://www.mla.gov.uk/) has conducted tests with archive users to determine how the project might best proceed. Subject to receiving funding, the project will be carried out between 2006 and 2009.

CONSERVATION AWARDS

The UK Conservation Awards Programme (http://www.consawards.ukic.org.uk/), which rewards excellence and innovation in the field of preservation of cultural heritage, has had its funding secured until 2009, thanks to support from Sir Paul McCartney. 'Our cultural heritage gives enormous pleasure to people of all ages and reminds us of who we are and where we have come from,' said the Director of the UK Institute for Conservation (UKIC, http://www.ukic.org.uk/) David Leigh. 'In our throw-away society, it has never been more important to preserve "the real thing", and that's what conservators do. The Awards have recognised the triumphs of conservation over the past decade and [this] support will help us celebrate many more.'

The Conservation Awards are open to projects from any heritage field, from institutions of any size, and the Digital Preservation Award, which debuted in 2004, attracted entries from all over the world. The 2004 Award for Conservation was won by the Hamilton Kerr Institute (http://www-hki.fitzmuseum.cam.ac.uk/) and the Digital Preservation Award went to the UK National Archives (http://www.nationalarchives.gov.uk).

In 2005, the UKIC and the Institute of Paper Conservation (http://www.ipc.org.uk/) will be incorporated into the Institute of Conservation, a larger professional body.

PAPERS FROM MUSEUMS AND THE WEB 2005

Papers from Museums and the Web 2005, which took place in Vancouver, British Columbia, Canada, between 13 and 16 April, are available online from http://www.archimuse.com/mw2005/. This includes the closing paper 'Converging Culture' by Ian Wilson (Librarian and Archivist of Canada), http://www.archimuse.com/mw2005/abstracts/prg_295000686.html, which addresses possible convergence of cultural repositories and programmes and explores future challenges.

News from DigiCULT's Regional Correspondents

Albania

Report by Bendis Balla

National Centre for the Cultural Property Inventory

The National Centre for the Cultural Property Inventory (NCCPI) was established in 1995 upon a decision of the Albanian Council of the Ministers and with initial funding from the World Bank. The Centre is part of the Ministry of Culture, Youth and Sports of the Republic of Albania (MCYS). Its basic mission is the creation of a computerised catalogue on the cultural heritage of the Albanian people by managing the registration of all cultural real and movable properties that are in the custody of the network of museums throughout the country, art galleries, private citizens, and local and national specialised and religious institutions.

History

The work started without the benefits of any real basis in computerising and cataloguing Albania's cultural heritage objects (only the central specialised institutions such as those under the Academy of Sciences held some simple 'passports' of objects).
Consequently the period from 1995 to 1999 served as an experimental phase without producing significant results but laying important groundwork for future work. The existing infrastructure at that time was very simple and not appropriate for remote use. Similarly, the software used, *Histbase*, was not appropriate for the searching and exploitation of data. Efforts were made to approach donors who could provide support in these fields (such as UNESCO (http://www.unesco.org) or the Unique Catalogue Italian Institute), but these efforts did not bring any concrete results.

The project *NCCPI 99/01 – Automatic System for the Management of the Cultural Heritage* was drafted in 1999. In 2000, the State Budget supported the initial phase for the replacement of the former material base as well as the creation of a national computerised network for the management of cultural heritage. Following this breakthrough, a methodological guide for unifying the criteria for classification, terminology and understanding for the disciplines and categories included in the database was distributed to the relevant national and local institutions. The new software compiled by the Centre was tested. New software (named *Onufri*, after the well-known Albanian painter of the Medieval era) was designed by Infosoft and tested. (Infosoft still offers its services on potential problems related to the updating or maintenance of the database.)

In 2002 the database was imported into a new one that is still in use today. This database contains an index that largely follows the Council of Europe Core Data Index.[20] The provision of the computerised infrastructure for the Centre and the seven regional branches was also completed in the same year, aiming to create a stable connection and fast information exchange.

Legislation

The new law on cultural heritage, approved by the Albanian Parliament in April 2003, extended the duties of the NCCPI regarding the control of movement and the management of this property. Although priority was given to the properties under MCYS, the project is now aimed towards the development of an automatic information system for the protection, study, preservation and management of cultural heritage objects, regardless of their ownership, which might be the State, religious institutions or private owners. The law establishes that 'computerised cataloguing' is the recording of data according to established scientific criteria on the identification and rapid administration of the cultural property, while the 'object's passport' is the identity card of a cultural heritage object, containing images, films, sketches, location, size, weight, material, the computerised code, the author, the preservation repository, description and history.

Project objectives

The major aims of the NCCPI are:
the full provision of objects' passports as well as the cataloguing of each cultural object's data for the purpose of their identification and management by a unique central database;
the remote communication of NCCPI with the regional centres and the main scientific institutions for the purpose of collecting and updating information on the movement and changes of the objects;
the management and control of the cultural properties based on the identification data;
the collection and preservation of scientific data and of the maintenance (restoration) data of the objects;
the facilitation of access to this information for interested parties, such as the State central administration (regarding each object's situation and scientific management), the protection and preservation institutions, customs (for identification and control of movement), as well as scientific and educational institutions, and the public (including Internet publication for the purposes of knowledge, study, dissemination, instigation of their use in the educational curricula and for cultural tourism).

The database is built on the descriptive data of the object passport. There are about 30 metadata fields, divided into a number of areas such as identification, scientific and administrative data. Each descriptive datum can serve as a potential search element, be it an individual search or combined through filters, as well as an element for report preparation. It will be used according to three use and security levels, following the three categories of concerned parties mentioned above.

While more than 95 per cent of the collections belonging to the network of national and regional museums (from the Cultural Monuments Institute,[21] to art galleries and studios) that depend on MCYS have already been registered in the 'Onufri' database, the work is targeted at the further inclusion of information from the institutes within the Academy of Sciences, such as the Archaeological Institute, the History Institute, the Institute of Popular Culture, and from all of those who administer Albania's cultural heritage.

[20] See "European Commission – Council of Europe Joint Programme: Integrated Rehabilitation Project Plan/Survey of the Architectural and Archeological Heritage (IRPP / SAAH)", Albania, March 2004, p. 9, 'Documentation'.

[21] The Cultural Monuments Institute has under its custody archaeological sites such as the ancient town of Butrint (see http://www.butrinti.org)

Since the best part of the infrastructure is provided by the State Budget, the Centre is working towards the scientific training of its staff[22] in the Centre and in the regions, the continued improvement of the software for distance communication with its associated limitations, the elaboration of the data and translation fields in which, despite domestic or international support offered to date, there is still room for improvement and updating. Communication with the regional branches was due to start at the beginning of 2005 after the necessary tests.

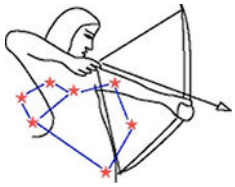
The Onufri database is divided into three informative catalogues:
Catalogue of the cultural properties that have Cultural Heritage status;
Catalogue of properties without this status, free from movement limitations;
Catalogue of lost objects (subject to restricted data).

The directors of the Centre are aiming at raising the status of the NCCPI to a National Cataloguing Institute.

BELGIUM

Report by Tijl Vereenooghe

CHIRON



© CHIRION, 2005

In Greek mythology, Chiron – originally a Thessalian god of healing – survived as one of the centaurs. Unlike the others of his race, Chiron was wise and had an extensive knowledge of the healing arts. He had been the tutor of, among others, Asclepius, Theseus and Achilles. Now, CHIRON is also the acronym for a new Marie Curie EU-funded project, which started in February 2005. CHIRON[23] will provide training fellowships for graduates wishing to start a research career in the field of IT applications to the research, conservation and presentation of tangible cultural heritage. The project will consist of a joint training programme and individual research carried out by fellows within a coordinated framework at the participating partner institutions. CHIRON has a duration of four years with an overall budget of about €2,300,000.

Marie Curie Host Fellowships for Early Stage Research Training are aimed at offering structured scientific and/or technological training as well as providing complementary skills to researchers with less than four years of research experience. Activity consists mainly of hosting young researchers, completing their training with courses and/or participation in research activity. Complementary skills include languages and management. The Marie Curie actions are intended to promote the development of human resources that constitute the underlying raw material on which research capabilities are built. The initiative also promotes the mobility of researchers between European countries and to Europe from outside.

In the field of material cultural heritage IT has proven to be an effective tool, using state-of-the-art technology adapted from other sectors or purposely developed to take into account the peculiar necessities of cultural applications. As interdisciplinarity is essential, young researchers need to be exposed to a variety of specialised techniques as well as to on-site experiences. Moreover, specialisation in a single discipline has placed the focus on individual steps of the production process, which leads from data acquisition, management and processing to exploitation and communication. Fragmentation is thus an issue to be addressed promptly and needs a holistic approach, which at present no training institution can provide.

The partners of the CHIRON network cover complementary parts of the potential training field. Their collaboration in a joint project will provide integrated coverage of the different specialisations, adding a common perspective and coherent approach. This will be developed through a joint training activity, in which all the young researchers will participate, and specific research projects carried out by individual partners in a coordinated way, which will be the main training/research activity of hosted fellows. CHIRON partners offer researchers a spectrum covering different archaeological periods and different locations – from Northern/Central Europe to the Mediterranean Area and the Near East. A wide range of technologies is covered, from database to multimedia and state-of-the-art 3D modelling and visualisation applications to cultural heritage. Economic and social implications will also be considered. According to the network implementation plans, there will be 7-12 fellows every year working on individual and independent, though coordinated, research plans. Their subjects will cover most, if not all, areas in the field.

EVENT REPORT: VAST2004

VAST2004 – The 5th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage – was hosted in Belgium. Organised by the Ename Center for Public Archaeology and Heritage Presentation (http://www.enamecenter.org), it was the largest VAST symposium to date, with more than 150 participants. It was dedicated to the theme 'Interdisciplinarity or *The Best of Both Worlds*: The Grand Challenge for Cultural Heritage Informatics in the 21st Century'.

The keynote talk by Neil Silberman (of the Ename Center) raised a critical voice against the easy temptation of being fascinated by the charm of technology and of rationalising from inside its use, just improving its performance. He reminded us that it is no use having a speedy car if one does not know how to drive it. Patricia Manson (EC Head of Unit) discussed some of the emerging issues for digital cultural content as we move from relatively simple developments based on access and discovery to environments that support new forms of user experience and of creative expression, and presented the role of IST research in addressing them.

Since the beginning, VAST has aimed at being a clearing-house for scientists of different provenance and this was reflected in the lecturers' contributions. There were a number of fully fledged papers, dealing with mature research in great scientific depth, while another set of shorter papers dealt with current projects, reporting on advances in ongoing research or exploring new research paths with preliminary announcements.

A third conference track for EPOCH showcases was also offered concurrently with the two paper tracks. These showcases aimed at demonstrating the potential of what is already available as a product of research. EPOCH also held its General Meeting in Oudenaarde where the results of the first year's activities were presented (see the first issue of the first EPOCH newsletter, which is available from http://www.epoch-net.org, for more information). In 2005 the VAST symposium will be hosted in Pisa, Italy, in October: http://www.vast2004.org.[24]

[22] See "European Commission – Council of Europe Joint Programme: Integrated Rehabilitation Project Plan/Survey of the Architectural and Archeological Heritage (IRPP/SAAH)", Albania, March 2004, p. 8, 'Management'.
[23] See the Web site at http://www.chiron-training.org for more information.
[24] This and other events can be viewed from DigiCULT's Events Database, available through http://www.digicult.info.

GREECE

Report by Efthimios Mavrikas

Use of photographs for recording statistical information regarding the cultivation of olives © Ergani, 2005, http://www.ergani.org.gr

OPERATIONAL PROGRAMME "INFORMATION SOCIETY" INVITATION 65: A REVIEW OF SELECTED PROJECTS FUNDED BY THE 3RD COMMUNITY SUPPORT FRAMEWORK 2000-2006

Invitation 65 for Culture – 146 projects funded under Information Society Action Line 1.3, 'Documentation, Exploitation and Promotion of Hellenic Culture' – has been under way for the past few months, and is steadily altering the country's cultural map. The Operational Programme 'Information Society' (http://www.infosociety.gr/) Secretariat has allocated some €50m at national level to public and private non-profit organisations, as well as academic and research institutions, which own and/or manage cultural assets, with the aim of bringing traditional methodologies and formats of heritage management, digitisation and promotion into the Information Age, and expanding the availability of these assets for research and public access to unprecedented levels.

This review focuses on the frontrunners of Invitation 65, starting off with a particularly interesting case – a state-of-the-art historical memory preservation organisation situated in a marginal region of Europe, not far from the Asia Minor coastline.

THE HISTORICAL ARCHIVE OF THE AEGEAN 'ERGANI': PRESERVING HISTORICAL MEMORY ON THE SEMANTIC WEB

The Historical Archive of the Aegean Ergani (http://www.ergani.org.gr/) is a civil non-profit organisation based in Mytilene, the capital of Lesvos island. Ergani has undertaken the Invitation 65 project entitled *Development and exploitation of a thematic digital collection regarding the modern economic and social history of the North Aegean region (1870-1930)*.

The object and historical study catalyst of the project is the Kourtzi family archive, a rich set of evidence that includes architectural and industrial plans, technical manuals and designs, diplomatic and accounting records, diaries and logbooks, personal and business letters, market studies and advertising material, newspapers and magazines, ethnographic period footage in the form of amateur Pathé Baby movies, photographs, postcards and works of art. The evidence sketches a family tale of entrepreneurship and innovation – overcoming troubled times and the unsettled confines of a declining Ottoman Empire and an embryonic Hellenic Kingdom to prosper in banking, commerce, shipping, mining, hôtellerie (spas) and the olive oil industry, founding model industrial processing plants on Lesvos island – which highlights, contrasts and eventually accords with Lesvos' progressive marginalisation and isolation from its natural geographic context – the Asia Minor coast, the eastern Mediterranean and the Black Sea – after joining the Hellenic state.

The Kourtzi family circa 1920 © Ergani, 2005, http://www.ergani.org.gr

Ergani follows a semantically consistent methodology for the electronic documentation and study of the archive, enabling knowledge-rich online and offline presentations of the Kourtzi family cultural property. The following two-fold description summarises these methodological choices:

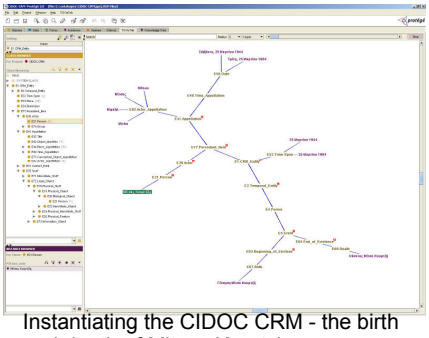
1. Ensure optimal coverage and accessibility of documentary evidence by detailing an initial structured and authoritative view of the archive through the comprehensive application of archival description and interoperability standards, extensive use of thesauri, and standardised signalling of personal, temporal and geographical information.
2. Open documentary evidence to historical analysis and materialise and creatively visualise the resulting discursive structures by interconnecting documentary evidence through the instantiation of a central knowledge model.

The former methodological choice is implemented by following RLG guidelines for cultural materials (http://www.rlg.org/en/page.php?Page_ID=555); primary archival documentation work produces ISAD descriptions (http://www.ica.org/biblio.php?pdocid=144) encoded in XML following the EAD DTD (http://www.loc.gov/ead/ead.html), and indexed in a native XML database together with OAI-PMH metadata (http://www.openarchives.org/OAI/2.0/guidelines.htm) extracted through the use of XSL stylesheets. Controlled documentation vocabulary is referenced to a selective list

ΙΣΤΟΡΙΚΟ ΑΡΧΕΙΟ ΑΙΓΑΙΟΥ "ΕΡΓΑΝΗ"

Ermou 220, 81100 Mytilene, Lesvos, Greece © Ergani, 2005, http://www.ergani.org.gr

of thesauri, including the Getty Art and Architecture Thesaurus (http://www.getty.edu/research/conducting_research/vocabularies/aat/), the Getty Thesaurus of Geographic Names (http://www.getty.edu/research/conducting_research/vocabularies/tgn/), the Library of Congress Thesaurus for Graphic Materials I (http://www.loc.gov/rr/print/tgm1/) and II (http://www.loc.gov/rr/print/tgm2/), and the UNESCO Thesaurus (http://databases.unesco.org/thesaurus/).



Instantiating the CIDOC CRM - the birth and death of Mitsas Kourtzis

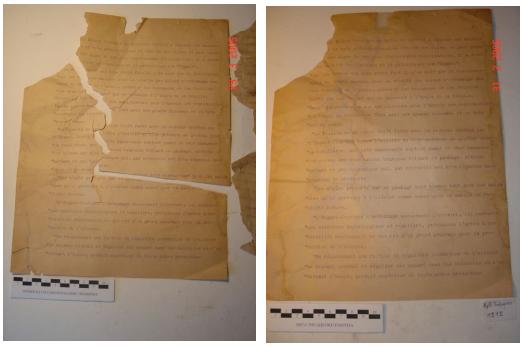
The latter methodological choice is implemented by placing the CIDOC Conceptual Reference Model (http://cidoc.ics.forth.gr/) at the centre of a scenario-driven information extraction approach supporting, and supported by, the historical analysis of documentary evidence. The CIDOC CRM enables the semantic linking of discursive network points extracted from intentional structures spanning the entire archive, which are backed or contradicted by subsequent historical analysis. This semantic basis acts as an enabler of knowledge-rich, contemporary historical discourse, and offers a very flexible way to explore and interpret historical information: on the one hand, the researcher may find the composition of an interpretative horizon with multiple narrative paths often fragmented and imprisoned by the archival evidence itself; on the other hand, the researcher may find the discursive multiplicity of interpretative and hermeneutical tools that the study of historiography demarcates. As a result, the interpretative grid created is always a new hermeneutic universe.

The complexity of relations, links and flows of events and discursive instances, and their representation as historical memory preserved within the Kourtzi family archive, composes the canvas of a continuous navigation which is none other than narration itself, each time different and unique. The epistemological goal of exercising and reconstructing narrative paths and structures has been set by Ergani in a framework of interdisciplinary research, which is touched upon by the following discussion with the organisation director, Christis Konnaris.

Correspondence between Ach. Kouppas Mechanical Engineering factory in Piraeus and PM Courtzis & Cie © Ergani, 2005, http://www.ergani.org.gr

About organisational aims...

'Ergani aims to collect, protect, restore and promote archival material, as well as actively support and contribute to research related to the history of Lesvos during the 19th and 20th centuries, with an emphasis on the historical role of the island in social, economic, political and cultural terms within the national and international context. Owing to its peripheral locality, Ergani can be considered unique, in the sense that it is the only non-profit organisation on the island which promotes historical research and preserves local heritage.'



Archival material before and after restoration - item 1212 © Ergani, 2005, http://www.ergani.org.gr

About the study of the past...

'Ergani aims to protect and preserve the archival treasures of Lesvos because we acknowledge the importance of preserving local collective historical memory, and the educational and social value of demonstrating the crucial historical role that the island played in the economic, cultural and political developments of a wide geographical area, encompassing Asia Minor, the eastern Mediterranean, the Balkans and the Black Sea. Any study of the past constitutes an attempt to influence the present and the future of a locality and its community. Therefore, one could say that Ergani wishes to participate in the economic and cultural development of the local community through a re-examination of its past, which can form the basis for future orientations.'

On power relations under Ottoman rule - Mitsas Kourtzis irreverent © Ergani, 2005, http://www.ergani.org.gr

About private archives...

'The interest taken by Ergani in private archives and collections, as opposed to archives maintained by public bodies, comes from our firm belief that this type of material and research can enrich our perspectives and broaden our historical narratives by allowing us to focus on the local, personal and private aspects of historical continuity and change. Without wishing to devalue the work and importance of state archives or archives of various public institutions, we strongly believe that the information contained within these sources needs to be re-contextualised and studied in comparison with the complexity of private archives. This re-elaboration of historical material against the backdrop of the documented lives and works of local people, and real-life personal histories seems to incorporate a complex network of historical truths and tensions that can alter and enrich historical discourse, and provide new insights into the reconstruction of our past. In methodological terms, we aim to establish a bottom-up approach to the study

of the past, as opposed to the top-down approach which has been dominant until very recently.'

About intermediation...

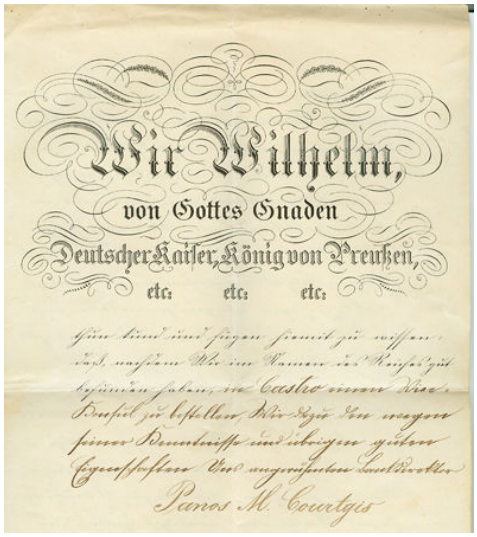
'Ergani is a hybrid organisation, in the sense that it is oriented both towards the promotion of research and towards the preservation of Cultural Heritage. This hybridity stems from our commitment to serve as a middle-agent between the world of research – mainly embodied by universities and other research institutions – and society in general. In other words, we commit ourselves to communicate recent scientific findings regarding the past to the general public through an elaborate and sophisticated popularisation, in a transformational process that renders expert knowledge comprehensible to the non-expert. To actively fulfil this commitment, we organise seminars, conferences, workshops, exhibitions and publications open to the public, making extensive use of information technologies to enhance the accessibility of novel and unexplored archival material.'

Postcards from the archive © Ergani, 2005, http://www.ergani.org.gr

About current activities and transdisciplinarity...

'We have recently embarked on a large-scale documentation and digitisation project which will result in the creation of a digital collection, a Web study space/portal and two educational electronic publications on DVD. This is, in fact, our first major undertaking so far, and the long-awaited result of our policy towards safeguarding an initiative with secure funding standards, rather than involving ourselves in more small-scale projects with insecure funding standards and vague targets – a pattern often followed by many non-profit organisations in our country. This choice, which is always subject to the struggle for viability that any NGO faces in terms of funding and public support – or, in certain cases, even social acceptance – has been a conscious one on our part; we wish to establish ourselves as a state-of-the-art heritage organisation that creates an environment encouraging real interdisciplinary scientific work. The OPIS Invitation 65 has offered us a sense of security, and the Greek State has indeed assisted us in achieving our principal goals. It has also given us the chance to form a working group of experts from a variety of academic and disciplinary backgrounds. We firmly believe that the documentation and analysis of archival material, and the adaptation of their results to a general/public or specialised/expert audience, requires a transdisciplinary approach to the formation of the project team, which should include archivists, historians, museologists, knowledge engineers and designers. Members of this small scientific community need to learn how to communicate with each other, therefore establishing a common language, a common epistemological and scientific code that constitutes one of the fundamental challenges of our present era and marks an attempt to produce valid interdisciplinary results. In our case, the University of the Aegean played a crucial role in ensuring the simultaneous presence of experts from various disciplines on a remote locality such as Lesvos – an invaluable human capital which gave us some of our closest collaborators.'

Restoration work on archival material © Ergani, 2005, http://www.ergani.org.gr



Appointment of Panos Kourtzis as Vice-Consul of Germany in Mytilene in 1896 © Ergani, 2005, http://www.ergani.org.gr

About participation and contribution in research...

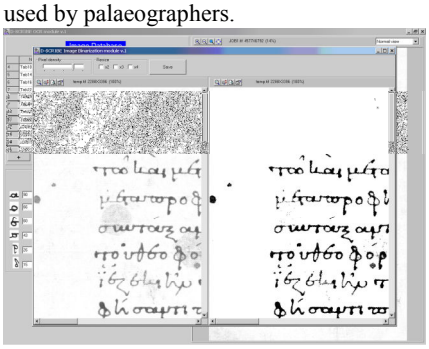
'We aim to achieve regular and dense interaction and cooperation with the national and international research community, since we acknowledge that one of the most devastating consequences of peripherality is the marginalising effect, the failure to participate in larger initiatives beyond local level on an equal basis. Indeed, there is an uneven element here, in the sense that a locality that has played such an active role in the past and has remarkable historical material to offer has been widely excluded and marginalised in the last 70-80 years. The study of its past, its history and the history of its people has a lot of insights to offer, not only for those interested in the history of this particular locality, but also to all those interested in important issues of European history, Ottoman history, Balkan history, the history of the Middle-East and the eastern Mediterranean; issues such as the formation of nation-states, the worldwide spread of capitalism, aspects of social and political change, and the impact of modernity, which are always current.'

Focus on research

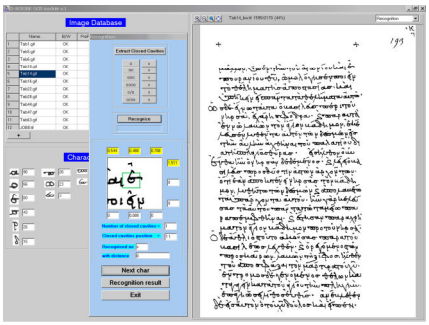
National Centre for Scientific Research (NCSR) 'Demokritos' Institute of Informatics & Telecommunications (IIT) The wealth of cultural information on the Internet and the growing development of digital libraries demand advanced techniques for intelligent data processing and organisation. The *Informatics Department* of IIT (http://www.iit.demokritos.gr/) conducts research and development work into support infrastructures for multimedia content management, and recognition and personalised access to cultural heritage digital material.

Processing and recognition of valuable manuscripts

Valuable manuscript collections are an important source of original information for classical studies. Because of the fragility of the material, new processes for preservation through digitisation need to be introduced. It is also important to provide easy access to the valuable material. The *Computational Intelligence (CIL) Programme* (http://www.iit.demokritos.gr/cil/) of the *Informatics Department* develops technologies for the digitisation, management, processing and recognition of valuable manuscripts. In collaboration with the *Mount Sinai Foundation* in the context of the project *D-SCRIBE* (Hellenic GSRT-funded R&D project: http://www.iit.demokritos.gr/cil/dscribe/index.htm), *CIL* aims to support and facilitate current and future efforts in old Hellenic manuscript digitisation and processing. A number of modules have been developed for the study and processing of the digital collections and are integrated under a common software environment. These modules include a document management system specifically adapted to manuscripts, a module for automatic processing and transliteration of manuscripts incorporating OCR techniques, and a self-study tool to be used by palaeographers.



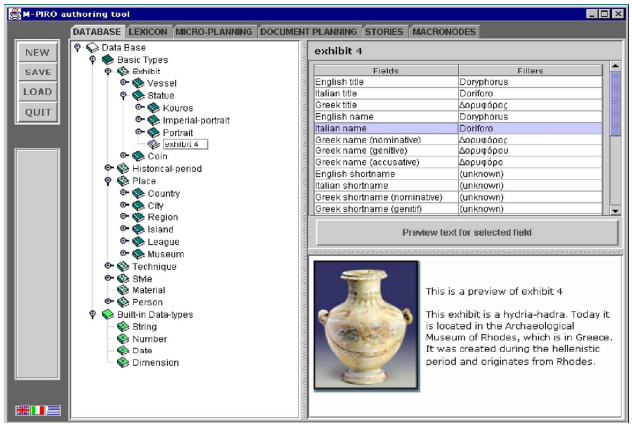
D-SCRIBE project: Image binarisation and enhancement. © D-SCRIBE 2004



D-SCRIBE project: Optical character recognition. © D-SCRIBE 2004

Intelligent management of cultural information

In a world of constantly growing information resources we need responsive and intelligent ways of locating and delivering these resources to the users. In the last few years, researchers at the



© M-PIRO, 2004

© M-PIRO, 2004

Software & Knowledge Engineering (SKEL) Laboratory (http://www.iit.demokritos.gr/skel/) at the *Informatics Department* worked together with other European R&D groups in the context of the *M-PIRO* project (http://www.ltg.ed.ac.uk/mpiro/) for the development of intelligent technologies that enable personalised access to cultural resources in a natural way. The project worked closely with museums, galleries and other memory institutions to develop technologies that allow written and spoken descriptions of exhibits to be generated automatically from an underlying language-neutral database and existing free-text descriptions.[25] The resulting descriptions, which are generated in English, Greek or Italian, are tailored according to the user's interests, background knowledge and language skills. For example, simpler vocabularies are used to respond to children, whereas more complex vocabularies are used to respond to mature subject experts. SKEL developed an innovative authoring tool, which allows domain experts (museum curators) to define or modify the structure and contents of the museum database used as input by the M-PIRO system, as well as to create or update the domain-dependent linguistic resources in the three languages of the project. This tool is currently being extended in order to enable the import of information from existing databases as well as from ontologies expressed in OWL Web Ontology (http://www.w3.org/TR/owl-ref/).

[25] The chapter on Human Language Technologies in *DigiCULT Technology Watch Report 3* (December 2004) considers how this could be achieved in and by cultural heritage institutions.

LITHUANIA

Report by Zinaida Manzuch

The 7th Congress of Baltic Librarians *Diversity in Unity: Baltic Libraries in the European Union*, which took place on 30 September – 2 October 2004 in Jumurda, Latvia, provided the occasion for an information exchange on new developments and achievements in Latvia, Estonia and Lithuania. It was not only a fruitful meeting with foreign collegues, but also a good opportunity to find out about current activities in Lithuania. Based on presentations by Lithuanian collegues, two major initiatives are reviewed as follows.

Martynas Mazvydas National Library of Lithuania (http://www.lnb.lt/) is moving towards an ambitious project involving a Lithuanian virtual library of cultural heritage. Fulfilling the functions of preservation of and access to Lithuanian cultural heritage, the National Library has been developing a virtual environment concept that will enable equal access to Lithuanian cultural assets for all, including both Lithuanian citizens and foreign users. The National Library received PHARE funding (http://europa.eu.int/comm/enlargement/pas/phare/) for the preparation of the project and its experts are engaged in the development of a *Virtual Integrated Library Information System*. In cooperation with 15 archives, 16 museums and more than 60 research and public libraries, the National Library of Lithuania will launch a huge digitisation initiative. Results are predicted to occupy 11 million digitised pages. The focus of the project is the creation of an integral digital space to represent and preserve Lithuanian cultural heritage and joint efforts and resources of Lithuanian memory institutions. The virtual library will be based on the LIBIS (*Lithuanian Integral Library Information System*, http://www.libis.lt) infrastructure. Currently the LIBIS network encompasses 60 libraries, enabling a joint environment for technological libraries' processes and user services, and provides numerous bibliographic databases for collective use. LIBIS has already been applied for

safeguarding digital cultural heritage. Its sub-system *Archive of Electronic Resources* is designed to collect and preserve digital resources for future generations. Some 4.5 million digital documents have already been archived.[26]

Fourteen Lithuanian higher education institutions and Riga Technical University (http://www.rtu.lv/) carried out an initiative called *Electronic Theses Dissertations (ETD) Lithuania Project as Pilot for Baltic States* (for further information, see http://www.labt.lt/index_projektas_en.php?psl=projektas/ETD_en.htm). The project aimed to create tools and methodology for the publishing and long-term preservation of theses and dissertations of master's and doctoral students. The project findings, disseminated to academic institutions in the Baltic States, provide a model for electronic publishing, access and preservation of students' research. A Lithuanian and international context in the field of full-text electronic databases was a prerequisite for the initiative. The *Lithuanian Academic Libraries Directors Association* (LABA) is implementing a long-term programme *Information Technologies for Science and Higher Education, 2001–2006*, with the support of the Ministry of Education and Science of the Republic of Lithuania. In 2002 a sub-programme *Electronic Publishing and Full-text Databases in Lithuania* encouraged a dialogue between ICT specialists, librarians and publishers. In 2003 LABA initiated work on Lithuanian electronic theses and dissertation (ETD) databases. Experts explored opportunities for publishing, archiving and providing access to ETD and underlying ICT tools and methodologies. An international initiative for *Networked Digital Library of Theses and Dissertations* (see http://www.ndltd.org/info/index.en.html) expanded the vision of ETD from the national to the Baltic region. The *ETD Lithuania Project as Pilot for Baltic States*, funded by UNESCO (http://www.unesco.org) and coordinated by the *Information Technology Development Center* of the *Kaunas Technological University* (KTU: http://www.ktu.lt/en/), was launched in December 2003 and implemented in March 2004. The project proposed the vision of a joint ETD information system that offers the same environment, tools, methodology and procedures for all higher education institutions. During the project the ALEPH ETD catalogue was configured, and ETD resources are also available at the virtual *MetaLib* portal (http://www.library.lt). Metadata and summaries of 46 master's theses maintained in KTU and a doctoral dissertation from Riga Technical University were placed in the system. Each project participant established an ETD Committee to coordinate the submission of electronic student papers. The organisational schema and technological infrastructure offered by the project allow students to submit their research to an electronic ETD database, which will increase the visibility of research performed by students in various countries and improve the quality of papers.[27]

[26] Regina Varnienė, Director, Bibliography and Book Science Centre, Martynas Mazvydas National Library of Lithuania, *New Strategic Activities of the Martynas Mazvydas National Library of Lithuania*. Presented at the 7th Congress of Baltic Librarians, Jumurda, 30 September 2004.

[27] Aleksandras Targamadze, Dean, Faculty of Informatics, Kaunas University of Technology /Adviser, National Programme 'Information Technologies for Science and Higher Education' (ITMiS);
Antanas Štreimikis**,** Deputy Head, Library Software Department, Information Technology Development Centre, Faculty of Informatics, Kaunas University of Technology / Coordinator, Lithuanian Academic Library Network (LABT) Project Group;
Vilius Kučiukas, Head, Library Software Department, Information Technology Development Centre, Faculty of Informatics, Kaunas University of Technology / Head, Lithuanian Academic Library Network (LABT) Project Group;
Lina Bloveščiūnienė, Head, Cataloguing Department, Kaunas University of Technology Library / System Librarian, Lithuanian Academic Library Network (LABT) Project Group:
Lithuanian Electronic Theses and Dissertations (ETD) Project as a Pilot Project for the Baltic States. Presented at the 7th Congress of Baltic Librarians, Jumurda, 1 October 2004.

THE NETHERLANDS

Report by Johan Oomen and Eelco Bruinsma

CATCH RESEARCH PROGRAMME

The Netherlands Organisation for Scientific Research (NOW, http://www.nwo.nl) recently launched a new initiative that aims to find solutions for persistent obstacles that are hindering progress towards continuous access to all cultural heritage collections. The CATCH programme (Continuous Access To Cultural Heritage) began in October 2004 and will run for a period of six years. The two central research issues in the CATCH programme are:
To what extent is it possible to develop innovative tools to connect knowledge and cultural objects (research question 1), to integrate scattered digitised cultural objects (research question 2), and to increase the accessibility of and interaction with our cultural heritage, supporting and improving the work of professionals (research question 3)?
Can we develop scientifically relevant methods to acquire new fundamental and applied knowledge about these processes and their IT-based solutions?

The challenges implied by the research questions are common to all cultural heritage institutions throughout the world. The CATCH programme joins ongoing international efforts to improve access to heritage resources. On the one hand, CATCH aims to develop tools to improve the specific situation for Dutch cultural heritage (research question 1), while on the other hand CATCH also wants to contribute new methods and techniques to the international research effort (research question 2).

In this demand-based programme the interests of the (potential) users of the research results are of outstanding importance. Hence the programme strategy has a two-fold focus: research and implementation. The CATCH research strategy concentrates on three research themes: semantic interoperability through metadata, knowledge enrichment through automated analyses, and personalisation through presentation.

The research and implementation will be carried out by research teams consisting of CATCH-funded temporary researchers, temporary scientific programmers and senior research staff (all employed by universities), and programmers and senior staff employed by cultural heritage institutions. With an estimated total budget of €12.5M in subsidies (to be realised in two phases), CATCH will be able to fund about 17 of these research teams. The programme will start with six research teams, each executing one of the six core projects that lay the foundation for the programme. The 11 remaining teams will be selected in competition on the basis of research plans. All Dutch universities can enter the competition, which will be organised by NWO. The participating cultural heritage institutions will contribute €2.8M in kind to the programme. The support programme provides for the transfer of knowledge and tools within the programme and to all other parties interested in the CATCH results.

More information is available from http://www.nwo.nl/catch.

From Elements to Elephants: A Review of Progress in providing Online Access to the United Kingdom's Archival Information

Amanda Hill, University of Manchester

In 1998 the National Council on Archives published a forward-looking paper called *Archives On-Line: the establishment of a United Kingdom archival network*.[28] This paper recommended '…the creation of a network that aims ultimately to provide access from a single gateway to *all* archival catalogues in the United Kingdom', hoping to '… avoid the problems of piecemeal investment in systems and practices which, in the event, may prove to be mutually incompatible'.

Unfortunately, the single gateway pictured by this report did not immediately materialise. Instead, in the years following the publication of *Archives On-Line*, the archival community in the United Kingdom developed a number of collaborative projects that have all been putting the information held in paper catalogues online. National projects include A2A (for England),[29] the Scottish Archive Network (SCAN),[30] Archives Network Wales (ANW)[31] and the Archives Hub[32] (covering records held in universities and colleges throughout the UK). Funding sources for these projects have been varied, reflecting the variety of institutions in which archive collections are to be found. There are also a number of collection-mapping projects, many of which originated in the higher education sector and were funded originally by Research Support Library Programme grants.[33] Many of these latter projects have a subject focus, but the largest, AIM25,[34] is geographically based, describing archives held in educational and learned institutions in the London area.

The impact of these different services has been significant, as the evaluation activities undertaken in the course of the last five years have shown.[35] Usage is high, and the services have been successful in attracting users who have not previously used archive catalogues and record offices. Of those who filled in the 'new user' survey on the A2A site in 2000-2002, over 60 per cent had not used archive catalogues before.[36] The success of the individual services has, however, to be balanced against the frustrations of their users, who frequently complain that there are now simply too many such Web sites, with sometimes overlapping remits and content. This is also a problem for the custodians of the archives, who may find that there are several copies of their descriptions available in a variety of online services. This increases their exposure, but can cause headaches when the descriptions need to be updated or deleted.

Plans for providing an overarching search mechanism for the principal online archival resources have been developed by The National Archives and a wide range of partners as part of the 'Linking Arms' programme.[37] This plan reflects the conclusions of the Government-commissioned Archives Task Force, which had the creation of an Archives Gateway as its principal recommendation.[38] There is widespread support for the concept of the Gateway among archivists, users and potential users of archives, but securing a financial commitment to develop and maintain such a service is proving more difficult.

The effectiveness of an Archives Gateway will depend to a large extent on the quality and comprehensiveness of the finding aids that are made available for searching. Standardisation is crucially important in the world of union catalogues, and the rapidity of the adoption by the UK archival community of the General International Standard for Archival Description (ISAD(G))[39] has been notable. Indeed, conformance to ISAD(G) was the first requirement of National Council on Archives' *Interoperability Protocol*, published in November 2003.[40] This document was drawn up with the future Archives Gateway in mind:

> "The aim of this protocol is to facilitate interoperability between these strands of the archival network in the UK. The protocol seeks to provide a standard of minimum conformity and full compliance, covering data structure, content and technical matters, to which existing strands of the current archival network have committed themselves and which newly developed resources should adopt."
> *Interoperability Protocol*

Other standards covered by the protocol include those for data exchange formats, subject indexing and name authorities. Encoded Archival Description (EAD) has become the accepted XML standard for encoding the elements of electronic finding aids in recent years and is now widely used all over the world for the exchange of archival data. Proprietary database systems are widely used in the UK, and these are increasingly offering EAD export and

---

[28] *Archives On-Line: the establishment of a United Kingdom archival network*, National Council on Archives, 1998, available online at http://www.ncaonline.org.uk/materials/archiveso nline.pdf
[29] http://www.nationalarchives.gov.uk/a2a/
[30] http://www.scan.org.uk/
[31] http://www.archivenetworkwales.info/
[32] The Archives Hub is funded by the Joint Information Systems Committee (JISC) and is available at http://www.archiveshub.ac.uk/.
[33] http://www.rslp.ac.uk/projects/
[34] http://www.aim25.ac.uk/

[35] Several of these are available from http://www.archiveshub.ac.uk/introduction.shtml.
[36] *A2A: Access to Archives Report, April 2000 – March 2002,* A2A Central Team, E-Access Department, Public Record Office, 2002.
[37] More information on Linking Arms can be found at http://www.nationalarchives.gov.uk/partnerprojects/linkingarms/.
[38] *Listening to the Past, Speaking to the Future*, Report of the Archives Task Force, MLA 2004, available online at http://www.mla.gov.uk/action/archives/00archives.asp

[39] *ISAD(G): General international standard archival description*, International Council on Archives, Committee on Descriptive Standards, 2000, available online at http://www.icacds.org.uk/eng/standards.htm
[40] The Interoperability Protocol can be found at http://www.ncaonline.org.uk/materials/interoperabilityprotocol.pdf.

import facilities in response to demand from their users.

Subject indexing has been another area where the archival community has seen rapid change over the last five years. Archivists in the UK have generally not been trained in subject indexing at postgraduate level, and yet subject access to archival resources is vitally important in an online context, where there are no friendly staff members to guide researchers to collections which match their area of interest. The UK Archival Thesaurus project (UKAT)[41] was developed during 2003 and 2004 to meet a need for a thesaurus that met the particular requirements of archivists in this country. The full thesaurus is now available for download from the UKAT project's Web site in SKOS-Core RDF format.[42] It should be noted, however, that it is unlikely that every archive repository in the country will adopt UKAT for subject indexing; a significant number of university archives, for example, use the Library of Congress Subject Headings, often in line with the practice of their university library. It might be that technology will provide the solution to the use of multiple thesauri, as it does with the 'Subject Finder' facility of the Archives Hub.[43] The results of a search in the Subject Finder on the word 'textiles' are shown in Figure 1, illustrating how terms from different thesauri are returned together, to give the user the best chance of finding appropriate collections.

Figure 1: Subject finder facility © Archives Hub, 2005

Despite overall adherence to the structure of ISAD(G), there are significant differences between the data of the existing online catalogues that will have an impact on the experience of the users of the Archives Gateway. Most significant, perhaps, is the proportion of full finding aids online, compared with summary descriptions of archival collections. In February 2005 the numbers of finding aids in those two categories within the main archive network strands are shown in Table 1:

Resource	Multi-level finding aids	Collection-level descriptions
A2A	57,000	30,000
AIM25		6,000
Archives Hub	50	18,800
Archives Network for Wales (ANW)		5,100
Scottish Archive Network (SCAN)		20,000
Total	**57,050**	**79,900**

Table 1: Number of full and collection-level finding aids in UK online archive networks

Resource	Finding aids	Subject index terms	Terms per finding aid
A2A	87,000	26,744	0.3
AIM25	6,000	65,867	11
Archives Hub	18,850	42,711	2.26
Archives Network for Wales (ANW)	5,100	11,168	2.1
Scottish Archive Network (SCAN)	20,000	0	0

Table 2: Number of subject index terms in UK online archive networks

As this table shows, more than half of the 134,850 finding aids currently available through the major archive network Web sites are at collection level only. Although some of these descriptions are very full, they obviously cannot contain as much detailed information as a full archive catalogue. Research carried out for the LEADERS project, based at University College London, showed that 76 per cent of users of archive services are looking for names of individuals, families, organisations or places.[44] The needs of

[41] http://www.ukat.org.uk/
[42] The SKOS-Core guide is available at http://www.w3.org/2001/sw/Europe/reports/thes/1.0/guide/.
[43] The subject finder can be found at http://www.archiveshub.ac.uk/searchsub.shtml.

[44] Anna Sexton, Chris Turner, Geoffrey Yeo and Susan Hockey, "Understanding Users: a prerequisite for developing new technologies" in the *Journal of the Society of Archivists*, April 2004.

THESES ALIVE PLUGIN FOR INSTITUTIONAL REPOSITORIES

Edinburgh University Library (http://www.lib.ed.ac.uk/) has added extra functionality to DSpace (http://www.dspace.org/), the digital library system that facilitates the capturing, storage and maintenance of repositories of university-level research. Developed with JISC funding (http://www.jisc.ac.uk/), the new system is called TAPIR (Theses Alive Plugin for Institutional Repositories) and was designed primarily to create a supervised authoring facility, for supervisors to observe, comment on and change ongoing work by their students. TAPIR is still under development; it will, however, be Open Source and freely available. The source code for both TAPIR and DSpace is available from SourceForge (see http://sourceforge.net/projects/tapir-eul). More information can be found in the excellent Ariadne articles: http://www.ariadne.ac.uk/issue41/jones/ and http://www.ariadne.ac.uk/issue38/jones/.

these users are not being fully met by the current resources.

The other major group of users, forming 19 per cent of the LEADERS project's respondents, are searching for records relating to particular topics. This group is overwhelmingly comprised of professional and academic researchers, whose needs are best met by providing information about subject strengths in collections. Unfortunately, subject indexing is extremely variable across the existing archive networks: some descriptions have no subject index terms attached, others very few, as Table 2 demonstrates.

Providing a subject index search across these resources will not produce reliable results, meaning that only simple keyword searches are planned for the first stage of the Archives Gateway. In the future, however, it may be possible to retrieve unindexed records in response to a subject search, by comparing keywords associated with a term from those records that are indexed, and then feeding those keywords back as search terms to the records that do not contain subject index terms.

Further funding may well be required to bring our existing online finding aids up to the standards required for a truly useful Archives Gateway, although improvements to the quality of electronic catalogues are already being carried out by the repositories who own them. One important step that should now be taken to make the Gateway more viable is to standardise the way in which the archive networks are populated with data. The best solution would be for each archive repository to own and maintain one 'master description' for each of its finding aids. Providers of archive networks then need to put appropriate procedures in place for regularly updating their copy of the finding aids from their contributing repositories, to ensure that their content does not become increasingly out of date. This will also guarantee that any improvements to catalogues that are made by the holding repositories will be mirrored in the networked version: this is often not the case in the current network strands. It should also be possible to make the content exposed in this way available to other resources, such as regional or subject-based gateways or portals.

Duplication of content will be a real problem in the Archives Gateway. Some catalogues are already available in more than one of the archive networks, often in slightly different forms. Data originally created for the Archives Hub, for example, have been provided to a number of other projects. Once these are drawn together into a combined gateway it will be extremely difficult for users to know which of the duplicate descriptions is the most current or comprehensive. This will be less of a problem if all the network services are able to draw their information from a master copy in the manner described above, as exact duplicates would then be easy to identify and to hide from display.

The Archives Hub is attempting to address some of these issues with its move towards a distributed model, where repositories are able to host their own EAD descriptions within a local version of the Archives Hub software (known as a Spoke), while still making them available to the main Archives Hub service, for users to search. The Hub will gather information from the Spoke each night, so that its data are never more than 24 hours out of date.[45] In this way, responsibility for maintenance of the electronic finding aids rests firmly with the repository, which also benefits by having Web and Z39.50 access to its own data. It will also be possible to provide access via SRW (search/retrieve Web services) in the next version of the software.[46]

Uncatalogued collections form another problem area in relation to the creation of an Archives Gateway. The Operation Logjam project, carried out in the north-west region of England during 2003, surveyed 30 of the area's main archive-holding organisations. The project team found that, on average, 29 per cent of archive collections held by repositories in the region are unavailable for research because they have not been catalogued.[47] It is likely that a similar proportion of uncatalogued material is to be found (or rather, not to be found) in the rest of the UK. The Logjam report estimated that it would take 65 archivist years to catalogue just those collections identified as of highest priority. The estimate for eliminating the entire cataloguing backlog in that one region was 299 years of professional work. For an archival gateway to be truly representative of the nation's archives, work on prioritising the cataloguing of these backlogs and making the results available online is crucial.

Perhaps all this is yawningly obvious for DigiCULT readers, particularly those who have had to deal with combining library catalogues, but for many archivists the issues surrounding standardisation, subject indexing and union catalogues are quite new. The EAD/Data Exchange group of the Society of Archivists works hard to raise awareness of these issues within the profession, running training courses for newly qualified archivists and providing information relating to other resources on its Web pages.[48]

The development of standards for the description of content and for the exchange of archival data has been essential in the journey towards the creation of national and international archives networks. Much progress has been made in the UK towards bringing information about archives online, but there is still a lot to do before the 1998 vision of *Archives On-Line* is realised. It is in the provision of training, tools, procedures and support for staff in the UK's archive repositories that the biggest challenges for the domain remain. Putting these key elements in place will help to ensure that the resources that are built from them in the future will be vibrant, living stores of the nation's memories: big grey elephants, rather than expensively produced white ones!

---

[45] More information about the Spokes software is available from http://www.archiveshub.ac.uk/arch/spokes.shtml.

[46] The SRW maintenance agency is the Library of Congress: http://www.loc.gov/z3950/agency/zing/srw/

[47] Janice Tullock and Alexandra Cave, *Logjam: an audit of uncatalogued collections in the North West*, North West Museums, Libraries and Archives Council, 2004. Available online at http://www.nwmlac.org.uk/News/PDFs/Logjam1.pdf

[48] http://www.archives.org.uk/groups/eaddataexchangegroup/usefullinks.html

(cont…)

KOSMOPOLIS DIGITAL COLLECTION

Dimitrios Gavrilis, Panos Georgiou, Fiori Papadatou and John Tsakonas
Library & Information Service, University of Patras, Greece

INTRODUCTION

Kosmopolis is a digital collection containing the full-text content of twenty Greek periodicals from the mid-nineteenth century up to the beginning of the twentieth century. The project is being carried out by the Library & Information Service of the University of Patras within the framework of the TELEPHAESSA project (http://www.lis.upatras.gr/LIS/TELEPHAESSA/telephaessa_EL.shtml), in partnership with the Department of Greek Literature of the University of Patras and The Hellenic Literary and Historical Archive (ELIA, http://www.elia.org.gr/), Athens, Greece. The project is funded jointly by the European Social Fund and the Greek Ministry of Education within the Third Community Support Framework.

The tasks of the project include digitisation, bibliographical and scientific processing of the documents, and delivery to the users via a Web interface. The basic objectives of the project are:
to extend public and scientific community access to rare and valuable content;
the enrichment of digital content in the Greek language;
to contribute to the preservation efforts of historical and valuable material;
to create and promote new research challenges in Greek literature and history through the scientific processing of the digital content.

As a multi-purpose project, the Kosmopolis workflow and procedures were designed according to a step-by-step approach. Basic issues to be dealt with were:
definition of the objectives and demands of the system, workflow design and resource management among the partners;
definition of the basic and advanced standards and requirements for issues such as digitisation, technology and infrastructure, access to and storage of the digital content;
development of the appropriate technical and organisational strategies for long-term digital preservation of the content;
structure and functions of the appropriate Web interface;
metadata schema;
service evaluation models.

ORGANISATIONAL CONCEPTS

Partnership
Kosmopolis represents a collaboration between the following partners:
Library & Information Service (LIS), University of Patras (http://www.lis.upatras.gr/): LIS is the project coordinator with financial and administrative responsibilities. Furthermore, LIS played the major part in determining the technical specifications in digitisation and preservation issues, as well as in the design and development of the integrated digital collection system (e.g. the software and Web site). Finally, LIS stores the digital material in its storage facilities while it hosts, operates and maintains the appropriate equipment.
Department of Greek Literature, University of Patras: The department is responsible for the scientific coordination of the project (including title selection and indexing). It simulates the role of the potential end-user for design and test purposes.
Hellenic Literary and Historical Archive (ELIA, http://www.elia.org.gr/): This partner provides the original printed material and performs the initial digitisation phase.

Digitised material
The selected material belongs to ELIA's collection and includes periodicals of the nineteenth and early twentieth centuries. Basic selection criteria were (among others) the physical condition of the original, the availability and completeness of the series, and the absence of copyright problems.

Kosmopolis periodical & magazine titles

	Title	Coverage	Pages
1	Poikili Stoa (Ποικίλη Στοά)	1881-1914	7290
2	Euterpe (Ευτέρπη)	1847-1855	4610
3	Chrysallis (Χρυσαλλίς)	1863-1866	2934
4	Ionios Anthologia (Ιόνιος Ανθολογία)	1834-1835	1218
5	Apothiki ton Ofelimon Gnoseon (Αποθήκη των ωφελίμων γνώσεων)	1837-1844	1565
6	Apothiki ton Ofelimon & Terpnon Gnoseon (Αποθήκη των ωφελίμων & τερπνών γνώσεων)	1847-1849	316
7	Nea Zoi (Νέα Ζωή)	1904-1927	4000
8	Grammata (Γράμματα)	1911-1921	3040
9	Alexandrini Texni (Αλεξανδρινή Τέχνη)	1926-1930	2086
10	Argo (Αργώ)	1923-1927	726
11	Melissa (Μέλισσα)	1886-1887	392
12	Chloe (Χλόη)	1869	400
13	Eklekta Mithistorimata (Εκλεκτά μυθιστορήματα)	1884-1894	7200
14	Hmerologio Skokou (Ημερολόγιο Σκόκου)	1886-1918	15,000
15	Asmodeos (Ασμοδαίος)	1875-1880	1428
16	Estia (Εστία)	1876-1895	19,228
17	Deltion Estias (Δελτίον της Εστίας)	1877-1889	-
18	Ioniki Melissa (Ιωνική Μέλισσα)	1850-1852	388
19	Mi Xanese (Μη Χάνεσαι)	1880-1883	4574
20	Noumas (Νουμάς)	1903-1931	10720

Collection requirements
The collection was designed to meet some basic initial requirements:
storage of digital content in appropriate file format to assure quality and conformity in relation to the original, compatibility with

current and emerging image processing technologies, and best storage space usage;
use of proper storage devices for long-term storage conditions, easy copying and transferring and compatibility with current and upcoming hardware and software technologies;
free and open access to digital content via the Web;
content delivery in image, PDF and (in the future) HTML file formats;
functional search, browsing and user help facilities

TECHNICAL SPECIFICATIONS

Digitisation
Taking into account the particular features and physical condition of the original material we used the following specifications to ensure good-quality digitisation:
The creation of two digital copies for each item (single page): a master file for preservation purposes and an access file for speedier delivery.
Master files were scanned at 300 dpi optical resolution, 1:1 size according to the original.
Master files were saved in jpeg format with the minimum possible compression.
Access files were created by processing master files to compressed jpeg files (25%) with the same resolution.
Image files are stored and delivered to LIS on DVDs.
Quality control was performed on the digitisation procedures to ensure that these standards were met.

Preservation of the digital material
For the long-term management and preservation of the collection and its content, LIS decided to adopt and develop gradually an OAIS Reference Model according to ISO 1472 (Consultative Committee for Space Data Systems, 2002). In this initial project phase, LIS adopted a series of actions and procedures aimed at maintaining the best possible conditions for the digital content preservation. These procedures included the storage of master files on DVDs, while the compressed files for access were stored in a RAID disk storage system based on Fiber Channel 2-Gbit technology as well as on DVDs. We also maintained a stable storage room environment with air conditioning to maintain a mean temperature below 20°C. We had a specific plan for periodical checking of the physical condition of content and storage devices, and developed a strategy for the preservation of digital data, whereby data are refreshed (or migrated) onto new storage devices of the same or similar technology every three years.

Metadata schema
Dublin Core was the metadata schema chosen by LIS for the Kosmopolis collection for the following reasons:
consistency with other LIS projects using the same schema, such as the subject gateway e-BGE (http://www.lis.upatras.gr/ebge/);
the wider acceptance and support of DC as a standard metadata schema worldwide;
the flexibility of the schema;
it is accepted by and interacts with the OAI – PMH 2.0.

Kosmopolis digital collection management software
In the early stages of the project, Greenstone digital library software (http://www.greenstone.org) was selected for testing and for the pilot operation of the collection. Most of the appropriate modifications were carried out very successfully and a pilot version of the system was released in late 2003. Soon afterwards, problems arose in the areas of cataloguing, Web delivery and presentation and, most importantly, we realised that Greenstone could not implement the OAI-PMH 2.0. At that point the working group decided to proceed with the development of a new application that would meet all requirements and specifications.

Kosmopolis digital collection management software is a simple application that has been developed in PHP using a MySQL database. While it is running, it creates and stores dynamically the appropriate document images' navigation thumbnails under a cache catalogue. Access files are automatically resized according to the user's screen resolution (800x600 – 1024x768). Furthermore, users have the option of creating dynamically a PDF version of the document for saving and printing purposes. In this case, document images are resized automatically to fit onto A4 paper using the function pdf_place_image from the PDFlib library.[49]

EVALUATION MODEL

Kosmopolis is the kind of digital collection that aims to support multiple education, research and cultural purposes. Therefore it has to be treated as an information service and we have to evaluate its use, its functionality, and finally (if possible) its effect on and outcome for the end-users. From an administrative point of view we have to define and/or forecast the real cost factors for developing, operating and maintaining the system as well as examine its viability. This approach demands an integrated evaluation framework focusing on these issues. In parallel, a detailed promotion plan has been designed aiming to reach all potential users worldwide.

In this initial phase we recorded, and are considering, a number of statistics and performance indicators such as number of accesses to collection, number of documents downloaded in total and per title, number of searches performed, downloads per various file formats (e.g. jpeg, pdf), among others.

CONCLUSIONS

The current version of Kosmopolis was launched in August 2004, with the content from three titles available to users via the Web. Since then the digital content of six more periodicals has been imported to the system, now totalling nine periodicals and about 7,500 documents. The rest of the content will be added gradually to the collection until the end of 2005. The promotion plan has already been set in action and by the end of 2006 all known potential users will be informed about the new service. The collection has been registered as an OAI-compliant Data Provider (http://www.openarchives.org/Register/BrowseSites) and it will gradually be registered with a number of known data harvesters, such as OAIster (http://oaister.umdl.umich.edu/o/oaister/).

Further reading and detailed information about the Kosmopolis project are available at: http://www.lis.upatras.gr/LIS/TELEPHAESSA/dlib_EL.shtml, where the final report of the project can be accessed (Greek only).

[49] More information on processing PDF files on the fly can be found at http://www.pdflib.com/.

GOOGLE SCHOLAR: NEW ACADEMIC RESEARCH TOOL

Adam Rusbridge, DCC Development Analyst

The academic community produces a large quantity of literature that is of interest to fellow scholars and public alike. Without explicit direction towards a resource it can be difficult to discover additional and relevant academic information. The Google Scholar service (http://scholar.google.com),

launched on 18 November 2004, aims to make this task easier.

Google Scholar differs from Google's better-known Web search services, as the results are limited to scholarly resources and Google has been granted permission to index a number of subscription-dependent resources (while the complete set of resources has not yet been announced, it is known to include the ACM, http://portal.acm.org/portal.cfm; IEEE, http://www.computer.org/publications/dlib/; and OCLC WorldCat, http://www.oclc.org/worldcat/). The full contents of restricted publications will still require a subscription for access. Now, however, at the very least an abstract will be available providing an indication of the publication's value. Proxy configurations of subscriptions[50] can enable access to many of these journals, making the process of finding and retrieving material seamless.

The service does not limit searches to peer-reviewed material, making the definition of 'scholarly' broader than some would desire. However, the service counts the number of (indexed) citations a publication has received. The number of citations received contributes to the relevance ranking a publication is given and, as the number of citations is visible, the impact a paper has had can (in part) be observed. A feature that is particularly useful for researchers is the ability to list each entry that cited the publication in question, making it very simple to explore a topic further.

The service also references a variety of printed publications. To enable access to these, Scholar links up with the OCLC WorldCat catalogue, a service that allows libraries to process, manage, share, and enable access to information resources. Correctly configured, this service integrates with a user's local institutional library catalogue, enabling resource discovery at an institutional level.

The service Google Scholar provides has previously existed for several specific fields:[51] Citeseer (http://citeseer.ist.psu.edu/) indexes computer articles and PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) medical articles. However, for academic fields without such dedicated databases, both academics and students turn to search engines to find resources and this service will help direct the community towards focused and relevant information. Hopefully, Google will have incited other search engine developers to add similar vertical search functionality to their own services. This will result in competition and, ideally, an improvement to the overall service quality.

While Google Scholar will not be a substitute for detailed and thorough research for relevant literature, it will assist many scholars to find pertinent information. Although alternative facilities have previously been able to perform this task (both on the Inter- and intranet level), the powerful indexing and filtering methods that Google provides and development of the field as a result of Google's brand association may make it easier to access more relevant knowledge than ever before.

[50] For example, access from behind a university proxy providing Athens authentication (http://www.athens.ac.uk/).

[51] It should be noted that librarians are still able to discover and make information available from a number of sources, and may be able to provide free access to information for which subscription sites may request payment. On an institutional level many libraries are developing local systems for resource sharing – these should also be examined for comprehensive academic research.

MUSIC-TO-KNOWLEDGE TOOLKIT

The International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) at the Graduate School of Library and Information Science, University of Illinois, has released the M2K (Music-to-Knowledge) Alpha 1.0 toolkit.
M2K is an Open Source Java-based framework designed to allow music information retrieval and music digital library researchers to prototype, share and evaluate experimental techniques in this field. The software downloads, installation requirements and instructions can be found on the project homepage at http://music-ir.org/evaluation/m2k along with the toolkit documentation.
M2K will be demonstrated at several of this year's conferences: 2005 ACM/IEEE Joint Conference on Digital Libraries; 2005 ACM Special Interest Group on Information Retrieval; 2005 Association for Computers and the Humanities/Association for Literary and Linguistic Computing Conference,[1] and two workshops. See the project Web page for details.

NEWS AND RESOURCES FROM JISC

The Joint Information Systems Committee (JISC) (http://www.jisc.ac.uk/) has developed several new programmes and resources during 2004 and 2005.
October 2004 saw the announcement of a Digital Preservation Programme comprising grants to the sum of £1M (around €1.43M) to nine higher and further educational institutions in the UK. The grants support digital preservation and asset management. As more institutions invest in digital materials, the need for long-term strategies for managing and preserving these resources becomes increasingly important and the JISC funding is aimed at developing the infrastructure necessary to maintain access into the future, and to encourage the amalgamation of long-term strategies into digital asset creation processes. Funded projects include the development of repository software, collaborative preservation networks, training, and assessment tools. The projects are expected to benefit digital curation across UK universities and colleges and in related initiatives such as the Digital Curation Centre (http://www.dcc.ac.uk/). More information is available from http://www.jisc.ac.uk/index.cfm?name=programme_404.
In November 2004, JISC announced the release of a new report 'Long-Term Retention and Reuse of E-Learning Objects and Materials,' which was commissioned by JISC to examine the long-term preservation and reuse of e-learning materials. The report is downloadable from: http://www.jisc.ac.uk/index.cfm?name=project_elo.

JISC COLLECTION

Online resources available from the JISC include:
The Academic Library (containing nearly 300 online book titles covering Anthropology and Development Studies; Cultural and Media Studies; Politics and International Relations, and Political Economy, available on a subscription basis: http://www.jisc.ac.uk/coll_academiclib.html).
Creative Club (an archive of 1.5 million advertisements across all media to be used in teaching and learning: http://www.jisc.ac.uk/coll_creativeclubcs.html).
Digimap (an EDINA (http://www.edina.ac.uk/digimap/) service allowing users to access Ordnance Survey products and map data for GIS and CAD applications).
New Towns Record Online (a comprehensive collection of information on the UK New Towns initiative: http://www.jisc.ac.uk/coll_newtowns.html).
The Ovid Arts Package (comprising the International Federation of Film Archives, the Bibliography of the History of Art and the Répertoire International de Littérature Musicale: http://www.jisc.ac.uk/coll_ovidartspackage.html).

Updates and links to all JISC resources are available from http://www.jisc.ac.uk/collections_latest_news_he.html. The JISC is investigating for possible future development what types of online resources are most wanted and needed by institutions.

DISPOSAL AND ARCHIVING OF ELECTRONIC RECORDS: THE DOMEA CONCEPT IN GERMANY

Andrea Hänger, Federal Archives of Germany

DOMEA, which stands for Document Management and Electronic Archiving in IT-supported business processes, is a wide-reaching German government information management requirement. The programme, also known as Konzept Papierarmes Büro (the 'paperless office concept'), was instituted by the Federal Government Co-ordination and Advisory Agency for IT in the Federal Administration (KBSt, http://www.kbst.bund.de/)[52] to assist government agencies in selecting suitable solutions or suppliers of content management, business process management and records management. It establishes the core requirements for an electronic records management system in public administrations.[53]

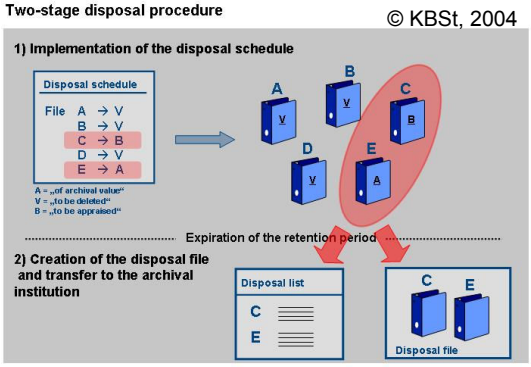
The DOMEA concept consists of three parts:

1. the organisational concept, which describes the procedures of records and business process management in an electronic environment;
2. the requirements specification, with about 250 requirements for electronic records and business process management systems;
3. and the certification for IT suppliers who wish their products to be evaluated against the requirements.

The concept forms a baseline that sets out the minimum standards necessary for a reliable electronic records management system. Each administration wishing to make use of these requirements will always need to consider its own specific business needs and context. The generic requirements must be tailored by adding special needs and selecting from alternative requirements according to corporate policy and practice. The concept describes the whole business process starting from the incoming letter, email or report, through its processing to its disposal.

Since 1996 DOMEA has become the most important guideline for implementation for electronic records and IT support of business processes. The rapid technological development and a considerable amount of new practical experience led the KBSt to review the concept. In September 2003 the draft of the revised concept was published on KBSt's Web site, where it could be discussed and commented on by the interested public. As well as the consideration of new technological developments, the new concept is more orientated towards the different steps of the business process, whereas the first concept focused especially on the implementation strategies. It shows how to manage records securely throughout a controlled lifecycle in compliance with relevant regulatory requirements.

The new concept is no longer focused on the federal administration, but integrates the requirements of the regional and local administrations. The first DOMEA concept was accompanied by a recommendation for the disposal and archiving of electronic records in federal agencies, published in September 1998.[54] The concept establishes a very detailed procedure for the disposal and archiving of electronic records. This procedure was never realised in practice, largely due to the fact that electronic records and business process management systems have only recently been implemented in the agencies or are still in the introductory phase. Even if agencies are already working entirely electronically, the retention periods have not yet expired, whereas the concept provided for a significantly shorter period of implementation.



In the framework of the re-adaptation of the DOMEA concept in January 2004, a new project group was set up to work on a revision of the disposal and archiving concept. Under the direction of the Federal Archives and with the assistance of the KBSt, a group of federal, regional and local archivists and consultants developed a new concept, which was successfully submitted to the German archival community in September 2004[55] and is applicable to every level of public administration.

---

[52] The Federal Government Co-ordination and Advisory Agency for IT in the Federal Administration (KBSt) was founded in 1968. Located within the Federal Ministry of the Interior, the KBSt is an inter-ministerial agency of the Federal Government intended to ensure that the federal administration optimises its use of information technology for specific fields and in organisational, economic and technical terms. The Federal Government's goals for IT use are to link IT strategically with other administrative areas such as management and organisation, to ensure that this complex technology remains controllable and manageable in the future, and to combine IT resources using new organisational approaches and the most comprehensive and uniform IT standards possible. To perform its mission, the KBST develops framework rules for the use of IT within the federal administration. Upon approval by the ministries in the Inter-ministerial Co-ordination Committee for Information Technology in the Federal Administration (IMKA), the KBSt publishes such rules as general recommendations.

[53] For the description of the first DOMEA project, see Andreas Engel and Michael Wettengel, "From project to practice, the DOMEA-project in Germany" in *ARMA*, Vol. 37, No. 4, pp. 49-56, July/August 2003.

[54] See Koordinierungs und Beratungsstelle der Bundesregierung für Informationstechnik in der Bundesverwaltung (KBSt). Konzept zur Aussonderung elektronischer Akten. Teil 1: Empfehlung des Bundesarchivs zur Aussonderung elektronischer Akten. Teil 2: Erfahrungen zum Aufbau und zur Ablage elektronischer Akten im DOMEA®-Projekt, Bonn 1998 (Schriftenreihe der KBSt, 40).

[55] The draft of the new concept is available on the KBSt Web site (http://www.kbst.bund.de). An English version will be available in 2005.

The recommendations will facilitate appraisal in an electronic environment and enable federal, regional and local agencies to meet their obligation to transfer non-current electronic records to archival institutions. The new concept describes all phases of the lifecycle through which documents typically pass, including creation, authoring, approval, release and ultimately revision or retirement and archiving. It defines the procedures and metadata that have to be implemented to make sure that the disposal can be managed as automatically as possible. A two-stage disposal procedure is recommended. In an initial stage, the archival institution delivers a disposal schedule. This establishes a timetable for the lifecycle of the records as provided for by legislation, regulation or administrative procedure, and provides authorisation for their disposal on the basis of the filing plan of the transferring agency. The schedule is implemented in the electronic records management system of the agency. For automatic selection of records, the metadata of documents, folders and files have to contain a *datafield disposal instruction* which can have the entry either 'to be deleted', 'to be appraised' or 'of archival value'. Information should be registered at the file level and 'inherited' downwards to all folders and documents created subsequently under that particular file.[56] On the basis of this information, in the second stage the agency automatically selects those files whose retention periods have expired and which are of archival value or have to be appraised, and transfers them to the archival institution.

The effectiveness of this two-stage procedure depends on the quality of the disposal schedule. Therefore the concept recommends that the archival institution should have a limited access to the records management system of the agency in order to review and continuously update the disposal schedule.

The concept defines a set of mandatory metadata that describe not only the documents, folders and files, but also the filing plan and the electronic records management system. This set of metadata has to meet the different corporate policies and practices on the federal, regional and local level as well as being sufficiently detailed and concrete to meet the needs of the IT suppliers. It stresses the importance of the audit trails that automatically capture and store all the actions performed on an electronic record, the user initiating and carrying out the action, and the date and time of the event, which have to be assigned to the records transferred to the archives. The new concept also defines a new metadatum, a time limit for the conversion. It makes sure that the conversion to an archival format takes place shortly after a file is closed, and not only at the time of the transfer. Given the current length of retention periods (around 30 years or more), this provision is absolutely necessary to prevent a loss of data.

The new concept will not only describe the procedures; it will also offer a practical solution for transferring principles into practice. The set of metadata will be transformed into an XML schema and implemented by IT suppliers in electronic records management systems. This will be part of the requirements of the new DOMEA certification in 2005. The aim is to standardise and, as a consequence, accelerate disposal procedures by making the electronic records management systems of public administrations interoperable with archival management systems. Realising this interoperability will be the major challenge in the years to come.

[56] The DOMEA concept is based on a three-level object hierarchy for documents, folders and files. Files are entered in a corporate information structure – in German administrational tradition the filing plan. On every level, metadata are defined. Their application follows the principle and functionality of 'inheritance' from the higher object to the lower. Filing a document in a folder or creating a new folder is analogous to associating it with the relevant part of the corporate information structure.

AFTER ERPANET

Georg Büchler, ERPANET Swiss Editor

'How does it feel, after ERPANET?'

'A bit sad, I would say. Knowing that the project's future wasn't fully secured, and that the close collaboration with a bunch of motivated, cheerful and bright people – my erpaColleagues – would come to an end, I couldn't help feeling melancholy. Still I was happy to see that we managed to keep our quality and even to increase our pace on the last leg... and I was lucky to meet most of my colleagues and directors again in the course of our final events. For we rarely actually met, although we were working in a constant and very close collaboration. Spread all over Europe, located in Glasgow, The Hague, Urbino and Bern, we got to know each other's qualities and style of work very well, and quick assistance was always at hand in the case of problems. And I think this "distant teamwork" was what impressed me most about the project.

'There's a second thing I really loved about working with ERPANET, and that's the young and fresh flavour of digital preservation. Imagine the newly graduated classicist I was in April 2002, flying to Glasgow for the content editors' kick-off meeting and ready to jump into a subject area I barely knew anything about. And imagine my relief when, during the first events I attended, I discovered that a good number of digital preservation professionals I met didn't have much more experience than myself. Granted, there were the pioneers that have been working in the field for many years (you know the names), and there were many archivists or computer scientists. But I also got to know people with backgrounds in history, archaeology, linguistics or natural science, often working for projects as young as ERPANET. No need to feel like an outsider. And while I always found it very hard to find a German equivalent for the term "digital preservation community," I completely understood its meaning. I am particularly happy to have made ties with this young and exciting community, and I am looking forward to continuing working with some of its members.

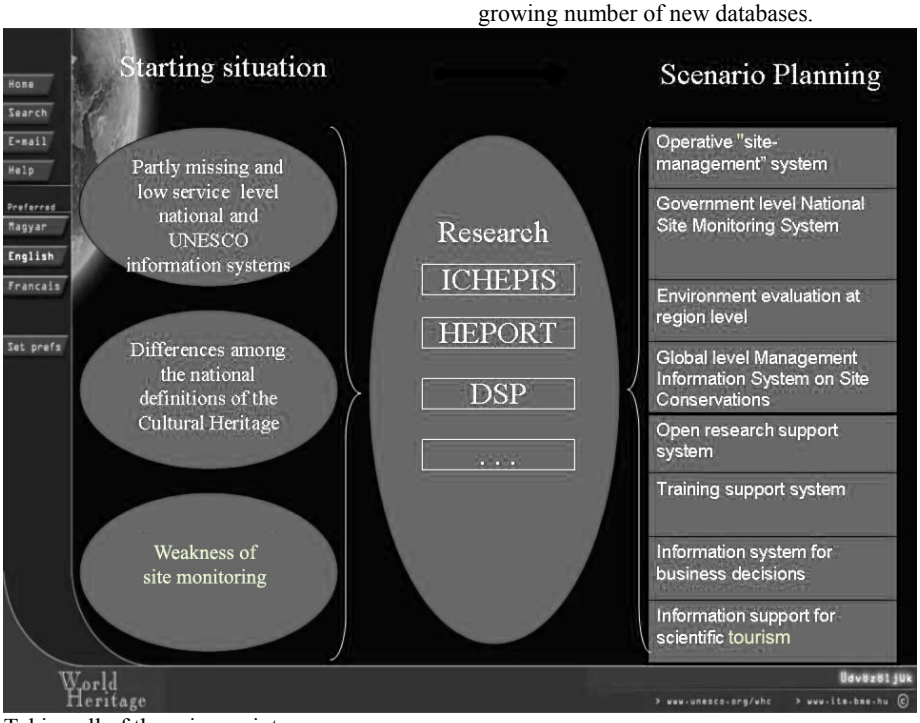
'So, what comes next? I myself will continue to be active in digital preservation. Just two weeks after my contract with ERPANET ended I resumed work at the Swiss Federal Archives (http://www.bundesarchiv.ch). I am now working on behalf of the archives of the Swiss Cantons that have joined forces to implement a coordination agency for digital preservation. As part of a two-person team I will assist them in coordinating their efforts to find and implement suitable preservation solutions for their holdings.

'As for ERPANET, just keep your eyes open. A number of events are forthcoming, and the European Commission seem to be quite keen on a follow-up project. ERPANET won't go away, nor should it. Looking back on the last three years, I am proud that we managed to make it an established brand in digital preservation.'

DIGITISATION AND ARCHIVING FROM A TECHNOLOGICAL PERSPECTIVE

Dr K. Ferenc, Budapest University of Technology and Economics

Cultural heritage research activities have been taking place at the Department of Information and Knowledge Management of Budapest University of Technology and Economics (http://itm.bme.hu/DesktopDefault.aspx?tabindex=0&tabid=1) since 1997. Our fields of research cover information management, IT support of heritage preservation, monitoring of cultural and natural heritage using remote sensing and direct IT-based measurement and data collection technologies, decision support for heritage protection authorities, applications of state-of-the-art IT, and knowledge management technologies. During this time, our research identified many technological, methodological and economic questions, which are summarised in the figure below.



Taking all of these issues into consideration, the Department of Information and Knowledge Management decided to set up a series of research projects to develop solution models and pilot services.

The first large research project was ICHEPIS (Inventorisation for a Cultural Heritage Periodic Information System, http://itm.bme.hu/DesktopDefault.aspx?tabindex=0&tabid=40), which involves the Department working as institutional partner in cooperation with the Hungarian Office for Cultural Heritage, the UN World Heritage programmes in Hungary and the Royal Angkor Foundation (http://www.angkor.iif.hu/). The direct objective of the ICHEPIS project was to define the scope and structure of different datasets that can provide flexible and user-friendly information and knowledge management solutions for all States Parties to the Convention Concerning the Protection of the World Cultural and Natural Heritage (1972, http://whc.unesco.org/world_he.htm), as well as for the World Heritage Centre of UNESCO (http://whc.unesco.org/) and the Advisory Bodies to the Convention. It was necessary to analyse how far the available public databases met these requirements. In addition, proposals had to be structured so that they could integrate the continually growing number of new databases.

During research, effective methods had to be found for integrating all text documents containing valuable scientific data into a data system and making them easily available and searchable. In connection with the amalgamation and standardisation of data, a review of the basic thesaurus of relevant areas and organisations had to be integrated, as far as possible, into a flexible Internet-based structure, to provide a means of information retrieval for more effective use of the existing knowledge base.

It is also of great importance to understand the utilisation of data such as remote sensing images, including the wide range of public data already accessible on world heritage sites. Thanks to the growing interest of space agencies, an increasing number of States Parties to the Convention and international research institutes (including Hungarian organisations) have been collecting data on world heritage sites for more than a decade, in order to track the state of their own and other cultural and natural world heritage sites on a regular basis. In addition to the crucially important legal concerns regarding the authenticity of collected data, technical problems also require coordinated research, as at present there is no approved methodology or procedure for such activities.

Eventually, our research led to the production of integrated data models, a structure for training and education, collection, distribution and use of such data, as well as recommendations for data migration. States Parties should be capable of creating comprehensive sites, regional or even thematic collections, while simultaneously learning and creating data processing and knowledge management standards, as well as developing rules to update information and provide metadata services on their own sites. These models should provide a suitable basis for approval by the individual States Parties' World Heritage Information and Knowledge Management System to be developed on a step-by-step and incremental basis. Such systems will enable the States Parties to recognise any relevant changes to the data, including those that point to an improvement or deterioration (perhaps disastrous) of world heritage sites in due course. It will also enable them to attract partners for mutually beneficial cooperation in the protection and development of these sites.

HUNGARIAN GOALS FOR NATIONAL CULTURAL HERITAGE INFORMATION RESEARCH PROJECTS

Nowadays more and more research centres, authorities, civil organisations, councils and private citizens are demanding access to an integrated, centrally maintained yet decentralised database, which contains updated information regarding the state of national cultural and natural heritage. A

new concept and new approaches were required to create such a system.

The most important goal of our research and development project is to produce a technologically sound, professionally accurate and internationally acceptable base model and data sample. This model is required to include all relevant Hungarian cultural and natural heritage components, and at the same time must meet the existing and evolving international standards of the information society.

In the course of research we will explore which basic elements of cultural and natural heritage should be included in the site data, and in the national, regional and international systems. The scope of data to be recorded to populate the agreed elements will be determined in this initial phase, along with the points of connection to additional databases (a theoretical solution to ensure data authenticity and safety). Another research task is to explore the likely scale and scope of potential users of the heritage database, and to determine and register users' changing and evolving demands for such databases. In the course of this project's development, a model database has been created, which is able to register the specified components and serve potential users in terms of functionality.

We intend to design and build the database in such a way as to ensure that it can store data that can be made immediately accessible to all levels of potential users on site, on national and international levels, for a variety of purposes, such as scientific research and advocacy. Thus, it will be possible to conduct research in this field with an unprecedented scope. Most of the required information technology infrastructure is available; however, it was necessary to establish a modern training facility during the project for those who wish to learn to use the technology for their own benefit.

By uniting two closely related aspects of the project, i.e. national and international standards, together with research and training demands, it is possible to achieve another important aim of the project, namely to encourage international cooperation and collaboration.

The first pilot software demonstrating our suggested model is under development by the HEPORT project (http://itm.bme.hu/DesktopDefault.aspx?tabindex=0&tabid=82); the current version is accessible online from http://itm.bme.hu/unesco.

During the ICHEPIS project, around 30,000 pages of official documentation were digitised and processed with OCR software.

© HATII, University of Glasgow, 2004

These papers contained data in 70 languages and involving typed, printed and handwritten text. During this subtask we had to address the well-known problems of managing long-term storage: which media and which file format should be used for the results, and issues related to the lifecycle and durability of these digital archives.

As a result of these challenges, the next project was called Digital Stoneplate (DSP), representing the apparent conflict between the durability of the ancient stoneplates and the relatively (or perceived) short lifecycle of the digitised databases. The goals of the Digital Stoneplate project are:
to define a methodology and workflow for digitisation to build long-term electronic document storage and retrieval processes;
to define a methodology for long-term data storage in technical (e.g. storage medium, migration strategy, file formats) and logical (e.g. structure, metadata, update handling) terms;
to develop the fundamentals of understanding of these topics for staff working in the target organisations, educating them in digitisation issues, technology and strategies.

Due to the current pace of technological development, it is unrealistic to think about archiving and digitisation today in the same terms that we have historically considered them. However, we are convinced of the fact that this issue cannot be simplified to a merely technological matter. It is very likely that we – due to the expected short lifecycle of storage formats – will be forced to opt for repeated migration, but upon accepting this constraint we also need to take into account the economic aspects of maintaining the data. Our further ambitions include the implementation of an archive interface for the digitally created documents, allowing easy access to, and updating of, data. This will help the data to remain useful during the period of development and migration.

Overall, the project can be defined as an optimisation task that may outline recommendations for practice in digitisation and archiving, and bring together technical, technological and economic points of view, integrated into a coherent model system.

AUSTRIAN DIGITAL PRESERVATION EVENT

9 March 2005 saw a meeting in Vienna of experts on long-term preservation in the digital era, organised by the Austrian Commission for UNESCO (http://www.unesco.at) and the National Library of Austria (http://www.onb.ac.at/). The meeting discussed UNESCO's *Charter on the Preservation of the Digital Heritage* (http://portal.unesco.org/ci/en/ev.php-URL_ID=1539&URL_DO=DO_TOPIC&URL_SECTION=201.html) and outlined steps for the development of a standardised strategy for preservation of digital content in Austria. The Charter, adopted in 2003, meets the challenge of preserving digital heritage, at risk from technology obsolescence, sporadic funding leading to uncertainty about resources, the changing issues of curation and maintenance of digital heritage, and the current lack of legislation for supporting preservation activities.

The Charter states:

> "Attitudinal change has fallen behind technological change. Digital evolution has been too rapid and costly for governments and institutions to develop timely and informed preservation strategies. The threat to the economic, social, intellectual and cultural potential of the heritage – the building blocks of the future – has not been fully grasped".

INTERPARES 2: A PROGRESS REPORT

Shaunna Moore and Luciana Duranti

In the very first issue of *DigiCULT.Info* in 2002, Luciana Duranti spoke to us about her work with InterPARES. For DigiCULT.Info's 10th issue, we asked for an update on this work.

Digital records only exist in the status of original at the moment when they are made or received. After a digital record has been saved for the first time, each retrieval produces a copy. These copies are still the creator's records, because they are used for action or reference in the usual and ordinary course of business of the person who makes, receives and accumulates them. However, they can only be considered reliable if the processes of creation and maintenance have been subject to strict controls and have respected authenticity requirements. When the creator's records are no longer active and are acquired for long-term preservation by some trusted custodian, we are confronted with a greater intellectual problem. In fact, it is not possible to preserve a digital record. It is only possible to preserve the ability to reproduce it as an authentic copy. Because copies made for purposes of preservation do not participate in the creator's course of business, they are no longer its records, although they may be endowed with the same reliability if the process of reproduction has been carried out by a competent person, following strict rules and properly documented.

Analysis of the relationship between authenticity and reproduction has always placed theorists in a precarious position. As the definition of authenticity includes both identity (the entity is what it purports to be) and integrity (the entity has not been corrupted), how can reproduction be undertaken in such a way that these qualities can be transferred from the original to the duplicate? Considering that a copy of an entity is never identical to its first instantiation, both facets of authenticity are immediately threatened upon attempts to imitate an original. As a result, the verification of authenticity in a reproduction necessitates a thorough understanding of the entity that is being reproduced in order to determine which specific parts, characteristics, attributes or formal elements are crucial to its identity, and to what extent alteration is acceptable.[57]

Nowhere is this more apparent than in the performing arts, which by their nature require the repeated presentation of interpretations of scores or scripts by different performers who execute the original idea in ever-new environments. In this context, the parameters of authenticity are unclear, for even in its original form a performance relies on a multitude of interactive components, which are never entirely determined by the creator of the work. The factors deemed relevant for the authentic reproduction of a performance will vary according to the circumstance. For instance, while some will consider the use of original musical instruments of the period to be of utmost significance in the performance of a baroque piece, others will focus on the articulation of notes, interpretations of dynamics, or even the ambience of the environment when evaluating whether the performance captured the essential quality of the composer's original score.

The concept of repetitive performance serves as a useful analogy in an electronic environment. The existence of entities that are fragmented into their digital components upon every storage action and reconfigured upon every viewing necessitates an in-depth analysis of the extent to which each reconfiguration may be deemed identical to the original entity. Clearly, this entails some degree of compromise with regard to the meaning of integrity, for, how can an object that has been torn apart and reconstructed truly be described as uncorrupted? The issue becomes further complicated by the highly interactive and dynamic nature of the modern digital environment. Authenticity in this case must take into account the ability to preserve some semblance of shifting contexts and relationships among and within digital objects.

INTERPARES 2

InterPARES (International research on Permanent Authentic Records in Electronic Systems, http://www.interpares.org/ip2_index.cfm) has been devoted to the development of the theoretical and methodological knowledge that will provide a framework for the long-term preservation of authentic records in electronic systems. As an international and interdisciplinary project,[58] the goal of the research is to ensure that society's recorded memory can be created in an accurate and reliable form, and be authentically maintained and preserved, for short- and long-term use by its creators and by society at large. The first phase of the project examined records created in databases and document management systems, whose preservation was mandated for the accountability and administrative requirements of organisations. InterPARES 2, which began in 2002 and will be completed in 2006, has focused on digital records produced in dynamic, interactive and experiential systems in the course of artistic, scientific and electronic government activities. Dynamic records are those whose content is dependent on continuously varying data, held in several databases and spreadsheets, while interactive records are created and maintained in interactive systems in which each user's entry instigates a response from the system, or triggers a particular action. Experiential records are defined by an essence that goes beyond the bits that constitute the digital object to incorporate the behaviour of the rendering system, or at least the interaction between the object and the system. This second phase of the project has necessitated a reassessment of some concepts in the context of systems in which fluidity is intrinsic to the nature of the records and essential to the accomplishment of their purposes as support for the actions of their creators. The issues revolving around the preservation of such digital objects are more complex than have been

[57] DigiCULT's first *Thematic Issue* deals with issues of integrity and authenticity in detail. The publication is free to download from DigiCULT's Web site at http://www.digicult.info.

[58] Researchers have backgrounds in: Archival Science, Chemistry, Computer Engineering, Computer Science, Dance, Diplomatics, Film, Geography, History, Information Studies, Law, Library Science, Linguistics, Media Studies, Music, Performance Art, Photography and Theatre. The countries actively involved are: Canada, United States, Australia, Belgium, China, France, Ireland, Italy, Japan, Netherlands, Portugal, Singapore, Spain and the United Kingdom. The Advisory Board also includes an archivist from South Africa.

encountered with traditional electronic records because the authenticity of their rendering is dependent not only on the objects themselves, but on the reproduction of their relationships to the environments in which the objects are experienced and acted upon.

Although based on interdisciplinary methods and approaches, the project is intended to be archival in purpose; thus, the results of the research activities carried out across disciplines are constantly translated into archival terms, associated with archival concepts, and developed into archival methods. In other words, the ultimate goal of the project is to articulate the characteristics, components and requirements of a trusted system for the creation and maintenance of digital records, and a preservation system that ensures the authenticity of their authentic copies over the long term. Nonetheless, upon completion of the research, the outcomes will be translated back into the language of each discipline so that these systems will be made comprehensible and accessible to records creators in the fields involved with the project. InterPARES researchers have retained the traditional archival definition of a record as any document created (i.e. made or received and set aside for action or reference) by a physical or juridical person in the course of a practical activity as an instrument and by-product of such activity. However, the characteristics established as essential for identifying the records and situating them in context have had to be reconsidered in the light of the nature of the digital objects under investigation.[59] For instance, fixity of form and stability of content are not characteristics of experiential, interactive or dynamic records, and the challenge has been to determine methods for preserving these records either in their dynamic state or in some alternative fixed form and stable content developed by the creator for the purposes of its activities.

CASE STUDIES AND GENERAL STUDIES

Research has been developed around a multi-method design, based on the methodology and tools considered most appropriate by each investigating team for the specific research activity it carries out. A wide range of case studies from the three focus areas (arts, science and e-government) forms the basis of the investigation of current practice with regard to the creation and maintenance of experiential, interactive or dynamic records. Examples include: *Obsessed Again…* an interactive electronic musical work written by Canadian composer Keith Hamel; the geographical information system developed and used by the Centre for Desert Archaeology in Tucson, Arizona; the Alsace-Moselle land registry system; the Antarctic Treaty Searchable Database; and the work of Stelarc, a multimedia performance artist who integrates digital media in his works performed in diverse environments. Each of the three focus areas has presented specific challenges based on the ways in which records are viewed within the particular context, the types of actions that the records are supporting or in which they participate, and the emphasis placed on representation of the records in their original form. For instance, in the arts, both visual and performing, authenticity is conceptualised as the ability of the digital object to reflect the artist's original intentions with regard to both content and form upon every subsequent presentation of the work. In the realm of science, creators are more typically concerned with ensuring the accuracy of data entered into, and maintained in, the systems than with the form of their aggregations, although the authenticity of scientific records is crucial to ensuring the integrity of the data within them. In e-government, accountability is of greater concern than in the other two focuses, but this often leads to confusion over the difference between authentication (a declaration of authenticity at a single moment in time) and the authenticity of the records through time. To date, eleven of the twenty-one case studies have been completed and represented in activity and entity models, and half of these have been analysed according to diplomatic principles in order to gain an understanding of the boundaries of the records contained in each system and to determine their essential characteristics.

To complement the case studies, general studies have been undertaken to address the issues relevant to each of the three areas of records creation. Examples of these include surveys of government Web sites, of the record-making and recordkeeping practices of digital photographers, composers and film makers, of the practice of preservation of interactive music, and of file formats and encoding languages used for non-textual materials, as well as a study of the San Diego Supercomputer Center's project to develop a prototype for persistent archives based on data grids for the National Archives and Records Administration (NARA). At the same time, three teams of researchers responsible for investigating the key concepts of accuracy, reliability and authenticity as understood within the disciplines encompassed by each of the focuses have produced annotated bibliographies and literature reviews, undertaken analysis of these concepts as discussed in the literature, and created bibliographic databases for the management of references.

TERMINOLOGY

A terminology research group has been working towards the standardisation of the vocabulary within InterPARES, so that all researchers and research assistants will be consistent in their usage of terms, thereby ensuring effectiveness of communication within the project, and consistency in dissemination activities. The work of the team has focused on the creation of lexicographic instruments: a Register of all terms and phrases used within InterPARES; a Dictionary including the definitions provided for those terms by all the disciplines involved in the project; a Glossary including the definition for each term chosen by InterPARES as the authoritative meaning for all the project's documents; and a Thesaurus that assigns the terms to a specific facet (e.g. agent, action, object, event, property), links it to other facets through hierarchical, equivalence or associative relationships, and points to the preferred term. The terminology group has worked closely with all other research units in order to ensure the greatest inclusivity while maintaining consistency and rigour of communication, and to keep the delicate balance between accurately describing the nature of concepts used across the various disciplines and prescribing a common language for the research project as a whole. The group is particularly aware of the political nature of its endeavours and of the partiality and bias potentially involved in the prescription of terms and the standardisation of meaning – a big challenge if one considers the encompassing nature of the research, both in terms of the diversity of cultures involved, and of the multiplicity of meanings often associated with terms used in the vast range of fields incorporated within the project. At present, the terminology research is focused on the evaluation of its lexicographic instruments, in order to assess their compliance with international standards, and their ability to meet the needs of external users. Future goals centre on the promotion of public access to the terminology database in order that researchers will be more aware of impediments to effective communication and of possibilities for increased

[59] InterPARES 1 identified the following characteristics of a record: (1) a fixed form, meaning that the entity's binary content must be stored so that it remains complete and unaltered, and its message can be rendered with the same documentary form it had when first set aside; (2) an unchangeable content; (3) explicit linkages to other records within or outside the digital system, through a classification code or other unique identifier; (4) an identifiable administrative context; (5) an author, addressee and writer; and (6) an action, in which the record participates or which the record supports either procedurally or as part of the decision-making process.

consistency and understanding among disparate fields of study.

MODELLING

The project has incorporated the use of modelling as a means to conceptually represent the activities and entities under investigation in order to assist in the analysis of case study data and to communicate findings related to creation, maintenance and preservation of the records under examination. The major efforts of the modelling team have been channelled into the development of a model which integrates the appraisal and preservation activity models from InterPARES 1 with the activity model produced by a previous research endeavour, commonly known as the UBC-DoD research project.[60] The amalgamated model (Manage the Chain of Preservation, or MCP model) depicts all of the activities involved in the management of electronic records throughout their lifecycle, from creation to permanent preservation. The model presents a clear visual representation of the ideal management process from the point of view of the preserver, providing the conceptual framework in which the realities of current practice may be measured. The model assists all other units of research by providing a means to identify preservation issues related to the case studies (which are walked through the model and validate it), and to assess the points of the records' lifecycle that require development of policies, procedures and standards (including metadata schemas and reproduction requirements). The group is currently engaged in testing the model and has begun to develop an activity model of preservation from the viewpoint of the creator.

POLICY

The policy research team has endeavoured to produce a conceptual framework for the development of policies that will impact each of the three focuses of investigation. The major area of enquiry centres on the identification of barriers to preservation that currently exist in laws, regulations, policies and standards on copyright and intellectual rights, privacy and freedom of information, authenticity and authentication, open standards and Open Source, and records and archival management. Legislation and standards originating in Canada, the United States, Australia, Singapore, Hong Kong and the European Union (as well as specific countries within it), which have a bearing on records creation, maintenance and preservation, are being analysed within the conceptual framework in order to formulate model policies and standards. Case studies have aided in the identification of areas in which policies are absent or insufficient to ensure the preservation of authentic electronic records created in the arts, science and e-government. Preliminary findings have indicated the need to return to the articulation of the concept of record within each environment – an endeavour that is currently a major focus of discussion among the researchers. A series of model policies and guidelines will then be able to ensure that the entities identified as records are managed appropriately throughout their lifecycles. One objective of this team is to establish liaisons with policy and standards-issuing organisations in order to ensure their effective implementation.

DESCRIPTION

The main objective of the research team focusing on the description of digital entities has been to evaluate existing and emerging metadata schemas, descriptive standards and metadata tools to determine their ability to meet requirements relating to the creation, management, appraisal, preservation and use of reliable and authentic records in the three areas. The purpose of this endeavour has been to recommend either the development of new schemas, standards and tools, or the extension of existing and emerging ones.

The resultant activities of this research group have consisted of a number of specific projects. The team has begun the analysis of the completed case studies in order to identify the various uses of metadata and other descriptive standards within the diverse spheres of activity for the purposes of identifying existing practices, setting a framework for the delineation of best practice, and determining those areas in which current standards are unable to meet the requirements identified. A specific product which is intended to both facilitate this research and be a direct deliverable of the analysis is a Metadata Schema Registry. The registry is a centralised repository of schemas, which will aid various professions and organisations to identify metadata sets, or the combinations of elements from several sets that are appropriate to serve their recordkeeping needs. As it currently exists, the database has nearly completed the design phase of its development, with implementation taking place in the summer of 2005. At present, approximately ten of the most relevant metadata schemas have been analysed with reference to ISO recordkeeping standards, InterPARES 1 requirements for the presumption of the authenticity of records, and the Australian Recordkeeping Metadata Schema, while a number of other schemas have been identified for registration and analysis.

Initiated by the description cross-domain, but built upon by all InterPARES research teams, a literary warrant database has been developed in order to facilitate the identification of authoritative sources relating to the ways in which metadata and archival description support record creation and record preservation activities. This database has been implemented and is currently being populated with references from these sources. The description team is also working closely with the modelling group to identify the requirements for the creation of metadata, the specific points within the MCP model at which they must be created, and by whom.

CONCLUSION

The large quantity of research material already generated by InterPARES 2 has effectively laid the groundwork for the last two years of the project, which will focus on the development of its deliverables – namely, guidelines for records creators and preservers, prototypes of appraisal and preservation systems, frameworks for the development of policies, strategies and legislation, descriptive schemas for digital entities, an existing metadata schema registry, and literature and terminology databases.

As mentioned previously, the experiential, interactive and dynamic nature of the digital entities under investigation has necessitated a re-examination of the conceptual foundations of archival theory, the meaning of authenticity in such a technological environment, and what it means to preserve digital records by reproducing them in an authentic way. To return to the analogy of the performing arts: in that context, preservation has traditionally been accomplished by keeping the script or the score, which, to varying levels of detail, specifies the creator's intention of how a work is to be reproduced. Due to the fragility of digital media and the rapid rate of obsolescence of digital technology, the current hypothesis is that the electronic records investigated in InterPARES 2 will require increasingly detailed scripts in order to ensure their authentic reproduction. Identity metadata inextricably bound to the data in the record, linked to a model of its form and a description of its digital environment (i.e. record functionality and system documentation), stored in a trusted recordkeeping system, will potentially be the only means for preserving the authenticity of records, thereby ensuring their continuing reliability and accuracy. While explorations of these processes and systems are still under way, to a large

[60] Luciana Duranti, Terry Eastwood and Heather MacNeil, *Preservation of the Integrity of Electronic Records* (Dordrecht: Kluwer Academic Publishing, 2002).

extent the project has already been tremendously successful, as it has promoted the collaboration of scholars and professionals from a range of cultural and disciplinary backgrounds, devoted to the preservation of their digital culture.

Trends in Digital Repository Management

The PREMIS Working Group (Preservation Metadata: Implementation Strategies, http://www.oclc.org/research/projects/pmwg/) released a report in September 2004 entitled *Implementing Preservation Repositories For Digital Materials: Current Practice And Emerging Trends In The Cultural Heritage Community*. The report presents the conclusions of an international survey on existing and future repositories for digital materials. The survey focused on current practices in managing metadata in digital archives, but also covered such key aspects as preservation strategy, funding, mission, and access policy. Discussion at the end of the report identifies emerging trends within the sector.

PREMIS is sponsored by OCLC (http://www.oclc.org/) and RLG (http://www.rlg.org/). The full report can be downloaded from http://www.oclc.org/research/projects/pmwg/surveyreport.pdf.

JISC Digital Repositories Programme

The Joint Information Systems Committee (JISC) (http://www.jisc.ac.uk/) has announced a programme aimed at facilitating the set-up and use of digital repositories by the UK's education and research communities. The programme will consider how repositories interact at an institutional, national and international level; the range of educational content held in digital repositories (for example, ePrints, learning objects, technical reports, multimedia and datasets); and the informal use of repositories. It will focus on:
projects which explore the differing roles of digital repositories (especially where the repository brings together different groups of people);
new technologies and tools;
pilot services, for example supporting information discovery, or services that could be shared in many different repositories;
the development of standards, specifications and frameworks;
supporting studies, such as evaluation, or related subject areas, e.g. IPR or data integrity and authenticity.

Related work and programmes by JISC include: FAIR (Focus on Access to Institutional Resources, http://www.jisc.ac.uk/index.cfm?name=programme_fair); X4L (Exchange for Learning, http://www.jisc.ac.uk/index.cfm?name=programme_x4l); MLEs for Lifelong Learning (http://www.jisc.ac.uk/index.cfm?name=programme_mle_lifelong2); Digital Libraries and the Classroom (http://www.jisc.ac.uk/index.cfm?name=programme_dlitc); Linking Digital Libraries with VLEs (http://www.jisc.ac.uk/index.cfm?name=programme_divle); and Digital Preservation and Asset Management in Institutions (http://www.jisc.ac.uk/index.cfm?name=funding_circular4_04).

Funding available through the programme is expected to total around £2.5M (€3.615M) per year, which will be available to proposals from UK HE institutions and their partners.

Open Access to Research

During 2004, a report *Scientific Publications: Free for All?*[1] was published in the UK to address the fact that scientific research is typically published through specialist journals, to many of which individuals and even academic libraries cannot afford to subscribe. This obvious obstacle to the dissemination of research findings both within and beyond the scientific community is a growing concern in the sector, and many have investigated open access as a solution.[1]

PubMed Central (http://www.pubmedcentral.nih.gov/) is a digital archive of life sciences and biomedical journal materials that is free to access online, funded by the USA's National Institutes of Health. At present, the National Library of Medicine (http://www.nlm.nih.gov/) is working with the Wellcome Trust (http://www.wellcome.ac.uk/) to create a European version of this resource. It is foreseen that depositing research articles in the online archive will become a requirement for those benefiting from Wellcome Trust research grants, and that additional funding will be supplied in order to cover the cost of producing PubMed-compliant metadata and the charges of open access publishers such as the Public Library of Science (http://www.publiclibraryofscience.org/) and BioMed Central (http://www.biomedcentral.com/).

More information is available from http://www.wellcome.ac.uk/doc_WTX022826.html and http://www.wellcome.ac.uk/doc_WTX022412.html.

THE BIRTH TELEVISION ARCHIVE: THE RESEARCH AND DELIVERY PORTAL FOR AUDIOVISUAL HERITAGE

Johan Oomen, MA, Netherlands Institute for Sound and Vision

The BIRTH Television Archive (http://www.birth-of-tv.org) is an innovative Web portal providing uniform access to digitised audiovisual material. Major European broadcast archives and specialised ICT companies joined forces to set up the basic infrastructure. Distributed content from various sources can be accessed from one central access point. Apart from moving image material, the BIRTH Television Archive also provides access to digitised programme schedules, stills, articles and much more. Particular attention is given to providing language-independent search possibilities and to offering the option to compare the different development paths in several countries across Europe.

The first version of the portal was launched earlier this year; a second release is planned for November 2005. The BIRTH Television Archive aims to encourage new institutes to join and contribute content or context information. The MEDIA PLUS programme of the European Commission provided funding during the first 36 months of the initiative.

INTRODUCTION

Gaining access to audiovisual material has been a fairly frustrating endeavour until recently. The material was stored in archives that were generally unwilling or unable to match the needs of external users. Viewing capacities were very limited and mostly only available at the archives' premises, due to the various formats used. Copyright issues made it difficult to show the material publicly. Furthermore, due to the various legislations in different countries, international exchange of audiovisual material was scarcely possible.

From a scientific point of view the audiovisual material often lacked a specific relevance as documentation was often poor and it was difficult to check the original sources – knowledge indispensable for any scientific source. As a result of these various obstacles, the scientific community itself has only begun in recent years to develop new methodological approaches to audiovisual material and this process is not yet over. The major aim of BIRTH was to tackle these problems where possible and to provide a wide user group with seamless access to a collection of meaningful and interrelated content.

The BIRTH Television Archive started in 2003 and brings together the original contributing archives: BBC (UK, http://www.bbc.co.uk), Sound and Vision (Netherlands, http://www.beeldengeluid.nl), ORF (Austria, http://www.orf.at/), RTBF (Belgium, http://www.rtbf.be), SWR (Germany, http://www.swr3.de) and technical partners Joanneum Research Graz (Austria, http://www.joanneum.ac.at), Hagenberg University of Applied Science (Austria) and Noterik Multimedia (Netherlands, http://www.noterik.nl/).

Each of the participating broadcast archives has hundreds of thousands of hours of audiovisual material in their collections, an invaluable resource for study and entertainment and a rich source for new creative outlets. Archives such as these are living organisms, growing at a steady pace every day.

Current developments in technology have proved the vision of continuous access to archive content to be feasible. During the past few years, media production companies have been transforming their tape-based production chains into fully digitised processes. Specialised so-called media asset management systems are gaining ground and playing a major role in transforming the archiving process; furthermore, they are key to providing ubiquitous access over networks. These changes in back-office processes, in combination with the now fairly widespread availability of broadband connections, also have an impact on end-user services. Already, content from various audiovisual archives is accessible online using streaming media technology. Undoubtedly, audiovisual content from the archives will be used more and more.

However, trying to accomplish uniform access to legacy material from different content sources raises further challenges. The list of technical and management challenges that needed to be tackled include:
aligning content (essence and metadata);
setting up business models for sharing and reusing archive content;
building (dynamic) multilingual access to content (a multilingual thesaurus);
building an infrastructure for uploading and editing;
setting up networks with academia and other archives;
and dealing with Intellectual Property Rights (IPR) issues.

This article will focus on several aspects of this work: how the architecture was set up and with which supportive measures to allow the BIRTH Television Archive to grow from its initial stage. Eventually, we hope that BIRTH will establish itself as an invaluable and 'living' resource, with services to cater for many user groups.

USER REQUIREMENTS AND FUNCTIONALITIES

In the first stage, the consortium conducted an extensive user survey in order to establish expectations on the services that could be offered by BIRTH. Over 400 potential users were approached. These users were clustered in three groups:
the general public;
academics;
broadcasting professionals.

Each group has its own specific demands regarding choices of functionality, metadata and content. The challenge was to translate these into functional requirements. Different scenarios were drafted and the development work on the portal began. Wherever possible, existing technologies and open standards were used. On the front end, people can browse through all collections in various ways using the metadata as a starting point, or alternatively using the Flash animated timeline. In detailed view, the programmes are represented by keyframes, and clips are available in several streaming formats. Hundreds of assets from between 1940 and 1970 are currently available, as well as hundreds of stills, vintage articles and programme guides. Some basic metadata elements have been translated into English and BIRTH uses a multilingual thesaurus, enabling users to browse through entire collections from different counties.[61]

[61] This is an excellent example of one of the uses of language technologies in the heritage sector. For more on multilingualism, see DigiCULT's (cont…)

In the final stage of the project, transaction modules will be added, allowing professional users to create profiles and request access to the original material for reuse. Currently, an innovative peer-to-peer architecture is in place to handle these requests and file transfers.

GETTING ORGANISED

An aim in designing the workflow was to provide a means for content suppliers to transfer material to the portal in an easy, effective and aligned way – for the archive partners as well as for the portal managers. For these reasons the Contribution Tool was developed. In general terms, the

Figure 1: Homepage and result list © BIRTH (www.birth-of-tv.org), 2005

Contribution Tool is a client application installed at every archive partner. Its basic functionality is to add information, annotate corresponding metadata and then upload this to the portal. Metadata and images are stored directly at the portal. For performance and legal issues, the video material presented on the BIRTH portal is stored on servers in the country of origin.

The Contribution Tool is at the heart of the BIRTH working processes. It supports various kinds of content: audiovisual material in a variety of formats, still images, audio, articles and generic documents (e.g. PDF). Advanced features such as generating keyframes and transcoding tools are fully integrated in the Contribution Tool, making it the one-stop-shop for publishing content on the portal.

As noted above, the archives already have content and metadata in their legacy catalogues. The Contribution Tool is able to import metadata in various formats, perform mapping processes and output uniform XML files, conforming to the BIRTH metadata model (i.e. Dublin Core with extra qualifiers). To transcend language boundaries, a multilingual thesaurus was set up. The Contribution Tool integrates the thesaurus by allowing the user to select thesaurus terms for several elements of the metadata. The actual management of the terms (such as translations or adding new entries) is carried out by the ThesauriX tool described below.

PROVIDING MULTILINGUAL ACCESS

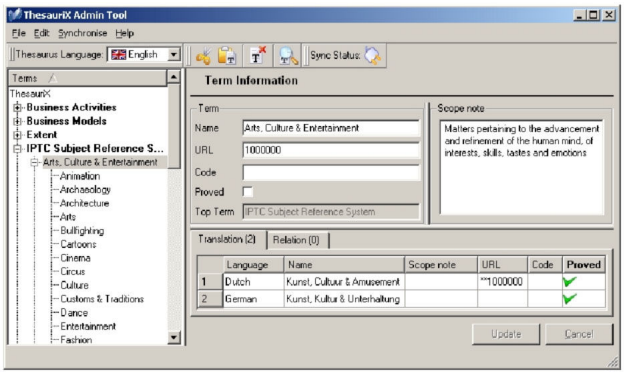
Since material from several countries can be found on the portal, to fully appreciate the content, users need to be assisted in crossing the language barriers that arise. The BIRTH Television Archive provides several means of navigating through the content in a multilingual way:
powerful use of multilingual thesauri;
translation of key elements;
timeline view;
media-rich articles on pre-defined topics.

This section will focus on what has been achieved regarding the multilingual thesauri. When a user types in a query in his native language, it is mapped to terms in the multilingual thesaurus. Subsequently, a new query is launched, containing translations of the original terms, retrieving items from different countries.

BIRTH has built – and adopted – thesauri for the metadata elements: Genres, Keywords and Location. With regard to the keywords, BIRTH offers users of the portal the option to browse a tree-structured list of thematic keywords, available in several languages. The consortium studied several projects and products providing a multilingual taxonomy to be adopted and used for the project. Eventually, the consortium decided to use the IPTC thesaurus (http://www.iptc.org/metadata/) and to add additional keywords to this thesaurus if needed. A similar approach was

Technology Watch Report 3 available free at http://www.digicult.info.

adopted for geographical names. Here, the ISO 3166 list proved useful.



ThesauriX AdminTool © BIRTH (www.birth-of-tv.org), 2005

The ISO and IPTC taxonomies were not available in all languages and in the format required. Handling the translation of well over a thousand terms and the export of these terms into machine-readable XML needed a specialised application. Joanneum Research developed the ThesauriX AdminTool, which has comfortable interfaces allowing the editing of the thesaurus content in a relational manner. The ThesauriX AdminTool enables the user to create and manage one or more thesaurus databases by providing functionality like hierarchical ordering of terms, multi-language support, adding relations (e.g. synonyms) between terms and more. Besides the creating and editing of multiple new terms and translations, the ThesauriX AdminTool provides the ability to synchronise several thesaurus installations with a master database. This offers the archive partners of BIRTH the ability to translate and extend the thesaurus in a simple way – without manually distributing their new terms to the other partners; this is done automatically. Currently, four languages are supported, but using the ThesauriX AdminTool more languages can easily be added.

The IPR issue

Rights issues for Web access to legacy audiovisual content have proved problematic to deal with. Contracts and rights relating to performances, music and programme distribution were drawn up before the advent of computers, the Internet or the digital age. Added to this situation is the diverse legislative landscape in each country relating to copyright and 'broadcasting' on the Internet. Mostly these rights are undefined and fragmented. As such, access to legacy material for Internet use has been inhibited, often leading to frustration for programme makers, academics and the public who find themselves excluded from really useful content.[62] However, the content owners see that this situation is slowly changing. The pressure to resolve the issue of Internet access has been growing in a number of areas both cultural and financial. Computers and Internet costs have decreased, leading to a growing number of households having access to high-speed Internet access. Broadcasters in general have moved heavily into digital channels, many of which rely on archival content. Some state broadcasters have begun to open up their archives as part of cultural and historical celebrations.

To support these growing demands on access, the pressure for clarification on legal definitions has surfaced at an international level, leading to a number of joint initiatives to establish a legal framework in which interested parties can operate. Examples of this include the Santiago Agreement (see http://www.gema.de/engl/communication/news/n163/santiago.shtml for more information) and the 'copyleft' movement, headed by initiatives such as Creative Commons (http://www.creativecommons.org).

End-users and the European Television History Network

As mentioned above, the BIRTH Television Archive reaches out to various user groups. Other major broadcasters are going to contribute content to the portal. The FIAT/IFTA organisation (International Federation for Television Archives) formally endorsed the initiative last year and the EBU (European Broadcast Union) have also shown their interest.

For archives that are interested in joining BIRTH, a 'cookbook' is available, outlining the workflow for publishing material online. In the next stages of the project, BIRTH will continue to focus on how to tackle existing copyright issues in order to make it easier to make contributions and to support a business-to-business service, including the distribution of high-resolution material for professional reuse.

Concerning the link to academia, BIRTH initiated the European Television History Network. The goal of the network is to cluster and integrate European research activities in the field of television history and to develop new research plans. No such network existed previously, making comparative research on a European level almost impossible. Utrecht University coordinates the network in close collaboration with the BIRTH Television Archive.[63] After a first preparatory meeting in October 2004, an expert workshop was organised for April 2005, the first meeting to offer the participants the possibility of discussing future research activities: content, approach, and projects that help to explore a comparative approach to television history in Europe.

The Birth Television Archive acts as the central space where cooperation between academics and archives is disseminated. The network is an excellent example of how archives could meet the needs of academics and conversely how academics could contribute to the knowledge contained in the archives. On the portal, the profiles of the leading academics are accessible and results from the research will be disseminated. Promotion will be one of the main focus points of the BIRTH Television Archive until the end of the project's lifetime. Thereafter, sustainability models and support from participating organisations will ensure the longevity of the initiative.

Conclusions

The BIRTH Television Archive brings valuable historic resources to the public. The original partners succeeded in setting up a rigid workflow for selecting, cataloguing, uploading and ordering the digitised assets. From this base, the BIRTH Television Archive will be extended in time and collections, growing to be an ever-important source for academic research and the enjoyment of the general public.

We are confident that the BIRTH Television Archive will have a significant impact as a rich repository of digital content, a promotion platform for member institutions, a best practice example of pan-European collaboration, and a showcase of how the use of advanced Web technology can increase accessibility of, and interaction with, our cultural heritage.

[62] For a similar example, and its solution, see the case study on AMICITIA in the chapter on Digital Rights Management in DigiCULT *Technology Watch Report 2*, http://www.digicult.info.

[63] More information on the European Television History Network can be found on the Web site of the BIRTH Television Archive at http://www.birth-of-tv.org.

NEMO: Network of European Museum Organisations

Jaana Rytkönen, Secretary of NEMO

NEMO (http://www.ne-mo.org) was founded in 1992 to draw attention to the museums' cause in EU decision-making and to link museum associations and similar organisations within Europe. Since then it has grown into a strong professional network involving 29 European countries. NEMO represents and lobbies on behalf of European museum organisations and provides them with information on relevant European Union initiatives and other issues of importance. It also encourages the exchange of information between museums and related organisations throughout Europe.

Museum size

There are thousands of museums in Europe and the majority of them are small, serving and working with their local and specialist communities. In this context, museums play a significant role in encouraging cultural diversity, identity, education, and awareness of our shared European heritage. Against this background, museums will have a major role to play in integrating Europe.

NEMO has actively promoted the potential of small institutions, such as museums, to participate in EU programmes. The network has closely followed the preparations of the new cultural programme Culture 2007-13, a follow-on programme from Culture 2000. NEMO considers it important that larger institutions should not be the only ones to benefit from EU support in order to develop their activities, but small organisations working locally should also be able to join and contribute to EU programmes.

NEMO activities

An annual meeting in November is the main event of the NEMO year. It is a forum where members can meet each other and discuss the most up-to-date issues affecting museums in Europe. In the meeting discussions, EU policy in the field of cultural heritage and museums is the main issue but participants have the opportunity to hear about the situations in other countries with regard to, for example, digitisation of collections, government support for museums, and developments in the museum field. The executive and the chairman are nominated for a period of two years at the annual meeting, the chair being occupied by all membership countries in turn. At the beginning of 2005, it was held by Finland. NEMO publishes a twice-yearly newsletter, *NEMOnews*, which is distributed through NEMO members to European decision-makers in cultural policy, and regularly meets EU officials in Brussels to promote museums' viewpoints.

Online museum partner search

In addition to lobbying in the EU, one of NEMO's major goals is to improve opportunities for cooperation between European museums. The main objective of NEMO's online partner search mechanism is to enhance the potential of museums to find partners in EU projects. NEMO previously published museums' partner advertisements in *NEMOnews*, and since 2003 NEMO has offered the same service online at http://www.ne-mo.org/partner. This mechanism enables museums and other organisations to publish and search for museums related to their projects regardless of contacts with NEMO member organisations. The shift from printed to digital service has seen a major improvement in the functionality and accessibility of the partner service. *NEMOnews* could never reach the entire museum community in the same way that the Web-based service does.

The service aims to be user friendly. The procedure of adding information to the database and searching available project ideas has been made as simple as possible. Those who need the service should not be forced to spend too much time and energy in learning to use it. The worst case would be that the user becomes frustrated and leaves the site when facing the first obstacle, so efforts have been made to make the site both appealing and easy to use. When the project information is added to the site, there are only five obligatory fields that must be filled in order to get the project idea published: the name of the project; description; the advertiser's country, name, and email address. The other seven metadata fields, which naturally improve the chances of finding partners, are optional: area of interest; the name of the funding programme; name of the organisation; Web site; postal address; telephone number; and fax number.

It is important with any type of Web-based service that it is up-to-date and reflects the needs of its users. In NEMO's partner search, for example, the options in the category of funding programmes were not fixed beforehand, but are inserted into the database as the project proposals come in, adapting to users' changing circumstances. When the user inserts the name of the funding programme in the project description, it will be added as an option to the partner search database if it does not exist there already. The partner search Web site also has basic information on the Culture programme: Calls; application procedures; and outcomes of the Calls. This information can be extremely useful to museums, especially those who have limited experience of submitting proposals in this way. Links to other services, which can be helpful when looking for partners in a specific field, and sources of information on EU funding can also be found on the partner search Web site.

Database of museum information

NEMO collects national museum information through its members about the number of museums (by type), ownership of museums, museum organisations, museum magazines, and legislation affecting museums. This information, which may be of interest to many of DigiCULT's readers, and the national

reports are available on NEMO's Web site at http://www.ne-mo.org.

Future plans

At the moment, mobility of collections and mobility of museum professionals are on NEMO's agenda. These themes are a focus of the next EU Culture programme beginning in 2007, and the next four EU Presidencies are also committed to promoting collection mobility. There are, however, obstacles relating to the practical arrangements of lending collections to other countries. NEMO is willing to cooperate with other European organisations in taking a further step towards addressing these obstacles. This includes surveying the present loan agreements in use in Europe, and promoting the best practices of lending and distributing information. A working solution needs to be found also for a good programme for the exchange of museum professionals. NEMO would like to encourage museum professionals to take the opportunity to visit and learn in other countries, especially those who will be working in the sector for a long time into the future.

NEMO continues its work as a lobbyist for museums in Europe and acts as a forum for collaboration between museums. The next large meeting of NEMO representatives will be in Manchester, UK, in November 2005.

European Workshop on Culture and Technology:

2-D and 3-D Technology Trends for Culture

Report on the Workshop held on 19 and 20 March 2005, Pistoia, Italy

The workshop, sponsored and supported by the Fondazione Cassa di Risparmio di Pistoia e Pescia, was held in the beautiful Fattoria di Celle in Pistoia. The opening and welcome were given by Dr Giuliano Gori, owner of the Fattoria di Celle, and by Professor Ivano Paci, President of the Fondazione Cassa di Risparmio di Pistoia e Pescia. The workshop presented a mix of expertise from academia, industry, governmental and cultural institutions.

The introduction by Vito Cappellini addressed quality (of products), protection and security (including safe systems and IPR), mobility (both geographical mobility and 'infomobility'), multilingualism, access (including new systems and devices), business and commerce, with a scope ranging from local to global. He explained the goal of the workshop: to discuss the state of the art and the future trends in the field with particular reference to European projects. He stressed the importance of Information Technology for Culture (including Cultural Heritage, Performing Arts and Contemporary Art) as a way of improving knowledge, preservation, dissemination and end result by using high-quality 2-D technologies and developing innovative 3-D ones.

Major themes of the workshop

Addressing the need for technology development

A recurring problem concerns the specification/clarification of technology 'needs' in the cultural field. Neither industry nor cultural actors are expert at expressing need in terms that both sides understand. What is required to push forward the research and development agenda? One solution may lie in enabling and encouraging cultural actors to select the appropriate technology from other sectors. Learning from other fields, for example medicine and the gaming industry, is regarded as essential. Researchers, technology developers and cultural partners should work together to explore the needs, questions and solutions identified in other sectors. It is, in other words, important to identify technologies that can be realistically utilised and to look at how to influence their further development. A pressing issue concerns, for instance, restoration and preservation control aided by digital technologies. Many concerned with restoration and preservation control have expressed specific needs for user-friendly tools to enhance and support their work.

Capturing and disseminating best practice, including the use and implementation of standards (not necessarily new ones), is still needed. In this respect, it is good to remember that technology is not merely there for its own sake.

Longevity and digital preservation

With the arrival of new technologies, the issue of digital preservation is becoming more pressing every day. The huge volumes of data that are being produced are increasing the problem. Also, new ways of combining data, data fusion, creation of rich, interactive and VR and 3-D environments, often based on multiple sources that change drastically over time, pose new problems of preservation. It is time this issue was put on the agenda since current research programmes mainly cover typical (structured) data relating to published documents.

In this respect, preservation is still a relatively new issue. Future use, reuse and repurposing of complex objects will have to be addressed, taking into account the prospective users, both within and outside the cultural institutions. Close cooperation between sectors adds to its significance.

Uncovering content, digging for meaning
With the emergence of new means of presentation, the usage of data is changing. Knowledge and interpretation, producing meaning, require different technologies and new approaches. Often within cultural institutions the awareness of the potential of new technologies to record this 'meaning' for public and future use is not apparent. Participants agreed that promoting this awareness should be an immediate priority.

It should be noted that, while 3-D representations and VR cannot replace the experience of the real object, they can help understanding if utilised in the right way. The development itself has to recognise this as it strives for further enhancements. The use of VR for recording culture (e.g. valuable information otherwise lost after an exhibition has finished) needs to be explored more widely. This technology enables the reconstruction of heritage from anecdotal memory.

3-D is fundamental to human perception and therefore essential to the understanding of culture. When applying these technologies, the governing principle should still be interpretation; users should not be constrained by the technology possibilities. The cultural sector needs to express these needs more explicitly to enable further and more user-focused development and research.

Imagination and creativity are needed on both sides to enable a more fruitful cooperation between industrial and cultural sectors. This would stimulate experimentation and the set-up of incubators where different disciplines interact with new technologies. This would be particularly important for those

concerned with 3-D representation and augmented environments. Many technological advances will help in this area, such as haptics and holography.

SHAPING KNOWLEDGE, UNLOCKING DATA

A characteristic of information management within the cultural sector is what could be called 'curatorial protection'. This protection means that potentially paramount information remains excluded from the digital knowledge that is accessible to everyone. It is obvious that universal access to information should be implemented everywhere to enable expanded usage of cultural information.

Access should also encompass availability for best use by everybody. This means that more emphasis should be placed on developments supporting personalisation. In terms of better access to information, we have to look to the reuse, repurposing, registration and data fusion that are commonly applied to vast heterogeneous sources in other disciplines. Technologies that will support this access need to be developed, or existing ones to be enhanced.

This stresses the importance of data harmonisation. If we wish to deliver effective services to the public, we should first look at how information is gathered. This is very important since each new generation uses information in a radically different way. Within the gaming industry, for instance, technology is being deployed from which the cultural sector can gain advantage. Within the same industry, content is developed to which the cultural sector can make a significant contribution. All of this emphasises that the cultural sector should adapt to the radical change in the use of information.

FUNDING

An important conclusion of this workshop was that the large-scale implementation of technologies and exploitation of cultural resources could be better promoted by closing the gap between technology users and technology developers. Technology users, typically the cultural institutions, should re-engineer their business. In this respect they should apply best practices from other sectors. As much as possible, exploitation of cultural resources should be based on creating added value. This applies not only for the cultural sector itself but also to its technological partners. Looking for the content value chain should help identify where opportunities can be found.

Also, as in former times of patronage, sponsorship could be extended and should be looked for more widely.

Often research and developments lead to solutions that deliver an optimum performance but may be too costly for general application across the cultural sector.
While taking into account the durability of the proposed developments, cost-efficient and affordable solutions need to be sought. This should be taken into account before, and not after, parties engage in new developments. Again, communication between sectors – academia, industry and cultural stakeholders – is as essential as the need to have a clear understanding of all users' expectations. In other words, the cultural sector should look more carefully into the long-term development of technologies that support exploitation as well as accommodate appropriate access.

DIGITAL RIGHTS FOR ART MUSEUMS: NEW REPORT

A report, entitled *Reproduction charging models & rights policy for digital images in American art museums,* has been published online. Funded by the Andrew W. Mellon Foundation (http://www.mellon.org/), the report surveyed one hundred American art museums to examine their cost and policy models for delivering digital images and rights services. The report draws conclusions about the financial policies of these museums, and demonstrates that museums do not undertake image digitisation or rights activities because it is profitable, but instead to provide a public service, educational opportunities, to promote the collection of the institution, and to provide a service for publishers and commercial users.

The study recommends that museums, with a better understanding of how these services relate to their overall mission, review their priorities regarding imaging and rights services. It suggests that staff be devoted to these activities on a full-time basis, that departments be appropriately credited for the work they do and that revenue generated be invested back into the department that created it.

Reproduction charging models & rights policy for digital images in American art museums by Simon Tanner is available from http://www.kcl.ac.uk/kdcs/USart.htm.

THE VICODI PROJECT: CREATING A SEMANTIC WEB APPLICATION FOR HISTORIANS

Professor Edward Acton and Mr Richard Deswarte, University of East Anglia Dr Jeremy Boulton, Dr Tim Kirk and Dr Jan Oosthoek, University of Newcastle

INTRODUCTION

Over the past few years the World Wide Web has become the main source of information for millions of people. The development of advanced search engines, such as Google and Yahoo, has made information searching and analysis much more productive. However, the massive increase in information available on the Internet has seen most queries return too many retrieved and/or non-relevant results. Although most search engines rank their results, nonetheless experience shows that users generally need to examine lots of hits from long lists in order to find the most relevant documents. This is time consuming and often frustrating. A second problem is that resources on the Internet, or almost any electronic retrieval system, are currently not interconnected. This results in a situation where a researcher interested in retrieving a particular document is not necessarily aware of the existence of additional relevant documents. Unfortunately, neither is the Internet at present.

In theory, the much-discussed Semantic Web will solve these problems but that seems a long way off. However, the recently completed EU-funded 'Visual Contextualisation of Digital Content' (VICODI: http://www.vicodi.org/)[64] project has been working on a prototype portal that addresses these issues directly. The project has developed a new visual contextualisation, interactive, semantic prototype application for document retrieval called *eurohistory.net*.

VICODI was an RTD project carried out by seven partners from six European countries under the European Union's Fifth Framework IST programme. It concluded successfully in August 2004. The project's main goal was the semi-automatic creation of contextual semantic metadata for digital resources. The context is visualised through historical maps and links shaded according to their degree of importance, thus improving user understanding of the resources. For its showcase domain, VICODI used European history from 500CE to the present time. A prototype Web portal (http://www.eurohistory.net) was developed that puts into practice the idea of contextualising historical resources, like articles and images, and visualising their context.

CORE CONCEPTS

The VICODI system is made up of two core concepts, which on their own are not revolutionary, but combined create a new way of representing the context of documents in a visual way. These two concepts are visualisation and contextualisation.

DEFINING CONTEXT AND CONTEXTUALISATION

The definition of context is highly debated and has been used in many different forms both in the humanities and in computing science. In the Web-based information systems developed from the 1990s onwards the notions of context and contextualisation have become commonplace. In these systems the term 'context' is often used to describe computational models that express domain context and user context. The user background, objectives and interests, including written language, are all part of the user context. Resource or domain context refers to the specialist domain information found in the resources, for the most part written historical texts in this project. This information is identified by using a domain ontology, in this case a history ontology. The VICODI ontology will be discussed in detail later, but it is necessary to recognise that the ontology provides a set of domain classes (the hierarchical grouping of key concepts) that allow for the modelling of context. In VICODI the term 'contextualisation' refers primarily to the process of interpreting and adapting content, content visualisation and navigation to the domain and user contexts.

USING CONTEXTUALISATION IN VICODI

The goal of VICODI was to enhance people's understanding of digital content on the Internet. Semantic meaning was added in two specific ways. First, relevant documents and the associated historical terms or entities found in the text and identified from the ontology as relevant to the context were interconnected with dynamic hyperlinks. Thus, the specific historical context of the actual document was used to filter and rank the results of subsequent queries. Secondly, we exploited the user context by allowing people to set various search and retrieval parameters such as language, location, time interval and category.

Wurman's LATCH (Location, Alphabetic order, Time, Category and Hierarchy)[65] properties were used in identifying mechanisms for storage, retrieval, presentation and navigation. For visualisation purposes, as will become apparent below, the part of the context that specifies Location, Time and Category is particularly important. Indeed, location and time are crucial features in modelling history since these properties represent historical change.

A key element of the contextualisation process was the automatic identification of historical terms in the text. This was combined with the automatic suggestion and listing of further additional historical terms or entities including locations or time intervals, which, while relevant to the resource context, were not explicitly mentioned in the text. The estimation and communicating of the relevance of related historical entities or concepts to resources such as texts and historical maps is essential in defining user context. It also represents the semantic goal inherent in contextualisation: adding extra meaning to resources.

Therefore in the VICODI contextualisation process it was possible, through the use of machine learning, training data and an ontology, to associate ontological terms relevant to the contexts of a text, to estimate the relevance of those entities to the text and maps and to provide additional features that transform and enhance the texts and XML-based visualisations. Further, a specially developed reasoning component

[64] IST-2001-37534.

[65] R.S. Wurman *et al.*, *Information Anxiety 2* (Indianapolis: Que, 2000).

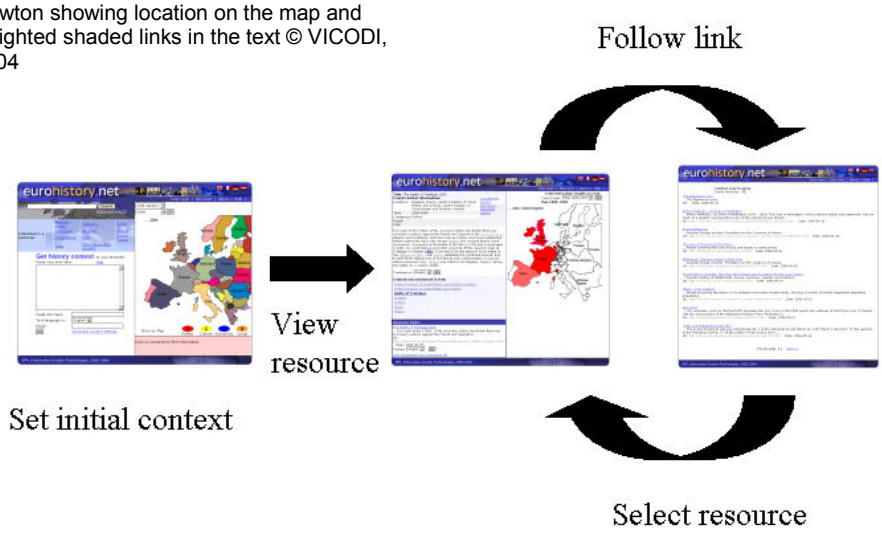
was used to estimate the most relevant locations, time intervals, and categories of a resource.

In order to estimate context automatically for resources it is necessary to build correlations between ontology entities (Correlation Index Vectors). Normally an ontology does not provide weightings that indicate, for example, how relevant a person was in a particular event. However, as indicated above, this is a key element of contextualisation in VICODI. Therefore it was necessary to train the system, using machine learning algorithms that build correlations between ontological entities. In practice, this was achieved by expert historians confirming and correcting the resource contextualisation data. Specifically, an expert using an annotation tool would be presented with a weighted set of elements or instances from the ontology. These weighted elements or instances would be confirmed, new elements added or inappropriate ones removed through an annotation tool. Thus, a historical explanation of the discovery of the laws of gravity would be modelled and weighted as follows:

```
{ Newton:1.0, 1660-
1680:1.0, England:0.8,
gravity:1.0,
apple:0.1, Scientific
Revolution: 0.7 }
```

Figure 1: VICODI interface and Figure 2: Portal workflow © VICODI, 2004

Figure 3: Contextualised document about Newton showing location on the map and weighted shaded links in the text © VICODI, 2004



In this example 'Newton', 'England', 'gravity' and 'Scientific Revolution' are terms associated with instances from the ontology, and the numbers represent relevancy weights on a scale between 0 and 1. Thus, in summation, these numerical weightings are visually represented through shaded colour-coded maps and text hyperlinks that make the context of documents and key terms visible in transformed texts and maps.

VISUALISATION

To represent the context of a document in visible format is another important aim of VICODI. Visualising the actual document context illustrates the semantics of the document and significantly helps in understanding the meaning of the document. By combining visualisation and context-based enhancement with the results produced by the system an approach was developed that we termed visual contextualisation.

Visualisation refers to the process of making the context of a document or key term visible through the colour coding (in our case shading) of hyperlinks or graphical aids such as maps. The purpose of visualisation of information is to make information more comprehensible. It also aids in the identification and analysis of data through visual exploration. The principle of visualisation is to map the characteristics of an abstract data structure to visual attributes such as position, colour and size, and to display the mapping (information).[66] To represent information VICODI developed an innovative interface that employs Scalable Vector Graphics (SVG) as well as a set of grey-scale weighted hyperlinks.

USER INTERACTION WITH THE VICODI SYSTEM

There are several ways to retrieve and visually contextualise documents using the VICODI prototype portal. The first element that can be used for retrieving documents and setting a context is the SVG historical maps of Europe. The user can specify the initial LATCH context visually by selecting a time period and clicking on a particular country in conjunction with the category buttons – politics, culture, economics and social (Fig. 1). Alternatively, users can start with a rough description of context or a full-text search by using the text search facility. Through this option users can build a context search query by adding instances from the ontology. Both a map-based search and a full-text search generate a Google-style list of related resources (Fig. 2).

When accessed by the user, the resources will be contextualised based on certain parameters (e.g. keywords, years, names) that the system uses to generate the appropriate context for the whole article. In the text, colour-shaded hyperlinks represent further possible search queries to the system. The user also receives contextual data in the form of interactive graphical maps. Thereby VICODI provides a visually

[66] A. Triesman, "Preattentive Processing in Vision" in *Computer Vision, Graphics, and Image Processing*, No. 31, pp. 157-176, 1985.

enriched interface to navigate both in place and time. For example, if the user specifies an article about Newton and his theory of gravity, the system will automatically generate a seventeenth-century map from predefined SVG drawings and will put links underneath the map as well as indicating England as the most important location.

The user can navigate further through the repository to find additional resources by clicking either on the contextualised (hyperlinked) terms in the document or on the map to the right of the text. As a result the user gets a new list of documents that the system considers the most relevant to the context of the map or hyperlinked term on which the user last clicked. For example, clicking on the link 'telescope' in a document about Newton will provide documents about this scientific instrument in the context of the Scientific Revolution.

Users may also upload their own historical documents for contextualisation and then use the hyperlinks to find additional and related texts. However, as the initial contextualisation process demands considerable memory and processing power, the results are not instantaneous but posted via email usually within several minutes. Previous results stored in the VICODI database are instantaneous.

Components of the VICODI portal

To make the VICODI system work and implement the workflow described above, several different components interact with each other. These various components are not innovative elements in their own right but combined they form the unique features of the VICODI prototype. The system is composed of the following elements: context engine, transformation engine, multilinguality, Scaleable Vector Graphics, and the ontology.

Context engine, transformation engine and multilinguality

An important functionality of the context engine is, in essence, to use text categorisation to build correlation scores between documents and the notions in the VICODI ontology. This allows the system to enhance the documents' visualisation and linkage to give users a faster and more intuitive understanding of a document's position among the notions represented in the ontology. The transformation engine processes the data of the relevant contextual information from the context engine and outputs it by either transforming it into SVG instances (dynamic maps) or generating hyperlinked (contextualised) documents for display in the user interface. The transformation engine supports both graphics-based visualisation and text-based presentation of the contexts. Another novel aspect of VICODI is the multilinguality offered by Systran.[67] The KAON ontology framework (http://kaon.semanticweb.org/) provides a lexical layer on top of a language-independent ontology core for each language. Using that feature it is possible to translate the language-dependent part of the ontology without disturbing its logical structure. In addition, retrieved resources displayed in the interface can be translated via an automatic translation tool into English, German, French and Latvian.

Scalable Vector Graphics (SVG)

In the VICODI system, locations found in a document are displayed through a graphical contextualisation interface based on SVG (http://www.w3.org/Graphics/SVG/). XML-based SVG is a distinct improvement over previous bitmap formats, such as GIF, PNG or JPEG, because it has smaller file sizes and independent scalability and will provide cross-browser compatibility in the near future.[68] Probably the most impressive visual elements in the VICODI interface are the historical maps of Europe. Over one hundred historical dynamic SVG maps of Europe dating from 500CE to the present have been created. As indicated above, these maps are dynamic and an integral part of the display, retrieval and contextualisation of the historical resources.

The ontology

At the heart of the VICODI system lies the history ontology. To compensate for the lack of extensive or comprehensive general electronic history repositories, an ontology of history based primarily upon an empirical deductive analysis of a 2,000-document corpus was built. The main purpose of the ontology for the VICODI project is to help machine algorithms in the automatic contextualisation task by storing relevant historical knowledge in machine-processable form. The task of devising an ontology of history from scratch was certainly daunting. On the one hand, it is always challenging to build an ontology covering a broad and very complex area of knowledge. On the other hand, history has several unique features that are problematic from the viewpoint of ontology construction. Above all, the complexity of history is immense and potentially requires an almost unlimited number of instances and property relations. To model this complexity practically, we introduced a shallow concept hierarchy starting from only seven basic concepts (called flavours), which are meaningful for domain/history experts: individual, event, abstract notion, organisation, object, social group and location (Fig. 4). The hierarchy below these concepts is shallow (2-3 levels) and stops at an abstraction level which, while still meaningful for historians, is still general enough to make the placement of new instances in the ontology easy, thus speeding up the population of the ontology with new historical knowledge. The

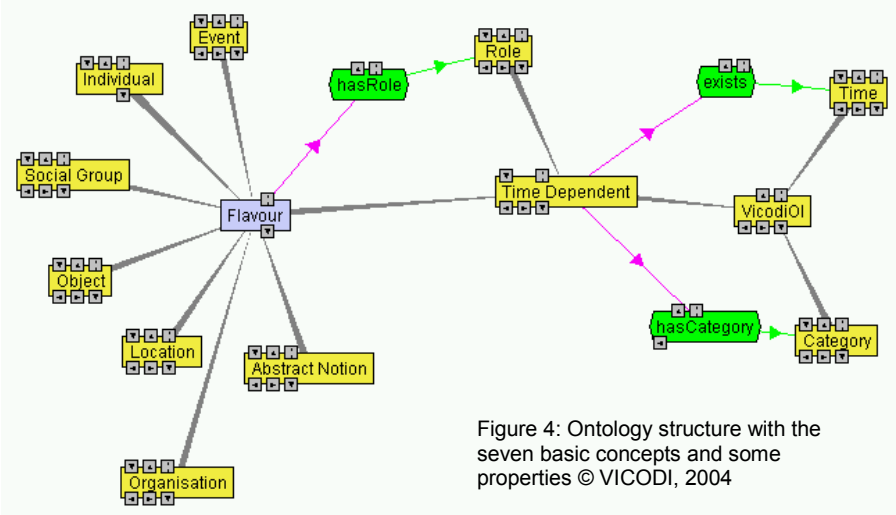


Figure 4: Ontology structure with the seven basic concepts and some properties © VICODI, 2004

complexity of history is further represented by connecting instances of these flavours with a limited number of property relations (15) and roles. The roles are used to model changes in the meaning and functions of instances over time. For example, Winston Churchill had numerous different roles such as Prime Minister, MP, journalist and author.

[67] SYSTRAN, WhitePapers, MT Summit 2001-2003. http://www.systransoft.com/Technology/WhitePapers.html

[68] At present a plug-in is needed in the majority of browsers to see SVG images. The eventual goal is to display SVG images without any need for plug-ins.

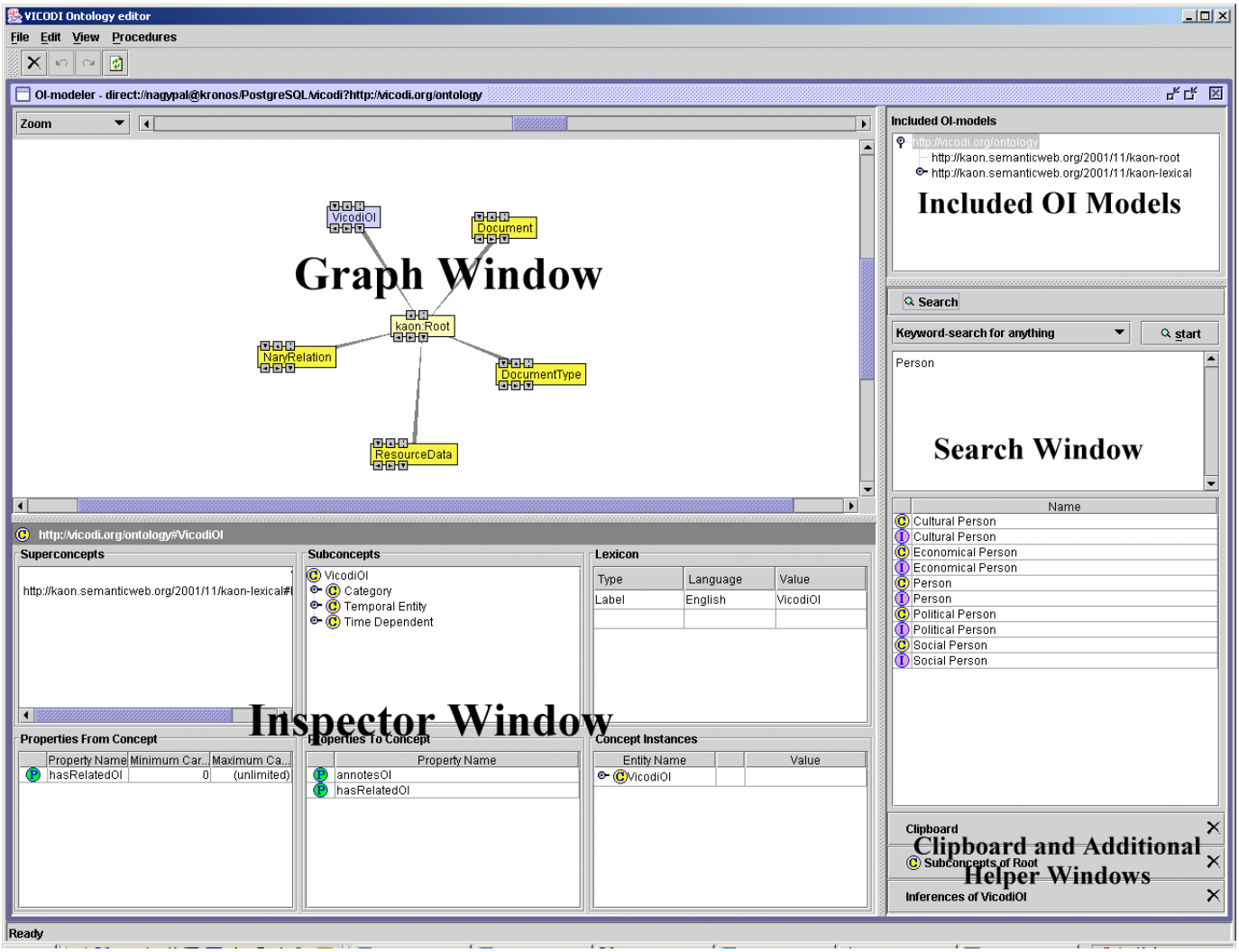


Figure 5: KAON ontology editor interface © VICODI, 2004

Another complicating factor in modelling history is the fact that historical time is often uncertain and debated. History includes many unknown dates, imprecise intervals and overlapping time (historical periods and events extending into each other without clear start and end dates). Moreover, many historical terms are time-dependent. To deal with these challenges but also to model change we added interval times, called existence time, to each instance in the ontology. In this way instances with a common time-dependent relation could be connected. For the VICODI prototype the intervals are precisely defined, although a novel fuzzy temporal model is now being examined.[69]

The ontology was developed using the Open Source KArlsruhe ONtology (KAON) framework, which provides an extension of the W3C RDFS standard.[70] Thanks to KAON's capability to provide programmatic access to the ontology it was possible to add a huge number of instances and concepts to the ontology by processing textual glossaries, databases or spreadsheets. This – together with our simple and intuitive concept hierarchy – significantly speeded up the ontology populating process, as the historians could use their preferred software tools (e.g. Excel) while codifying their knowledge. The manual functions of the ontology editor (Fig. 5) were therefore only needed to carry out advanced operations, like relocating existing concepts and instances, adding new connections and visualising the existing ontology structure. Nevertheless, the process of developing and populating the ontology was an extremely time-consuming process taking over 24 person-months.

The VICODI history ontology is freely available on the VICODI Web site (http://www.vicodi.org) under a GNU Free Documentation License (GFDL).

LESSONS LEARNED

The present VICODI prototype combined several innovative technologies to create an experimental history knowledge space for the Semantic Web. Although the outcome of the project was successful, we believe that VICODI was in several respects over-ambitious. In terms of proof of principle the goal of creating a functioning semantic humanities portal was most definitely achieved. However, in terms in creating a usable tool for historical research and learning, the project was less successful. As emphasised in the user feedback there are a number of elements that need to be improved. The contextualisation process is very demanding of server capacity and this resulted in a slow system. The present interface is overcomplicated for the average user and the present database of resources is much too small for either research or learning purposes. Users also indicated that the 'shade coded' hyperlinks representing weighted instances were ambiguous,

[69] Boris Motik and Gábor Naypál, "A Fuzzy Model for Representing Uncertain, Subjective and Vague Temporal Knowledge in Ontologies" in Robert Meersman and Zahir Tari (eds) *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA and ODBASE*, pp. 906-923 (Heidelberg: Springer, 2003).

[70] See http://www.w3.org/RDF/. Boris Motik, Alexander Maedche and Raphael Volz, "A Conceptual Modeling Approach for Semantics-Driven Enterprise Applications" in Robert Meersman and Zahir Tari (eds) *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA and ODBASE*, pp. 1082-1099 (Heidelberg: Springer, 2002).

therefore reducing the added context value. Our problem in this regard was how to better rescale the relevance rankings for each instance so that they could be expressed in the few colours used by the portal. More user feedback is required to create a better solution.

The different components did work together but did not always produce the high-quality search results needed for teaching or research purposes. We believe that the different components need to be further developed independently before attempting to reintegrate them. In addition, a more extensive document corpus should be compiled on a thematic basis rather than trying to cover all of European history.

Nevertheless, the VICODI project did achieve one very important result. The creation of the VICODI prototype proved the principle that it is feasible to build a working humanities Semantic Web portal. The project also raised a number of new research questions, not least those related to the construction of a history knowledge system – the ontology. The VICODI ontology development process has shown that a complex humanities domain can be represented through a shallow ontology structure and a limited number of concepts and properties. One of the questions that still remains is whether constructing an ontology for any humanities domain is too labour intensive and too costly. VICODI has shown that, by applying mass upload techniques (through, for example, Excel spreadsheets), it is possible to construct an ontology of over 15,000 instances in a relatively short time-scale. Furthermore, the ontology is extensible and could be expanded or integrated with other ontologies in the future. Finally, the project showed that the development of a history portal requires the input of historians to achieve useful historical applications. Without historians it would have been almost impossible to model the history ontology in a meaningful way. To create an ontology you need both technical and subject specialists.

The project also raised the fundamental question of what the Semantic Web can do for history and the humanities, an area of research that is still very much under-explored.[71] It is not yet clear if ontologies will really make searches in complex text collections more efficient than through other available technologies. VICODI also raised the important questions of whether meaningful 'context' can be created and whether making the context of historical documents and relevant terms visible will result in a better understanding of those historical sources. These are methodological questions that require further research.

The VICODI project has shown how a semantic interface and system for a humanities subject could be formulated and constructed. Yet, at the same time its weaknesses, including slowness, an overcomplicated interface and mixed search results, illustrate the massive gap between proof of principle and an end-product.

ACKNOWLEDGEMENTS

We would like to thank Bob Mulrenin from Salzburg Research who provided invaluable comments on the content and the written text and Gábor Naypál of FZI for the useful information he graciously provided.

CONTACT THE PEOPLE BEHIND DIGICULT

UNIVERSITY of GLASGOW

salzburg research

Director, HATII
Seamus Ross, s.ross@hatii.arts.gla.ac.uk
DigiCULT.Info Content & Managing Editor
Daisy Abbott, d.abbott@hatii.arts.gla.ac.uk
HATII/DCC Intern in Multimedia and Digital Curation
Marta Ribeiro, m.ribeiro@hatii.arts.gla.ac.uk
DigiCULT Web Developer
Brian Aitken, b.aitken@hatii.arts.gla.ac.uk

DigiCULT Project Manager
John Pereira

DigiCULT Thematic Issue Scientific Editor
Guntram Geser

The Humanities Advanced Technology and Information Institute, Glasgow University © HATII, University of Glasgow, 2002

[71] See also Onno Boonstra, Leen Breure and Peter Doorn, *Past Present and Future of Historical Information Science* (Amsterdam: NIWI/KNAW, 2004).